

ISA² ACTION 2017.01

**STANDARD-BASED ARCHIVAL DATA MANAGEMENT,
EXCHANGE AND PUBLICATION**

STUDY

FINAL REPORT

DOCUMENT METADATA

Property	Value
Release date	15/06/2018
Status:	Final version
Version:	V1.00
Authors:	Susana Segura, Luis Gallego, Emmanuel Jamin, Miguel Angel Gomez, Seth van Hooland, Cédric Genin
Reviewed by:	Lieven Baert, Julie Urbain, Annemie Vanlaer, Belá Harsanyi, Razvan Ionescu, Reflection Committee
Approved by:	

DOCUMENT HISTORY

Version	Description	Action
0.10	First draft	
0.90	Version for Review	
0.98	Second version for Review	
0.99	Third version for Acceptance	
1.00	Final version	

TABLE OF CONTENTS

Table of Contents	3
List of Figures	8
List of Tables	10
1 Executive Summary	14
2 Introduction	16
2.1 Context	16
2.2 Objectives of the Final Report.....	17
3 Approach and methodology	18
3.1 General approach.....	18
3.2 Sources	19
3.2.1 Archives.....	19
3.2.2 Initiatives and projects.....	20
4 Business Processes	25
4.1 Objectives.....	25
4.2 Approach	25
4.3 Stakeholders	25
4.4 Functional Model	26
5 Standards and Best Practices	29
5.1 Objectives of the inventory	29
5.2 Main findings	29
5.2.1 Overview of standards and gaps.....	30
5.2.2 Cartography of standards.....	31
6 IT Tools and Solutions	33
6.1 Approach and methodology	33
6.2 Overview of IT Tools	34
6.3 IT Tools inventory template	36
6.4 Assessment.....	37
6.4.1 Conclusions of the gap analysis.....	38
6.4.2 Cartography of IT Tools	39
7 Linked and Open Data Initiatives	40
7.1 Introduction.....	40
7.2 Context and methodology	41
7.2.1 Introduction to Linked and Open Data	41
7.2.2 Relevance of Linked Data for the archival community	48
7.2.3 Relation to the automation of Records and archival practices.....	53

7.2.4	State of play within the EC on Interoperability	58
7.3	Inventory of current solutions	63
7.3.1	Cartography of the Linked Open Data initiatives	70
7.3.2	Potential for Linked Open data at the Archives of the EC	71
8	Conclusions	76
8.1	Inventory of Business Processes (D02)	76
8.2	Inventory of Standards (D01)	77
8.3	Inventory of IT Tools (D03).....	78
8.4	Inventory of Linked Data and Linked Open Data initiatives (D06)	79
8.5	Comparative Assessment Tool (D04)	80
9	Recommendations for the Proof of Concept	82
9.1	Architecture of the overall solution	83
9.2	Proposition for a Proof of Concept	86
9.2.1	Approach	86
9.2.2	Roadmap	92
10	Annex A: Business Processes	95
10.1	Delivery.....	95
10.1.1	Pre-ingest (digital archives).....	95
10.1.2	Transfer (non-digital archives)	97
10.1.3	Digitisation.....	98
10.2	Acquisition	99
10.2.1	Ingest (digital archives)	99
10.2.2	Accession (non-digital archives)	100
10.3	Archives Processing.....	101
10.3.1	Appraisal & Selection	101
10.3.2	Sensitivity Review	103
10.3.3	Description / Indexation / Enrichment	104
10.4	Preservation	104
10.4.1	Storage Management.....	104
10.4.2	Preservation planning (digital archives)	106
10.4.3	Elimination.....	108
10.5	Administration.....	109
10.5.1	Reporting.....	109
10.5.2	Request Management.....	109
10.5.3	Master data management	110
10.5.4	Authority Lists Management.....	112
10.6	Access	113

10.6.1 Reading Room	113
10.6.2 Standard Publication	114
10.6.3 Open data publication	115
10.7 Consultation	116
10.7.1 Search & Request	116
10.7.2 Loans.....	117
10.8 Data Exchange.....	118
10.8.1 Data Import.....	118
10.8.2 Data export.....	119
11 Annex B: Standards and Best Practices	121
11.1 Introduction.....	121
11.2 Archival description	122
11.2.1 Data content standards	123
11.2.2 Data structure standards.....	135
11.2.3 Data value standards	145
11.3 Archives Interoperability	155
11.3.1 Semantic Web Standards	155
11.3.2 Standards for Repository Interoperability.....	160
11.4 Records management.....	163
11.5 Preservation	169
11.6 Security.....	182
11.7 Storage Information.....	185
11.8 Metrics and key performance indicators	188
11.9 Practices in other areas.....	190
Automatic data classification	190
File Analysis.....	191
12 Annex C: IT Tools and Solutions	192
12.1 AMLAD	192
12.2 Archivematica	193
12.3 Preservica	195
12.4 RODA	196
12.5 ArchivesSpace	198
12.6 AtoM.....	200
12.7 CuadraSTAR.....	203
12.8 Eloquent Archives	204
12.9 Axiell CALM.....	206
12.10scopeArchiv.....	208
12.11ArchiDoc.....	210
12.12Archeevo	211
12.13Preservica – ArchivesSpace Connector.....	214

12.14	Preservica – Axiell CALM Connector	215
12.15	ArchivesSpace – AtoM – Archivematica connector	216
12.16	E-ARK Extraction Tools.....	217
12.17	E-ARK Access tools	218
12.18	Archivematica (DIP provider)	220
12.19	Preservica OAI-PMH API	221
12.20	AtoM OAI-PMH plugin	222
12.21	CONTENTdm	223
12.22	Omeka	224
12.23	Neatline	225
12.24	HPE ControlPoint	226
12.25	IBM Watson.....	228
12.26	Moriarty.....	229
12.27	SAS	230
12.28	Blancoo.....	231
12.29	OpenText Redact-it	233
12.30	List of archives management systems	234
13	Annex D: Self-Assessment Tool	238
14	Annex E: Linked and Open Data Initiatives.....	239
14.1	Inventory of Open Data and Linked Data initiatives	239
14.1.1	Open Data Portal project (Publication Office)	239
14.1.2	Pan-European Open Data Portal (DG CONNECT)	240
14.1.3	LOGD (Linked Open Government Data).....	242
14.1.4	Estonian – Open Data portal.....	244
14.1.5	Europeana.....	247
14.1.6	Archives Portal Europe Foundation	250
14.1.7	H2020 Holocaust archiving	253
14.1.8	Swiss – Federal Archives.....	256
14.1.9	Cellar (Publication Office)	263
14.1.10	Archives Hub (UK).....	266
14.1.11	Social Archive - SNAC project	269
14.1.12	Bibliothèque Nationale de France (BNF)	271
14.2	Inventory of data structuration approaches and tools	274
14.2.1	CIIM.....	274
14.2.2	AD tool – Preservica	276
14.2.3	GraphDB	277
14.2.4	Silk framework.....	279
14.2.5	LinkedPipes (ETL for Linked Data)	281
14.2.6	Publication Office Preservation (Cellar approach)	282
14.2.7	CORDIS – Data Structuration process	284
14.2.8	Project Open Data.....	288
14.2.9	Guidelines on Open Government Data for Citizen Engagement	290

14.2.10 TopBraid EDG (Enterprise Data Governance) - Data Governance for LOD	292
Terms and acronyms	297
Glossary.....	297
Acronyms and abbreviations	306
References	309
Desk research sources.....	309
Bibliography	315
Interesting readings	318

LIST OF FIGURES

Figure 1: Stakeholders map	18
Figure 2: Desk research	19
Figure 3: 2017 Magic Quadrant for Master Data Management Solutions	24
Figure 4: Proposed Model	27
Figure 5: Methodology of the study	34
Figure 6: Cartography of IT tools	39
Figure 7: Overview of some data quality issues within the data set SG DORIE data made available	43
Figure 8: Overview of data models (van Hooland & Verborgh, 2014)	44
Figure 9: Serialisation formats (van Hooland & Verborgh, 2014)	45
Figure 10. RDF triple example	46
Figure 11: Comparison of four data models (van Hooland & Verborgh, 2014)	47
Figure 12: Example of the usage of Linked Data within the Google search interface	48
Figure 13: Example provided by Popovici (2016)	50
Figure 14: Proposal of an extension for Schema.org by the W3C Archetypes working group	52
Figure 15: Search interface of the COM files of the EC archives	56
Figure 16: Manual labelling of TM results with Eurovoc	57
Figure 17: Roadmap of linked open government data (Ding, Peristeras, & Hausenblas, 2012)	61
Figure 18: General approach to analyse the Open Data and Linked Data initiatives	64
Figure 19: Magic Quadrant for Linked Open Data initiatives	70
Figure 20: High-level overview of data architecture, from EC internal documentation	72
Figure 21: Metadata encoded on file level, from EC internal documentation	73
Figure 22: Overview of the archival filing plan	74
Figure 23: Technical Architecture	84
Figure 24: Technical Architecture	85
Figure 25: Step 1 – Internal Processes	87
Figure 26: Step 2 – Authority Lists Management	87
Figure 27: Step 3 – Data Exchange	88
Figure 28: Ranking of archives management tools by business process criteria	90
Figure 29: Ranking of archives management tools by standards criteria	90
Figure 30: Report on users' community of Archives Space	91
Figure 31: Clients of Archivematica	92
Figure 32: Software stack	92
Figure 33: PoC Roadmap	93
Figure 34: Pre-Ingest	95
Figure 35: Transfer (non-digital records)	97
Figure 36: Digitisation	98
Figure 37: Ingest (digital archives)	99
Figure 38: Accession (non-digital archives)	100
Figure 39: Appraisal & Selection	102
Figure 40: Sensitivity Review	103
Figure 41: Description / Indexation / Enrichment	104

Figure 42: Storage Management	105
Figure 43: Preservation Planning.....	107
Figure 44: Elimination	108
Figure 45: Reporting	109
Figure 46: Request Management	110
Figure 47: Master Data Management	111
Figure 48: Authority Lists Management	112
Figure 49: Reading Room	114
Figure 50: Standard Publication.....	114
Figure 51: Open Data Publication.....	115
Figure 52: Search & Request.....	116
Figure 53: Loans	117
Figure 54: Data Import	118
Figure 55: Data Export.....	119
Figure 56: Sections of ISAD(G)	123
Figure 57: Example of transfer record included in ISAD(G)	123
Figure 58: OAI-PMH Interchange	161
Figure 59: Standards in Records Management scope	164
Figure 60: Interaction with OAIS	170
Figure 61: E-ARK project research	181
Figure 62: GDPR Regulation Key Topics	183
Figure 63: Text- based classification	190
Figure 64: Semantic-based classification	191
Figure 65: File Analysis.....	191
Figure 66: ArchiveSpace Architecture	199
Figure 67: Scope Archive product range	209
Figure 68: Functional architecture of Archeevo	213
Figure 69: Access tools E-ARK.....	219

LIST OF TABLES

Table 1: E-ARK Desktop Research	21
Table 2: ICA Desktop Research	22
Table 3: ICSTI Desktop Research	22
Table 4: Functional Blocks	28
Table 5: Mapping between standards and business processes	31
Table 6: Cartography of standards	32
Table 7: IT Tools Template	37
Table 8: Principles of the EIF	60
Table 9: Linked Data workflow	63
Table 10: Template for the description of the initiatives	65
Table 11: Inventory of solutions	69
Table 12: Archimate standard elements legend	83
Table 13: Archimate specific archiving elements legend	83
Table 14: List of PoC activities	94
Table 15: Pre-Ingest (digital archives)	96
Table 16: Transfer (non-digital archives)	98
Table 17: Digitisation (non-digital archives)	98
Table 18: Ingest	100
Table 19: Accession	101
Table 20: Appraisal & Selection	103
Table 21: Sensitivity Review	103
Table 22: Description / Indexation / Enrichment	104
Table 23: Storage Management (Non-digital archives)	106
Table 24: Storage Management (digital archives)	106
Table 25: Preservation Planning	107
Table 26: Elimination (non-digital archives)	108
Table 27: Elimination (digital archives)	109
Table 28: Reporting	109
Table 29: request Management	110
Table 30: Master Data Management	112
Table 31: Authority Lists Management	113
Table 32: Reading Room	114
Table 33: Standard Publication	115
Table 34: Open Data Publication	116
Table 35: Search & Request	117
Table 36: Loans	118
Table 37: Data Import	119
Table 38: Data Export	120
Table 39: Inventory template	122
Table 40: ISAD(G)	126
Table 41: ISAAR(CPF)	128

Table 42: ISDIAH	129
Table 43: ISDF	130
Table 44: DACS	132
Table 45: RAD.....	133
Table 46: RiC-CM	134
Table 47: EAD.....	137
Table 48: EAD 2000.....	137
Table 49: EAD 3.....	138
Table 50: apeEAD.....	140
Table 51: EAC-CPF	141
Table 52: EAG.....	142
Table 53: ISO 15836 Dublin Core DC.....	144
Table 54: MARC Family	145
Table 55: Conceptual definitions of different lists of terms (Society of American Archivists, 2017)	146
Table 56: EuroVoc	147
Table 57: LCSH.....	148
Table 58: NAF	149
Table 59: LCC	150
Table 60: VIAF	151
Table 61: UKAT.....	152
Table 62: UNESCO Thesaurus.....	153
Table 63: TGN.....	154
Table 64: GeoNames	155
Table 65: Semantic Web Standards (W3C, 2017).....	156
Table 66: RDF	157
Table 67: RDF Schema.....	158
Table 68: OWL	159
Table 69: SKOS	160
Table 70: SPARQL.....	160
Table 71: OAI-PMH.....	162
Table 72: OAI-ORE	163
Table 73: ISO 15489.....	166
Table 74: ISO 30300 Series.....	167
Table 75: MoReq Family	168
Table 76: ISO 23081	169
Table 77: OAIS (ISO 14721)	171
Table 78: PAIMAS (ISO 20652)	172
Table 79: PREMIS	173
Table 80: METS	174
Table 81: MODS	175
Table 82: LoC Format Guidelines	176
Table 83: FADGI Federal Agencies Digital Guidelines Initiative.....	178
Table 84: DCC Curation Reference Manual	179
Table 85: DPC Digital Preservation Handbook.....	180

Table 86: ISO/IEC 27038	185
Table 87: AMLAD	193
Table 88: Archivemata	195
Table 89: Preservica	196
Table 90: RODA	198
Table 91: ArchivesSpace	200
Table 92: AtoM	203
Table 93: CuadraSTAR	204
Table 94: Eloquent Archives	206
Table 95: Axiell CALM	208
Table 96: scopeArchiv	209
Table 97: ArchiDoc	211
Table 98: Archeevo	214
Table 99: Preservica – ArchivesSpace Connector	215
Table 100: Preservica – Axiell CALM Connector	216
Table 101: ArchivesSpace – Atom – Archivemata connector	217
Table 102: E-ARK extraction tools	218
Table 103: E-ARK access tools	220
Table 104: Archivemata (DIP provider)	221
Table 105: Preservica OAI-PMH API	222
Table 106: Atom OAI-PMH plugin	223
Table 107: CONTENTdm	224
Table 108: Omeka - top-down approach	225
Table 109: Neatline	226
Table 110: HPE ControlPoint	227
Table 111: IBM Watson	229
Table 112: Moriarty	230
Table 113: SAS	231
Table 114: Blancoo	233
Table 115: OpenText Redact-it	234
Table 116: Open Data Portal Project (Publication Office)	240
Table 117: Pan-European Open Data Portal (DG CONNECT)	241
Table 118: LOGD (Linked Open Government Data)	244
Table 119: Estonian Open Data Portal	247
Table 120: Europeana	250
Table 121: Archives Portal Europe Foundation	253
Table 122: H2020 Holocaust Archiving	256
Table 123: Swiss Federal Archives	263
Table 124: Cellar (Publication Office)	266
Table 125: Archive Hub (UK)	269
Table 126: Social Archive - SNAC Project	271
Table 127: Bibliothèque Nationale de France (BNF)	274
Table 128: CIIM	275
Table 129: AD Tool – Preservica	277

Table 130: GRAPHDB	279
Table 131: SILK Framework	281
Table 132: LinkedPipes (ETL for Linked Data)	282
Table 133: Publication Office Preservation (Cellar approach)	284
Table 134: CORDIS - Data structuration process	287
Table 135: Project Open Data	290
Table 136: Guidelines on open government data for citizen engagement	292
Table 137: TopBraid EDG (Enterprise Data Governance) - Data governance for LOD	296
Table 138: Glossary	306
Table 139: Acronyms and abbreviations.....	308
Table 140: Main sources	314
Table 141: E-ARK deliverables for IT tools inventory	315

1 EXECUTIVE SUMMARY

“A people without the knowledge of their past history, origin and culture is like a tree without roots.”

Marcus Garvey.

ISA² is the European Commission's programme which supports cross-border and cross-sectorial public services for public administrations, businesses and citizens in Europe. One of its aims is to develop, maintain and promote an integrated approach to interoperability in the EU and to contribute to the development of reusable IT solutions at European, national, regional and local levels of public administration.

In this context, the ISA² action 2017.01 *“Standard-based archival data management, exchange and publication”* was launched by DG DIGIT (DIGIT.B.2) and OIB (OIB.OS.1.002, being the Historical Archives Service of the European Commission (HAS)). This study forms the first phase of this action and has the following objectives:

- On the one hand, provide an overview of the current landscape as far as the business processes and the use of data standards and IT tools implemented by archival institutions are concerned. This information can be used by the HAS for the selection of their future archives management system, and can serve as a reference and inspiration for other organisations and institutions in Europe that deal with archives management.
- On the other hand, analyse the latest trends and options for publication of relevant archival data as (Linked) Open Data. In the analysis, elements that will streamline interoperability, such as the use of shared reference data or authority lists get special attention.

The project team performed desk research and conducted a series of interviews with recognised archival institutions from all over Europe to gather information about their business processes, standards, best practices and IT tools. These institutions also shared their experiences and the lessons learned from the challenges and difficulties that they encountered during the implementation of their archives management systems. This led to the elaboration of three inventories¹, whose details are provided in annexes, which were the first step towards the creation of the global overview.

Based on these inventories, and in order to fulfil the objectives of the study – the selection of an archives management system –, a self-assessment tool was developed. This tool can be configured to generate a series of recommendations on which IT tool(s) fits best the specific needs of the archival institution.

The research underscored the high degree of heterogeneity existing among archival institutions and the lack of significant standardisation efforts in the sector. This added complexity to the definition of an approach applicable to all institutions. Nonetheless, a high level functional model was defined that can serve as a generic model for other archival institutions and as a starting point for the next steps of the ISA² action, namely the implementation of several pilots.

Alongside the collection of information, a workshop with representatives of several archives services of European Institutions was organised to allow each organisation to express its needs with regards to

¹ Inventory of Business processes (available in Annex A); Inventory of Standards and Best Practices (available in Annex B); Inventory of relevant IT Tools and Solutions (available in Annex C).

interoperability and exchange of data. The conclusion after this workshop was that there is a shared interest and willingness to work together towards more standardisation, in particular by managing and sharing authority lists and reference data, and that the Historical Archives of the European Union based in Florence could play a pivotal role in the coordination and standardisation of this process.

The study came to the following conclusions:

- **Business processes.** A generic functional model has been defined which extends the OAIS model to also support the full spectrum of archives management activities. It identifies the main functional blocks and describes the key business processes that are primarily performed by archival institutions and to some extent by other stakeholders. Special emphasis is put on the needs for enhanced interoperability and data exchange.
- **Standards.** On the one hand, a large set of generally adopted national and international standards is available for archival description. On the other hand, there are not many standards for other traditional archival processes. Each archival institution is addressing this gap by developing its own guidelines and custom-made procedures.
- **IT tools.** A combined implementation of archives management and preservation systems can cover most of the internal business processes of the archival institutions. Nonetheless, custom development might be needed to cover their full range of activities, especially with regards to publication and exchange of data.
- **Open Data and Linked Data.** Although these technologies, when fully implemented, could foster the publication of archival data, they are still in an early stage of development, so their adoption should be progressive and should complement traditional publication approaches rather than replace them.

Based on these conclusions, a Proof of Concept (PoC) is proposed to explore the viability of the recommendations that will be presented later in this study for the implementation of an archives management system in the HAS. The proposed PoC consists of three main steps:

- Implementation of the key business processes performed by archival institutions;
- Implementation of an application to manage authority lists shared with other stakeholders;
- Set up of data exchange processes and interfaces, focusing on the exchange and enrichment of metadata between HAS and the Historical Archives of the European Union in Florence.

Three software stacks are proposed in order to test the suitability of the open source and commercial solutions available on the market.

2 INTRODUCTION

2.1 CONTEXT

The ISA² programme, which is running since 1 January 2016 to 31 December 2020, supports the development of digital solutions that enable public administrations, businesses and citizens in Europe to benefit from interoperable cross-border and cross-sector public services.

In the context of the 2017 ISA² Work Programme, the action 2017.01 "Standard-based archival data management, exchange and publication" aims to improve the interoperability of archives services and develop a methodology that can be reused by any archive service in Europe. It begins by obtaining an overview of how public archives services – at EU, national and local levels – are managing their archives and how they make them available to the general public. These activities have been broken down into two Work Packages (hereafter WP1 and WP2).

WP1 starts with a study aiming at providing a generalised business process model for archives management related activities, identifying existing data standards in archival information management and exploring supporting IT tools and services. This analysis will allow determining how these standards are applied in different contexts and how they can be used when managing the archives of the European Commission, i.e. if they are paper-based, digitised or born-digital files. This WP deals also with the analysis of the interoperability requirements among the Historical Archives of the European Commission compared to those of the Historical Archives of the European Union (managed by the European University Institute in Florence). Additional stakeholders such as the Archives Portal Europe are also being taken into consideration.

WP2 is dedicated to the analysis of available options to publish the relevant part of the content of EU archives in a Linked Open Data format, through the different available platforms (like the Open Data and the European Data portals) allowing Member States, researchers and citizens to interoperate with them and reuse them in different contexts.

Prior to these two Work Packages, an extensive data collection task was carried out, consisting of desk research and interviews with key stakeholders in order to gather the necessary information to prepare the required inventories.

Work Package WP1 focuses on building three inventories for core business processes of the archives management workflow, applicable and/or useful standards, and supporting IT tools. Three tasks were conducted in parallel and led to three intermediate deliverables whose outcomes are listed in Annexes A, B and C:

- Task-1.1 - Discovery of archival standards: the identification and categorisation of commonly used and accepted international and national standards related to archives management and exchange of archival data (Annex B);
- Task-1.2 - Discovery of business processes: the identification of common business processes – and their corresponding tasks and activities – in archives management (Annex A);
- Task-1.3 - Discovery of archives management IT tools and services: the identification and analysis of standard-based IT tools and services used to manage archives (Annex C).

Results of Tasks 1.1, 1.2 and 1.3 provided the main source of information for the next tasks of the Work Package WP1:

- Task-1.4 - Design of an evaluation framework: A comparative assessment tool that could easily be used by public administrations to assess the best tools for their needs. It should enable the

archival community to define the minimal requirements for archives management systems and archival data exchange.

- Task-1.5 - Conclusions and recommendations: a document containing the relevant conclusions for the Historical Archives Service of the European Commission (HAS) and recommendations for setting up Proof of Concepts (PoC) in the next phase of the ISA² action.

Work Package WP2 covers the analysis of the existing initiatives on Linked Data and Open Data, and aims to identify opportunities for publishing archival data as Linked Open Data. The following activities were performed:

- Task-2.1 - Discovery of Linked Open Data initiatives and methodologies: The identification of existing initiatives and best practices of archival data being published as Linked Open Data.
- Task-2.2 - Conclusions and recommendations: The provision of recommendations and requirements for turning archival data into Linked Open Data emphasizing also on tools that exist for this purpose.

2.2 OBJECTIVES OF THE FINAL REPORT

The Final Report summarises the key findings from four inventories and reaches a conclusion after the research carried out for the elaboration of the deliverables belonging to WP1 (Inventory of Standards, Inventory of Business Processes, Inventory of IT Tools, Comparative Assessment Tool) and WP2 (Inventory of Linked Open Data and data transformation tools). Finally, it proposes a set of recommendations for the execution of a proof of concept implementing several Use Cases that are particularly interesting to verify the feasibility of key features of a global interoperable archives management solution.

3 APPROACH AND METHODOLOGY

3.1 GENERAL APPROACH

To carry out this study, information was gathered from desk research, alignment meetings and interviews carried out with relevant stakeholders at EU level.

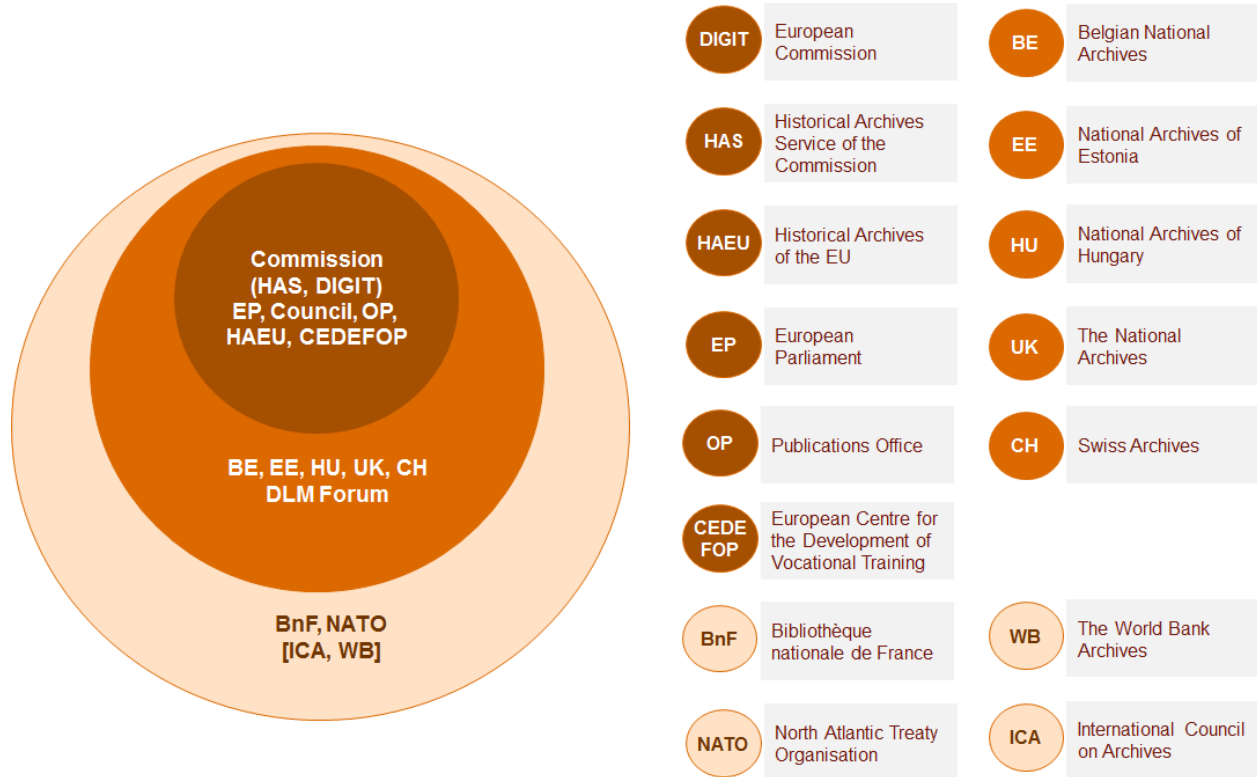


Figure 1: Stakeholders map

The methodology used to obtain general knowledge of current practices and standards consisted of an initial stage of primary research and initial interviews, followed by knowledge gathering via available published sources and contacts with main stakeholders through workshop and meetings. The information obtained from interviews and reports or publications was then analysed and cross-referenced.

The information sources used were national and international agencies like the International Association for Standards (ISO), the International Council on Archives (ICA) and other groups that have developed standards in the domain of records and archives management. The ISO, for instance, has issued a range of standards related to records management, digital preservation and metadata management. Other sources were relevant projects and initiatives in the context of archives management and publication activities.

The most representative information sources for the standards inventory are listed in Figure 2: Desk research below. They are divided into the following categories:

- Websites of archives associations (national; international);
- Websites of national archives;
- Websites of relevant projects;
- Websites of relevant forums;

- Glossaries, publications and other documentation produced by archive institutions;
- Recently published books in the area of archives, metadata and standards.

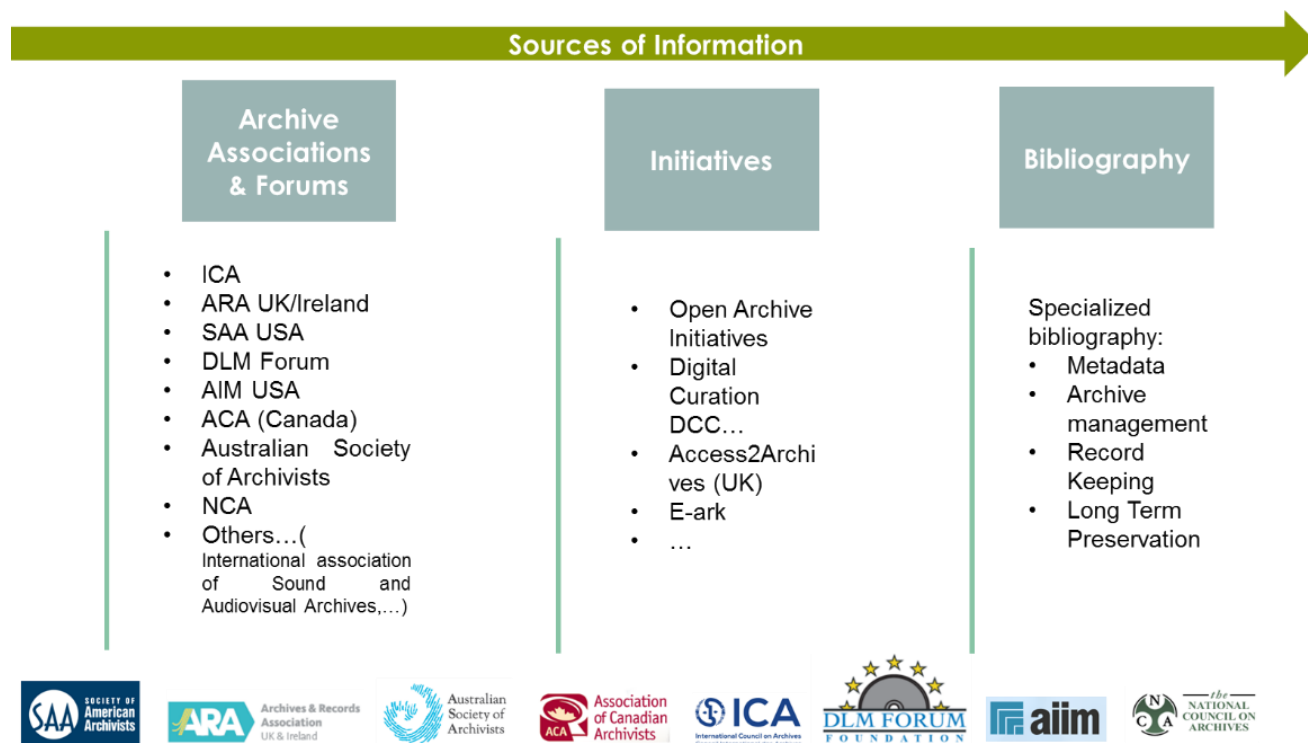


Figure 2: Desk research

3.2 SOURCES

Amongst several sources of information, the outcome of the E-ARK project (E-ARK Project, 2017) about a large set of European archives has been further analysed², and a special focus has been placed on two particular archival institutions. These two institutions are the HAS, which was chosen due to the possibility of accessing the information regarding its business processes, and the UK National Archives, as it is considered as the reference archive in Europe due to its well defined processes.

3.2.1 Archives

Historical Archives of the European Commission (HAS)

At the European level, as in other international contexts, the complexity related to the organisations themselves, the types of documents they produce and the kind of metadata required to describe and preserve documents is very high and can be an important obstacle to archives processing and interoperability between institutions.

In this context, the Historical Archives Service of the European Commission (OIB, 2017) has proven to be an interesting case among archives services of international organisations.

² The following E-ARK deliverables are particularly relevant: 3.1 E-ARK Report on Available Best Practices, 4.1 E-ARK Report on Available Formats and Restrictions, 5.1 E-ARK GAP report between requirements for access and current access solutions.

Compared to its counterparts at international level, the HAS is responsible for the long term preservation of its archives, but, its archives, together with those of the other EU institutions, have to be made available via the Historical Archives of the EU, located in Florence, Italy, This brings in the issue of working in a cross-border context in which a large group of heterogeneous stakeholders have different levels of participation, due to their varying business processes.

UK National Archives

The UK National Archives (UK National Archives, 2017) have been selected because of its important web presence and the level of availability of documentation of their business processes covering both paper and digital archives. They even offer some best-practice manuals for some of the most common archival processes, like the appraisal and selection of documents (Archives, 2018). Another aspect that is interesting for this study is their established practice in the area of open data. One of their flagship projects is without a doubt their open data effort to publish the UK's legislation on legislation.gov.uk. Their collaboration with the UK Data Archive (UK Data Archive, 2017), which is the institution that manages the largest collection of social and economic data in the UK also creates parallelisms in both institutions. This can specially be appreciated in the *Assessment of UKDA and TNA Compliance with OAIS and METS Standards* (JISC, 2017), for example.

3.2.2 Initiatives and projects

Several specific initiatives and projects have been used as a source of information for the identification of the standards.

E-ARK project

The European Archival Records and Knowledge Preservation project, best known as E-ARK, was a research project conducted in the context of archiving by European National Archives in co-operation with commercial systems providers. It developed a pan-European methodology and created seven pilots covering the main steps of archives management for digital archives and combining existing national and international best practices (E-ARK Project, 2017). The project ended in February 2017 and the DLM Forum acquired the responsibility to help to sustain the project outputs, enhancing their longevity (DLM Forum, 2017). Furthermore, in the context of the CEF programme run by the European Commission, a new Building Block on eArchiving has been defined. Dedicated activities on this Building Block will start in 2018 (European Commission, 2018). In addition, a project partner, *The Digital Preservation Coalition* (The Digital Preservation Coalition, 2017) has promoted the best practices in this area.

E-ARK aimed to synchronise the methods and technologies of digital archiving, in order to achieve consistency across Europe. Tackling a range of problems associated with independent record-keeping technologies, systems and practices, E-ARK focused on the development of internationally accessible archives through: the provision of technical specifications and tools; the development of an integrated archiving infrastructure; the demonstration of improved availability, access and use; and the rigorous analysis of aggregated sets of archival data.³

<p><u>E-ARK Project</u> <i>Guidelines on pan-European e-archiving system as part of EC e-infrastructure</i> <i>Open Archival Products (tools, services, framework, metadata specifications)</i> <i>Open Technical Products (tools, services, metadata</i></p>	<p>http://www.eark-project.com/</p>
---	--

³ <http://www.eark-project.com/>

<p>specifications) <i>Open Operational Products (ingest and access tools, services, metadata specifications)</i> <i>Open Access tools, services, metadata specifications, including data mining tools for business intelligence</i></p>	
<p>D3.1 E-ARK Report on Available Best Practices provides an overview of the current situation of the digital archiving best practices.</p>	<p>http://www.eark-project.com/resources/project-deliverables/6-d31-e-ark-report-on-available-best-practices</p>
<p>D2.1 General pilot model and use case definition describes the conceptual framework or E-ARK General Model as well as the process to put it in practice and its use case definitions. The final version of the General Model has been published by the DLM Forum.</p>	<p>http://www.eark-project.com/resources/project-deliverables/5-d21-e-ark-general-pilot-model-and-use-case-definition http://kc.dlmforum.eu/gm</p>

Table 1: E-ARK Desktop Research

ICA P-COM experts group

ICA Expert Groups are groups of experts on any matter of professional interest or concern in the scope of archives management. This Programme Commission (P-COM) experts group drives the ICA's technical and professional programme, including monitoring the Branches, Sections and expert groups. The Programme Commission is responsible for ICA specific programmes, as well as the selection of projects submitted to ICA by members or non-members.

Specific projects of ICA P-COM have been consulted in order to obtain information useful to include in the desktop research process:

<p><u>ICA Expert Groups</u> <i>Groups of Experts on any matter of professional interest or concern in the scope of archives (ICA Expert Groups, 2017)</i></p>	<p>https://www.ica.org/en/our-professional-programme/expert-groups</p>
<p><u>EGAD-Archival Description</u> <i>For the term 2012-2016, the EGAD is charged with developing a formal conceptual model for archival description that identifies and defines the essential components of archival description and their interrelations to promote a shared understanding of archival description, to facilitate the development and use of archival descriptive systems, to enable national, regional, and international collaboration, in the archival community as well as with allied cultural heritage communities. (EGAD, 2017)</i></p>	<p>https://www.ica.org/en/about-egad</p>
<p><u>EGA – Appraisal</u> <i>Drafting guidelines for the selection of digital content for long-term preservation by heritage institutions.</i> <i>The Guidelines are designed primarily for libraries, archives and museums, including professionals and administrators who deal with digital materials and plan to prepare a policy on selection of digital</i></p>	<p>https://www.ica.org/en/expert-group-appraisal-ega</p>

<i>materials for long-term preservation.</i> (EGAD, 2017)	
<u>EG-MDPR - Expert Group on Managing Digital and Physical Records</u> <i>The Expert Group on Managing Digital and Physical Records</i> (EG-MDPR, 2017) <i>was created in early 2017 by the merging of the Expert Group on Records Management (RMEG, 2017) and the Expert Group on Digital Records (DREG, 2017)</i>	https://www.ica.org/en/expert-group-on-managing-digital-and-physical-records-eg-mdpr
<u>Records in Contexts (RiC):</u> a standard for archival description developed by the ICA Experts Group on Archival Description	https://www.ica.org/en/records-in-contexts-ric-a-standard-for-archival-description-presentation-congress-2016 https://www.ica.org/en/egad-ric-conceptual-model

Table 2: ICA Desktop Research

ICSTI

<u>ICSTI</u> <i>The International Council for Scientific and Technical Information is a forum that promotes the interaction between organisations that create, disseminate and use scientific and technical information.</i> (ICSTI, 2017)	http://www.icsti.org/
<u>Report of the ICSTI 2009 Public Conference held 9-10 June 2009 at Library and Archives Canada, Ottawa, Ontario, Canada.</u> (ICSTI, ICSTI Annual Conference - "Managing Data for Science", 2009)	http://www.icsti.org/IMG/doc/ProceedingsOTTAWAV-F.doc

Table 3: ICSTI Desktop Research

Open Archive Initiatives

The Open Archives Initiative (OAI)⁴ develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. OAI has its roots in the open access and institutional repository movements. Continued support of this work remains a cornerstone of the Open Archives programme. Over time, however, the work of OAI has expanded to promote broad access to digital resources for eScholarship, eLearning, and eScience.

DCC Digital Curation Centre (DCC)

The Digital Curation Centre (DCC) is an internationally-recognised centre of expertise in digital curation with a focus on building capability and skills for research data management. The DCC provides expert

⁴ <https://www.openarchives.org/>

advice and practical help to research organisations wanting to store, manage, protect and share digital research data and provides access to a range of resources including guidelines, case studies and online services.

DPC Digital Preservation Coalition

The DPC is a not-for-profit company limited by guarantee and registered in England and Wales. DPC is a membership organisation established in 2002 as a cooperation between a number of agencies operating in the UK and Ireland.

DPC's mission is to enable members to deliver resilient long-term access to digital content and services, helping them to derive enduring value from digital collections and raising awareness of the attendant strategic, cultural and technological challenges they face.

Personal Data Protection

The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) is a European Union (EU) regulation intended to strengthen and unify data protection for all individuals within the EU as well as addressing the export of personal data outside the EU. During the research phase this piece of legislation, as well as related presentations and interpretations were consulted.

Organisations are required to make changes in their policies, processes and contracts, as well as in technical and organisational compliance measures, to demonstrate and sustain compliance.

Technology advances in this field have led to the development of a wide set of IT solutions that are useful and cost-time effective in helping to locate, identify, classify and protect all data – personal and non-personal – implement and maintain data inventories, data maps and to ensure GDPR compliance.

Content analytics

Content analytics works with unstructured data such as text in documents and can be used to improve archive processes that need to analyse big volume of content.

Content analytics uses different big data technologies; ranging from semantic analytics or ontological analysis to discover correlations and patterns in the content, to natural language processing or sentiment analysis to place the data in the right context to multilingual search or linguistic modelling. In addition, text-analytics or text mining will help to extract meaning in the text.

The archives are considering the use of these technologies to apply them in their regular management processes which can bring many changes in the near future.

Master Data Management

Master data management (MDM) is the comprehensive method used to consistently define and manage the reference data of an organisation to provide a single point of reference for data required to operate across several applications or organisations, such as for Customer Relationship Management (CRM), Enterprise Resource Planning (ERP) or Supply Chain Management (SCM). Such reference data may include customer data, product data and other lists of standardised data regularly used by these applications (e.g. list of ISO country codes, postcodes, etc.).

MDM has the objective of providing processes for collecting, aggregating, matching, consolidating, quality-assuring, persisting and distributing such data throughout an organisation to ensure consistency and control in the ongoing maintenance and application use of this information.

MDM solutions are software products that:

- Support the global identification, linking and synchronisation of master data across heterogeneous data sources through semantic reconciliation of master data.
- Create and manage a central, persisted system of record or index of record for master data.

- Enable delivery of a single view of one or more subject areas to all stakeholders, in support of various business initiatives.
- Support ongoing master data stewardship and governance requirements through workflow-based monitoring and corrective-action techniques.
- Are "agnostic" with regard to the business application landscape in which they reside; that is, they do not assume or depend on the presence of any particular business application(s) to function.

Processes commonly seen in master data management include all processes related to the management and control of data created, used and maintained across different applications and support tasks such as source identification, data collection, data transformation, normalisation, rule administration, error detection and correction, data consolidation, data storage, data distribution, data classification, taxonomy services, item master creation, schema mapping, product codification, data enrichment and data governance.

A recent market study by Gartner (Gartner, 2017) reveals as leaders specific vendors out of the scope of the management of non-structured data (text, documents), the main business case for historical archives. At this moment the main focus of MDM is ERP, CRM and SCM or specific business applications that works with structured data (located in relational databases or in other structured locations).

No study or project has been found that makes use of products in archives management or interchange between archive institutions. However, the most mature archives management systems do provide some features for master data management in the archiving context, in particular support for authority lists management.



Figure 3: 2017 Magic Quadrant for Master Data Management Solutions

4 BUSINESS PROCESSES

4.1 OBJECTIVES

The goal of this section is to provide an overview of the business processes carried out by a set of reference institutions in the context of archives management. This overview will serve as the basis to establish a common framework that is used in the next two chapters – *5 Standards and Best Practices* and *6 IT Tools and Solutions* – and their corresponding annexes. The assessment tool provided as a separate deliverable also refers to this list of processes.

Research was carried out targeting archives services within the European Institutions and at national level in several European countries. These institutions are all facing similar challenges in ensuring long-term preservation of and accessibility to their archives, using digital means and this for both non-digital and digital-born archives. This research has led to the identification and description of a set of common functionalities and business processes for managing archives.

4.2 APPROACH

Recent projects in archives management, to be understood as “the general oversight of a program to appraise, acquire, arrange and describe, preserve, authenticate, and provide access to permanently valuable records” (Society of American Archivists, 2017), were evaluated to gather information about the metadata used, the procedures, standards and IT tools used in the field, and the best practices adopted. Furthermore, future projects and developments regarding archives were also examined to gain a better understanding of new trends in this field.

The chapter follows a two-layer approach:

- High-level overview: in the first layer, the aforementioned research has led to the development of a functional model that aims at identifying a set of functional blocks which regroup the most common business processes. These processes can be applied by organisations of various sizes, allowing a high level of flexibility to adopt them.
- Focus on specific processes: in the second layer, the proposed functional model is used as the reference for structuring and describing the archival business processes carried out. Each process is described, and divided into sub-processes if needed. This is mostly the case when the handling of digital archives differs from handling non-digital archives. For each process, a definition, the responsible entity, the input and the output are described.

4.3 STAKEHOLDERS

This section identifies each of the stakeholders that participate in the archival process, and gives a description of their respective roles and responsibilities in the sequence of this process.

Producer

A Producer is any entity that provides the records that start the archival process. In the public sector, a Producer corresponds to any public organisation/service that creates records that must be preserved and remain accessible over the long term. The entity can be found at any level, from local and regional administrations to national or European institutions.

Archival Institution

An Archival Institution is an organisation or organisational unit that manages archival materials as its primary business function. Its main purpose is to acquire, preserve and provide access to collections of archival materials once it has received the records from the Producer.

It may be also in charge of supervising the archival processes carried out by the Producer.

Interoperability Stakeholder

Any other (archival) institution that exchanges archival data or other archival information with the Archival Institution may be considered as an Interoperability Stakeholder (IO Stakeholder). The need for this type of stakeholder emerged from the observation that information exchange takes place between (archival) institutions. Among others, the following examples can be highlighted:

- Cooperation between national and regional archives within a country;
- Publication of archives from multiple sources on the Archives Portal Europe;
- Mandatory interaction between the archives services of the European Institutions with the Historical Archives of the European Union in Florence.

Consumer

A Consumer is any person or institution that accesses the archives managed by an Archival Institution to consult or retrieve information from them.

4.4 FUNCTIONAL MODEL

This section presents a high level functional model that aims to regroup the main business processes that are performed in archives management. Although this model was developed using the OAIS reference model (ISO, 2012) as a basis to group the processes into logical functional blocks, it integrates input from several other sources that were consulted during the study and combines them to produce the model presented hereunder. One of the reasons for this integration is that, although the OAIS reference model describes the processes regarding digital preservation in a precise way, it does not cover the processes related to the more classical archival processes (i.e. management of non-digital records) and processes for interoperability needs. The objective is therefore to ensure that traditional activities as well as data exchange activities performed by archival institutions are also covered by one single functional model, de facto extending the OAIS model with its main focus on the preservation of digital records. The sources used are outlined in greater details in section 3.2 Sources and include:

- The outcome of the E-ARK project,
- The documentation set on business processes published by the UK National Archives,
- The internal research carried out by HAS as well as the detailed documentation of their internal business processes,
- The information from the interviews carried out in the first phase of this study,
- The professional glossaries such as the glossary maintained by the Society of American Archivists (Society of American Archivists, 2018) or the Multilingual Archival Terminology database maintained by the International Council on Archives (International Council on Archives, 2018) with definitions of the business processes carried out by archival institutions.

The proposed model is a modular model making various configurations of business processes for archiving possible. This choice was made because no standardised description of business processes for archives management exists. Only for digital processes the E-ARK project has defined a framework of business processes to establish in the context of archiving, focusing on several key archives management processes (E-ARK Project, 2017).

Given these facts and the complexity of the context of archives, in which digital and physical business processes are carried out simultaneously and with interactions taking place between them, a modular framework was developed, covering most of the activities performed in general by archival institutions...

The proposed model is the following:

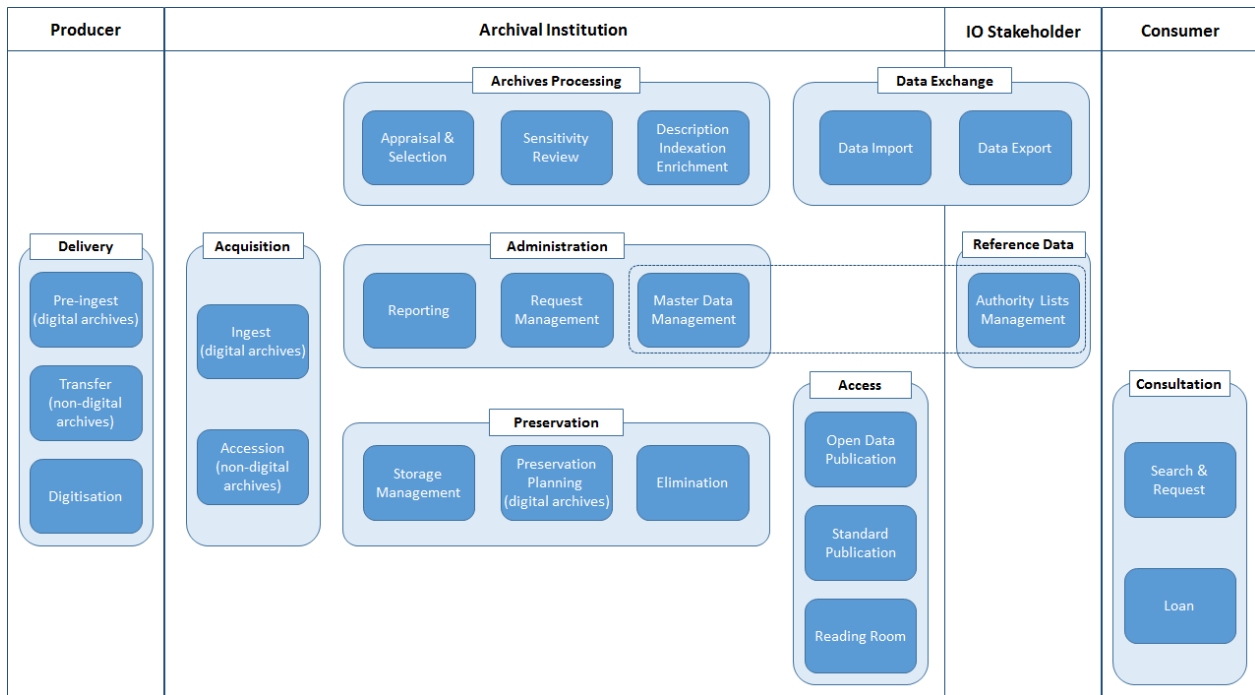


Figure 4: Proposed Model

The proposed model describes the most basic structure considered to cover the main functionalities of an average archival institution. The stakeholders in this model are the Producer of the information, which can be any institution that transfers or delivers archives; the Archival Institution itself, that will be in charge of preserving, managing and providing access/publishing the archives once received; the different Interoperability Stakeholders that will exchange data for various purposes with the Archival Institution; and the Consumer, which is the final user that will extract the information from the repositories maintained by the Archival Institution.

The following table describes the functional blocks and indicates the stakeholder(s) involved. The detailed description of each business process is available in Annex A: Business Processes.

Functional Block	Stakeholder	Description
Delivery	Producer	Activities performed prior to the transfer of the archives to the Archival Institution

Acquisition	Archival Institution	Activities performed upon reception of the archives in order to document the transfer and formalize the shift in control.
Archives Processing	Archival Institution	Core activities performed by the archivists to appraise archives, file, describe, index them, identify sensitive data and make them ready for search and consultation
Preservation	Archival Institution	Activities related to the long term preservation of archives including their disposal when preservation is no longer needed
Administration	Archival Institution / IO Stakeholder	Supporting activities needed to guarantee the smooth and efficient operations of the systems managed by the Archival Institution
Access	Archival Institution	Activities that consist in the preparation of the metadata and archives for consultation by the Consumer
Consultation	Consumer	Activities related to the consumption of the information maintained in the systems managed by the Archival Institution
Data Exchange	Archival Institution / IO Stakeholder	Activities ensuring the exchange of metadata between Archival Institutions in a standardised way
Reference Data	Archival Institution / IO Stakeholder	Activities related to the ease of data exchanges between IO stakeholders.

Table 4: Functional Blocks

5 STANDARDS AND BEST PRACTICES

5.1 OBJECTIVES OF THE INVENTORY

The purpose of this inventory is to collect and document a list of standards currently available⁵. It aims to provide an overview for archivists of the existing standards and the institutions that develop and maintain them, and to serve as a reference tool for archives that need assistance in the selection of relevant standards.

The inventory identifies and categorises commonly used and accepted international and national standards related to archives management activities (such as description, arrangement, appraisal, selection, preservation, publishing, etc.) and to data exchange and it provides a solid baseline of the actual use of standards.

In the context of this study, 'standard' is used in the broad sense of the word: the standards inventory includes standards, such as generic ISO standards, conventions and rules developed by specialised associations and institutions, but also widely accepted best practices, procedures and guidelines for which experience and research has demonstrated that they are an optimal and efficient mean to produce a desired result⁶.

Most of the entries of the current list of standards are the product of project research or the results of the work of an institution which received either broad review within their organisations or by potential users before they were adopted or published.

5.2 MAIN FINDINGS

During information collection for the elaboration of the standards inventory, it became clear that there is a large set of generally adopted national and international standards defined for archival description. However, other traditional archival processes such as appraisal, selection, elimination, transfer, storage or of physical space management and control of access and security are less prone to standardisation.

As far as standards that define traditional archival processes are concerned, no widely adopted standard(s) exists yet at international or European level. Each institution covers this gap through a) the development of its own guidelines and best practices and b) the use of the methodologies and definitions from standards belonging to other areas, like records management, long term preservation, or more recent areas like IT security.

For archival description, there is a need for (further) normalisation and the progressive use of standards developed and implemented traditionally in the scope of libraries and museums for improving archival description and information discovery can be observed. Examples of these standards would be the ones used for bibliographic description and cataloguing; or value standards for authorities or subjects in the form of controlled lists, thesauri or ontologies. The usage of linked data and the semantic web entail the use of these standards and will enrich archival description by making it interoperable, granular and more accessible to users.

Changes with regards to data security, including the new General Data Protection Regulation add another set of standards to be taken into account in the management of archives: the data protection standards. These standards are not new: they come from other domains such as informatics and cyber-security, but archival institutions are becoming more and more aware that they can be applied to traditional archival processes such as access control and content security.

⁵ The comprehensive inventory may be consulted in Annex B.

⁶ Best practices are not necessarily formal standards, but they may be considered ad hoc or de facto standards.
<https://www2.archivists.org/glossary/terms/b/best-practices>

Finally, another interesting area is the rise of new techniques for massive data and content management, such as content analysis and automatic classification, based on methodologies from Artificial Intelligence (AI) and statistics. These could help to simplify archives management processes and maybe the development of new functionalities in the near future. There are no existing standards in this field yet but there are methodologies under development that show big potential for archives.

5.2.1 Overview of standards and gaps

The standards inventory has been divided in 8 broad categories based mainly on the categorization that is used in specialized information sources in the area of archives, metadata exchange and technology applied to archives, libraries and museums.

The overview includes standards for the list of business processes identified in the previous section such as provisioning, acquisition, archives description, archives management, metadata management data exchange, access, consultation and preservation.

In addition, the list of standards covers other related areas such as security applied to archives, physical space management and quality management.

- The *Archives description* category covers descriptive standards mainly used in arrangement and description processes, but also in other processes that produce records with descriptive content related to appraisal and selection activities, such as information about transfers, acquisitions, elimination, etc. There are data content standards, data structure standards and data value standards, and their use brings benefits for data exchange and data integrity:
 - *Data content standards* are a set of formal rules that specify the content, order, and syntax of information to promote consistency. All of them are national and international cataloguing rules in the scope of archives.
 - *Data structure standards* are formal guidelines specifying the elements into which information is to be organised. Finding aids, inventories or content management systems use them to provide access.
 - *Data value standards* are established lists of normalised terms used as data elements to ensure consistency. They provide lists or tables of terms, names, alphanumeric codes, or other specific entities whose use might be mandatory for specific data elements. Their use greatly facilitates the exchange of data between archives.
- The *Archives Interoperability* category covers standards used in the business processes of access, publication and data exchange activities. It comprises standards of the Semantic Web for Linked Data, Linked Open Data and for interoperability between repositories as providers or aggregators of data.
- The *Records management* category covers standards related to the lifecycle of records: creation or capture, use and management, disposal: transfer or elimination. It includes a set of frameworks and can be used in the standardisation of traditional archival processes.
- The *Preservation* category covers standards for business processes associated with certain functions of the OAIS ISO standard: pre-ingest, ingest, archives management and dissemination of content. Specific guidelines and initiatives about technical formats and technical preservation requirements are also covered in conjunction with reference manuals and best practices developed by institutions specialised in data curation and preservation activities.
- The *Security* category lists the standards applicable to a range of security techniques. In order to provide standards for the business processes of sensitivity review, several techniques for protecting personal data are provided by this group of standards.

- The *Storage Information* category regroups the standards in use for managing content, indexation and storage in related domains such as libraries and museums.
- The *Metrics and key performance indicators (KPI's)* category lists the set of standards to use in the monitoring of systems being part of the archiving environment.
- Finally, the section about *Practices in other areas* provides information and best practices for activities related to automatic data classification and file/content analysis using machine learning techniques that can be useful to automate some archiving activities.

The following table summarises the mapping between each group of standards (as defined in Annex B: Standards and Best Practices) with the business processes identified in Chapter 4 and described in detail in Annex A: Business Processes.

The table shows that for most of the business processes standards exist, except for the specific business processes of Access and Consultation of non-digital archives.

Standards / Business Processes	Delivery			Acquisition		Archives Processing			Administration			Reference Data	Preservation		Data Exchange		Access			Consultation		
	Pre-ingest (digital archives)	Transfer (non-digital archives)	Digitisation	Ingest (digital archives)	Accession (non-digital archives)	Appraisal & Selection	Sensitivity Review	Description / Indexation / Enrichment	Reporting	Request management	Master Data Management	Authority Lists Management	Storage Management	Preservation planning (digital archives)	Elimination	Data Import	Data Export	Open Data Publication	Standard Publication	Reading Room	Search & Request	Loan
Archival description																						
Data content standards	✓	✓		✓	✓	✓		✓	✓				✓			✓	✓	✓	✓	✓	✓	✓
Data structure standards	✓	✓		✓	✓	✓		✓				✓				✓	✓	✓	✓	✓	✓	✓
Data value standards	✓	✓		✓	✓	✓		✓				✓				✓	✓	✓	✓	✓	✓	✓
Archives Interoperability																						
Semantic Web Standards																		✓				
Standards for Repository Interoperability																✓	✓		✓		✓	
Preservation																						
Open Archival System Model - OAIS (ISO 14721)	✓			✓		✓			✓	✓			✓			✓	✓		✓		✓	
ISO 20652:2006 (CCSDS 651.0-B-1:2004) Space data and information transfer systems -- Producer-archive interface -- Methodology abstract standard	✓			✓		✓										✓	✓					
Data Dictionary for Preservation Metadata (PREMIS)	✓			✓		✓		✓			✓	✓										
Metadata Encoding & Transmission Standard (METS)	✓			✓		✓		✓			✓					✓	✓					
Metadata Object Description Schema (MODS)	✓			✓		✓		✓			✓											
LoC Format Guidelines - Digital Preservation at the Library of Congress	✓		✓	✓																		
FADGI Federal Agencies Digital Guidelines Initiative	✓		✓	✓		✓																
Curation Reference Manual	✓			✓		✓					✓		✓	✓								
DPC Digital Preservation Handbook	✓		✓	✓	✓	✓					✓			✓	✓				✓		✓	
E-ARK Information Packages SIP, AIP, DIP	✓			✓		✓										✓	✓		✓		✓	
Records Management	✓	✓			✓										✓							
Security								✓							✓							
Storage Information										✓			✓									
Metrics and KPI's									✓													
GAP / NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP

Table 5: Mapping between standards and business processes

5.2.2 Cartography of standards

In order to provide an overview of how standards are used by the archives community, the inventory was split in two major areas:

- **(Meta)Data management**, regrouping all standards for archives management, including content description and metadata;

- **Exchange**, regrouping all standards used in the exchange of (meta)data between systems.

Within each area, three categories were defined depending on their relevancy in the context of archives management:

- **Focused on archives management**, regrouping the standards developed specifically for archives management and covering the key business processes performed by archival institutions such as archival description;
- **Adopted by archives management**, regrouping standards developed in areas closely related to archives management, such as record management or preservation, and used by the archives community;
- **Somewhat related to archives management**, regrouping additional standards created within other fields of knowledge without a clear intention to support archival processes, such as content analytics, and with potential use for the archives community.

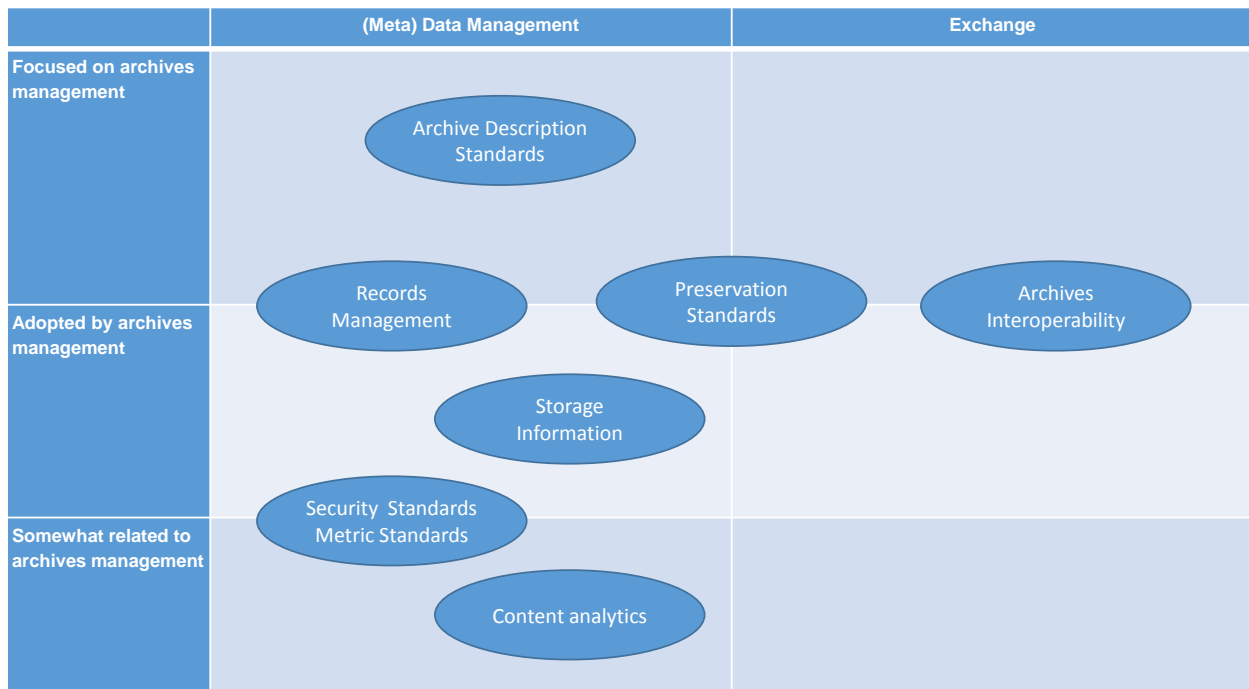


Table 6: Cartography of standards

6 IT TOOLS AND SOLUTIONS

6.1 APPROACH AND METHODOLOGY

The IT tools inventory follows a top-down approach and divides them in several categories based on their generic characteristics. It also provides additional information about mapping to and coverage of the business processes identified in Chapter 4 and the standards listed in Chapter 5, licensing schemes, availability of support services, size of the user community and main IT characteristics, such as infrastructure requirement, dependencies with other tools and detailed architecture specifications.

During the analysis, identifying and assessing possible combination of and integration between IT tools in order to satisfy the requirements of current and new standards and business processes got special attention.

The methodology that was followed consisted of an initial stage of desk research, followed by the collection of more detailed information from publicly available sources, market studies and reports. This approach can be translated into the following steps:

- Step 1: General desk research to identify IT tools in the different categories.
- Step 2: In-depth research including:
 - Market analysis studies of products: recent market studies from Gartner, Ovum and other consultancy companies specialised in the assessment of IT products, as well as research studies from archival institutions or organisations;
 - Vendor official Web-sites: specific documentation and published information about vendor surveys, vendor white papers and vendor product/component technical specifications;
 - Websites of national and international archives associations and institutions: information and officially published documentation about archives management systems and preservation;
 - Recent projects and initiatives in the domain of archives management.
- Step 3: Elaboration of the inventory: the information extracted from the different sources is reformatted in the form of tables/fiches, in order to standardise the information...
- Step 4: Conclusion of the research: At this point of the study, the most important gaps were identified and documented.

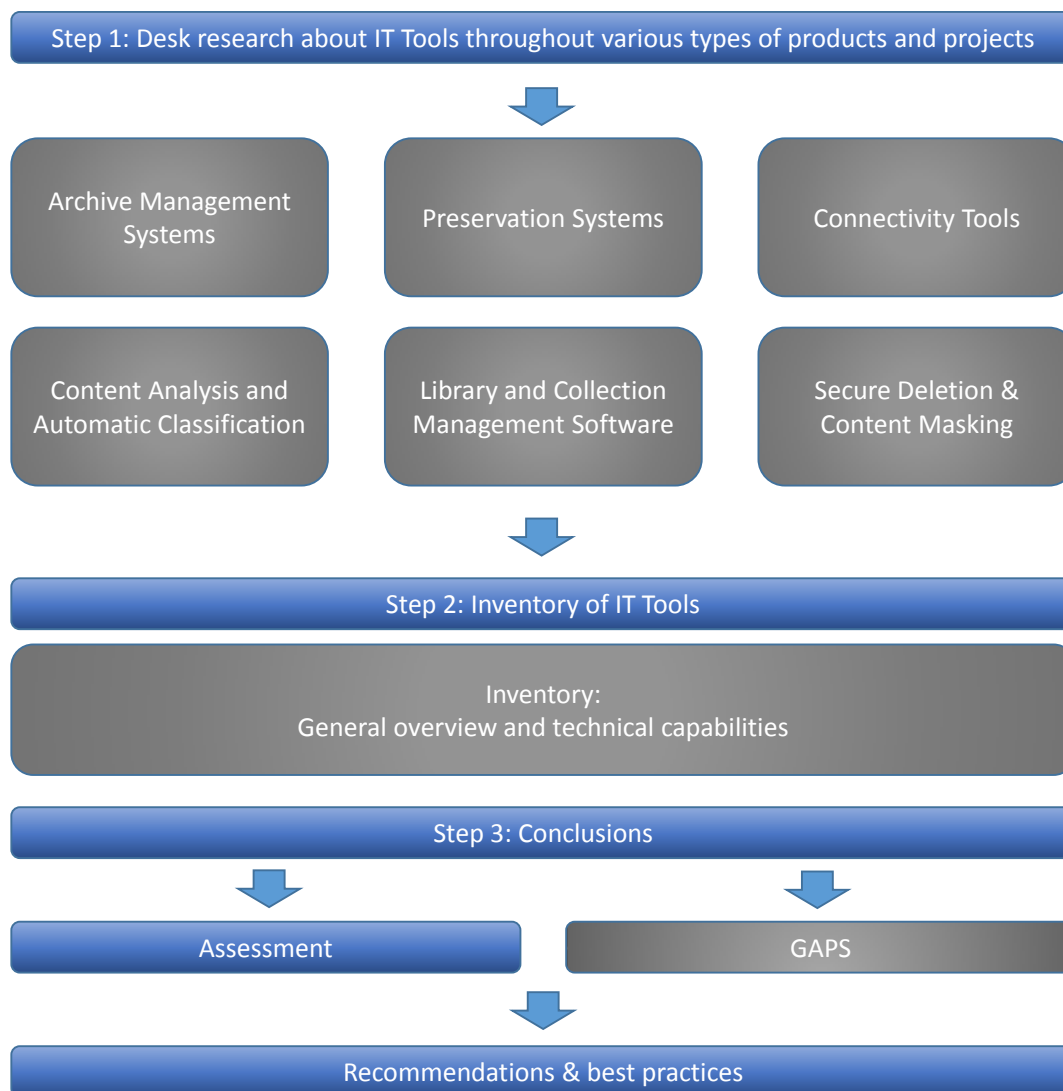


Figure 5: Methodology of the study⁷

It should be noted that the IT tools inventory details the software that supports the archives management business processes identified during the study. As such, although some of them are listed, IT tools supporting the preservation aspects are not at the core of this inventory.

6.2 OVERVIEW OF IT TOOLS

The inventory comprises commercial and open-source software from different domains, uses and typologies, including tools developed for specific projects, which could be re-used, customised or improved by the archives community.

Most of the tools with a clear archival focus started as specific initiatives or projects, created in the scope of a public project involving public archival institutions of different countries and evolved towards open-source products maintained by an established community of developers and users.

⁷ Recommendations & best practices are covered in D05 – Conclusions & Recommendations

The overview of IT tools covers the following domains:

- Archives management systems⁸: software (commercial or open-source). It also includes software applications falling into the category of:
 - archival description and integration with standardised controlled vocabulary lists;
 - products for creating and publishing encoded archival description (EAD) and including support for linked open data (LOD);
 - finding aids supporting the archival business processes, which can be integrated as a component;
 - collection management;
 - digital asset management;
 - records management integration;
 - preservation integration.
- Preservation systems: software (commercial or open-source) that provides preservation capabilities. Based on ISO-OAIS Reference Model, this type of software supports the technical aspects of a digital object's life cycle: ingest, archival storage, data management, administration, access and preservation planning. Although preservation tools are out of scope of the objectives of the study, they are listed in this inventory since their interaction with archives management systems is a key criterion for selecting the right system. Interaction between these two types of tools can either be through direct interfaces or via a connectivity tool (listed below).
- Connectivity tools: IT tools developed for the integration between archives management systems and preservation systems, usually in the form of open source or commercial software components (add-ons⁹) that provide additional functionalities and can cover one or more of the following categories:
 - Implementation of technical connection standards / protocols (like CMIS);
 - Metadata mapping (schema A to schema B);
 - Connectivity with external systems to acquire the records to be archived – e.g. databases, file-systems, raw systems, email systems, content management applications or records management systems, etc..
 - Connectivity with bespoke systems/portals to publish data (which often does not need to involve the preservation systems but only archives management systems – such as sending metadata to Archives Portal Europe);
- Library and collection management software: General IT tools that provide a wide range of capabilities to store, manage and publish content, but not necessarily developed in the context of archives management. They usually offer specialised search and retrieve capabilities for different content formats (video, images, sound, etc.) and incorporate Linked Data and Linked Open Data features.
- Content analysis and automatic classification: file analysis software to access, understand, classify and control all information.

⁸ A program to appraise, acquire, arrange and describe, preserve, authenticate, and provide access to permanently valuable records.

⁹ A piece of software which enhances another software application and usually cannot be run independently

- Secure deletion and content masking:
 - Secure deletion: software that manages the elimination of documents in a secure way, includes such features as managing permanent data deletion, tracking the elimination process and creating evidences of deleted information.
 - Content masking: software that provides capabilities for redaction, anonymisation or pseudo anonymisation of text in documents.

6.3 IT TOOLS INVENTORY TEMPLATE

The inventory template provides a brief overview of each IT tool, covering their key attributes and general characteristics, identifying their typology and describing the use of the tool in the archives management process and other relevant features.

Category	Evaluation criteria ¹⁰	Evaluation criteria values
Basic description	Name	<i>Official name</i>
	URL	<i>Website</i>
	Basic description	<i>General description of the tool or project</i>
	Owner	<i>Organisation or company and brief description</i>
	Category	<i>Archives management systems; preservation systems; connectivity tools; library and collection management software; content analysis and automatic classification; secure deletion and content masking</i>
	License and delivery model	<i>Open source or commercial software On-premises or Cloud</i>
Functional & Technical capabilities	Business Process	<i>List of business processes identified in this study that are covered by tool</i>
	Standards	<i>List of standards identified in this study that are supported by the tool</i>
	Authority Control	<i>Use of authority control lists</i>
	Controlled vocabulary	<i>Use of controlled vocabulary lists</i>
	Access management	<i>Management of security access and users/ roles</i>
	Reporting and statistics	<i>Management of information related to actions performed and reports</i>
	Functional architecture	<i>High-level functional components</i>
	Infrastructure platform	<i>System requirements such as technical architecture components (database, internal applications, ...)</i>

¹⁰ Some key points are provided by the study: https://www.clir.org/pubs/reports/spiro/spiro_Jan13.pdf

	Migration	<i>If the solution provides tools that can support the migration of data with functionalities that allow exporting and/or importing data in specific formats.</i>
	Integration	<i>Information regarding integration with third-party products or add-ons.</i> <ul style="list-style-type: none"> • <i>Integration with preservation systems</i> • <i>Integration with records management systems</i> • <i>Integration with publishing systems</i>
	User support	<i>Type of user support (support, training, documentation)</i>
Relevant aspects	Remarks	<i>Some important remarks about archival data management capabilities</i>
	Strengths	<i>List of unique features that allows the comparison with other products or tools in the same category/typology</i>
	Weaknesses	<i>List of features not covered compared to others products or tools in the same category/typology</i>

Table 7: IT Tools Template

6.4 ASSESSMENT

Once the business processes were defined and the IT tools selected, to what extent these tools cover the business processes was analysed. The main groups of IT-tools identified were mapped to the different functional blocks and processes. The result obtained is shown in the following table. With the exception of the Delivery process¹¹, which is under the responsibility of the producer, it shows that no gaps were detected and that it is possible to select a set of tools to serve the functionalities of the entire functional model. However, since there is no single tool that offers all the required functionalities, it will be necessary to combine several tools. Furthermore, custom development is likely to be needed, especially for the Data Exchange block.

Standards / Business Processes	Delivery			Acquisition		Archives Processing			Administration				Preservation		Data Exchange		Access		Consultation			
	Pre-ingest (digital archives)	Transfer (non-digital archives)	Digitisation	Ingest (digital archives)	Accession (non-digital archives)	Appraisal & Selection	Sensitivity Review	Description / Indexation / Enrichment	Reporting	Request management	Master Data Management	Authority Lists Management	Storage Management	Preservation planning (digital archives)	Elimination	Data Import	Data Export	Open Data Publication	Standard Publication	Reading Room	Search & Request	Loan
Archive Management Systems	✓*	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓*		✓	✓	✓	✓	✓
Preservation Systems	✓*			✓		✓		✓	✓					✓	✓	✓*	✓*	✓	✓		✓	
Connectivity Tools	✓*			✓				✓			✓					✓*	✓*	✓	✓		✓	
Library and Collection Management								✓	✓		✓	✓	✓		✓	✓*		✓	✓		✓	
Content Analysis & Automatic Classification	✓*			✓		✓		✓		✓			✓									
Content Masking							✓															
Secure Deletion															✓							
GAP / NO GAP	(NO GAP)	(NO GAP)	GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP	NO GAP

✓* Custom development might be needed

Table: Summary of IT Tools and features

¹¹ It should be noted that the inventory of IT tools does not cover the digitisation tools. This explains why it is highlighted as a gap in the matrix.

6.4.1 Conclusions of the gap analysis

The study comes to the following conclusions on the key topics:

- With regards to archives management systems:
 - Products in the category of archives management tools are the ones that cover traditional archives activities better.
 - Most products offer a platform of interconnected components in order to offer solutions covering the full scope of 'traditional' archival processes (appraisal, accession & transfer, description, manage locations, reference services and publishing with finding aid tools).
 - Most of the main archival description metadata are covered and transformation from content standards (ISAD(G), DACS, RAD) to structural standards (EAD, MODS, MARC21, DC) are offered.
 - Some products are more oriented towards serving as collection management tools and offer additional capabilities for digital asset management (images, videos and traditional documents).
 - For this category of products, the major gaps identified are those related to LOD, integration with controlled lists and implementation of EAD-CPF or other standards used to support the definition of relationship and context information of archival items.
 - Some products are integrated with both records management systems and preservation systems, thereby giving support to the full lifecycle management.
- With regards to preservation systems and connectivity tools
 - The preservation tools implement a large part of the identified processes that serve the born-digital or digitised archives. Tools that cover the complete workflow, from pre-ingest to publication, have been identified. However, many of these have no incorporated LOD in the publication process and other IT-tools are used for it. For some preservation tools, new features such as online analytical processing (OLAP) or geographic information systems (GIS) are implemented to address specific users' needs.
 - Most of the Preservation Tools provide connectors to share information between both systems: the traditional archives management system tools and the preservation system tool. The goal is to share data in order to facilitate tasks and to avoid duplication of information and metadata in multiple systems.
- With regards to library and collection management systems:
 - These IT tools offer new functions in line with linked and open data needs and other new forms of access and interaction: integration with GIS or with databases and OLAP.

A list of new tools (coming from the GDPR environment and Big Data technologies) in the domain of automatic classification, content analysis and security, such as content-masking tools and secure deletion, are included in the inventory. Currently, connections between archives management systems, preservation systems and the above mentioned tools have not been identified.

These last tools can be used to facilitate work with large volumes of documents, for example, to facilitate automatic indexing and classification. They could also be useful in more traditional processes like appraisal, archival description or publication. It is a field that is inciting but there are not many examples yet of practical use in the archives management domain.

6.4.2 Cartography of IT Tools

The inventory of IT tools was split in two major areas:

- **(Meta)Data management**, regrouping all tools for archives management, including content description and metadata.
- **Exchange**, regrouping all tools used in the exchange of (meta)data between systems.

Within each area, three categories were defined depending on their relevancy in the context of archives management:

- **Focused on archives management**, regrouping the tools developed specifically for archives management and covering the key business processes performed by archival institutions such as archival description;
- **Adopted by archives management**, regrouping tools developed in areas closely related to archives management, such as record management or preservation;
- **Somewhat related to archives management**, regrouping additional tools created within other fields of knowledge without a clear intention to support archival processes, such as content analytics.

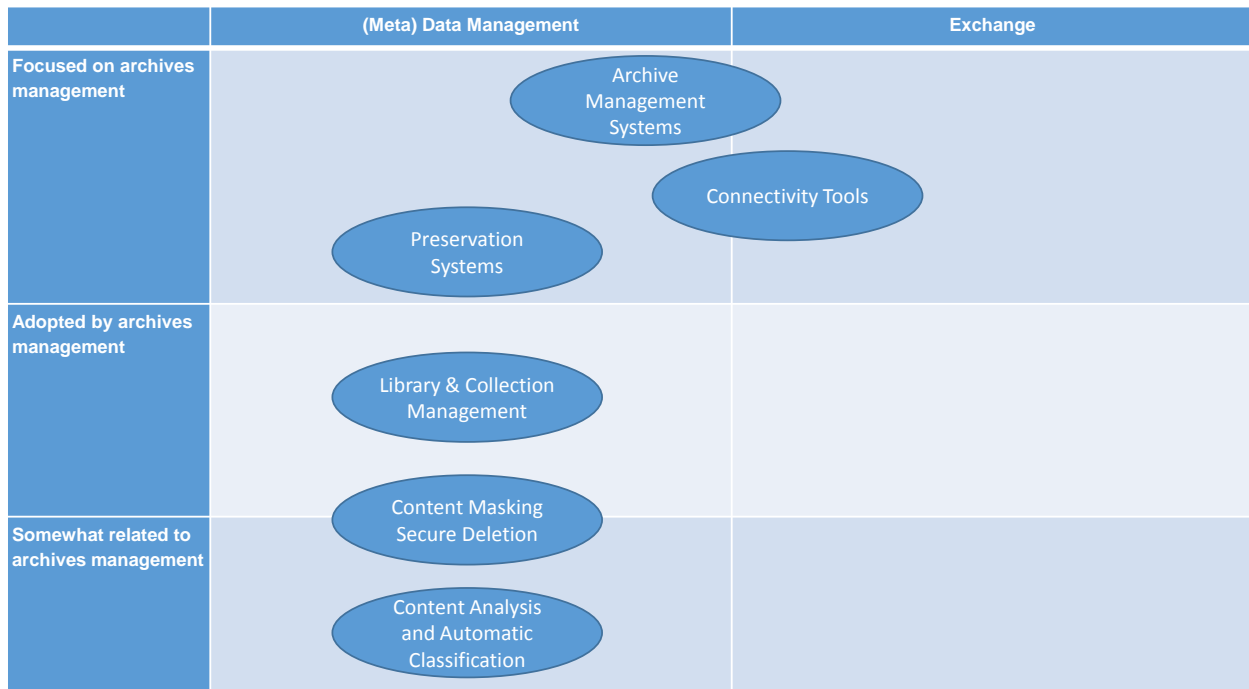


Figure 6: Cartography of IT tools

7 LINKED AND OPEN DATA INITIATIVES

7.1 INTRODUCTION

Linked Open Data (LOD) collates two different concepts: Linked Data and Open Data. The first term refers to a set of principles to publish structured data on the Web, the latter to the practice of publishing data on the Web which can be freely re-used. Combined together, LOD can be defined, in the words of Tim Berners-Lee, as “Linked Data which is released under an open license, which does not impede its reuse for free” (Wikipedia, 2018). The European Commission, as well as the other institutions of the European Union, offers a wide variety of archival information online. In parallel, the Historical Archives of the European Union (HAEU) located in Florence, aggregate information from these institutions and offer a single, authoritative, multilingual online platform to explore the historical archives of the European institutions¹². Over the last years, subsets of these data have been made available through portals such as Archives Portal Europe¹³. With all these various online publication channels already in place, *what is the relevance of LOD?* The answer is relatively simple and straightforward: interoperability. For example, in addition to provision of finding aids in the common EAD format on the Archives Portal Europe, Linked Data facilitate the re-use of common vocabularies for places, periods and names. By doing so, new research questions can be developed on data which originates from various sources.

Other parts of this study address the standardisation work which the archival community has undertaken over the last decades to make archival finding aids interoperable. Despite the consensus on the body of standards to use, the reality on the terrain forces us to acknowledge the heterogeneity of practices. The workshop organised on the 24th of November 2017 at the EC archives, which brought together archival and IT experts from various EU institutions, helped to understand the variety of practices and views on both the structure and the content of archival finding aids. *How can LOD help?* This information architecture paradigm holds the promise to create meaningful links between documents, files, and archival holders across heterogeneous collections. This report will give an overview of existing initiatives and best practices of archival data being published as LOD. Also, requirements and recommendations for turning archival data into LOD and an overview of tools that exist for this purpose are provided.

However, as desk research, visits and interviews conducted in the context of this study demonstrate, the actual implementation of LOD tends to be challenging. The ambition of this deliverable is to give a solid overview of existing initiatives from the archival community, but also to bring a sense of pragmatism to the debate. This study will point out the low hanging fruits currently available, but also identify potential issues and areas where it is uncertain that investments in LOD will deliver short-term benefits for the archives of the EC and its stakeholders. As a profession and discipline, the archival community has been working hard over the last decades to streamline descriptive practices. The rise of the Web in the 1990s obliged archivists to pick up pace with standardisation efforts of metadata schemes and controlled vocabularies, which were initiated since the introduction of databases for creating finding aids throughout the 1980s and 1990s. At the same time, budget cuts and fast-growing collections are obliging archivists to explore automated methods to provide access to resources. Archives are expected to gain more value out of the metadata patrimony they have been building up for decades. The expectations surrounding archival LOD herald this approach as the way to go. However, to what extent can archivists, as a discipline and a profession, take Linked Data at face value? Technology is a means and not an end. Opportunities rise from new technologies, but also behold risks. Linked Data principles are often misunderstood and need to be implemented in a well-reflected manner. Linked Data present tremendous challenges in regards to the

¹² See <http://eui.eu>

¹³ See <http://www.archivesportaleurope.net>

quality of our metadata, so it is fundamental to develop a critical view and differentiate between what is feasible and what is not.

After a substantial introduction to the general principles of LOD, the study will give an overview of how the archival community has been experimenting with various LOD principles and what the open-standing questions are. In a second part of the study, a more structured inventory is presented of various LOD initiatives, projects and tools.

7.2 CONTEXT AND METHODOLOGY

7.2.1 Introduction to Linked and Open Data

In a now famous 2009 TED talk entitled “The next Web”, Tim Berners-Lee introduced and evangelised the concept of Linked and Open Data (Berners-Lee, The next web, 2009). Before introducing LOD, it is important to explain the context in which it was introduced. Almost a decade earlier, Berners-Lee published in *Scientific American* an article on how the notion of the Semantic Web, “*a common framework that allows data to be shared and reused across application, enterprise, and community boundaries*” would revolutionise the Web functions (Berners-Lee, Hendler, & Lassila, 2001). Despite major academic and industrial efforts from 2000 onwards, the Semantic Web failed to become a reality, due to its inherently ambitious scope of formalising the meaning of all sorts of data to be published on the Web by making use of ontologies. Varying definitions of the term exist, but we can refer to Wikipedia, which defines ontologies as “*formal naming and definition of the types, properties, and interrelationships of the entities that exist in a particular domain of discourse*” (Wikipedia, 2018). Even if ontologies demonstrated their use within very specific domains with vast resources, such as the medical sector, standardising and formalising the semantics of entities across all sorts of application domains at the scale of the Web has been a failure. In his 2009 TED talk, Berners-Lee decided to rebrand the idea of the Semantic Web by lowering the conceptual and operational barriers of publishing semantically meaningful data on the Web. For this purpose, the 5-stars principles of the Linked and Open Data paradigm were introduced (Berners-Lee, Bizer, & Heath, 2009), allowing a gradual approach to publish data on the Web, going from the lowest barrier of publishing data under an open license to the more complex task of formalising semantics. These five characteristics are the following:

1. **Available on the Web** (in any format) under an open license
2. **Structured**: make it available as structured data
3. **Open**: make it available in a non-proprietary open format
4. **Shared**: use URIs to denote things, so that people can point at your data
5. **Contextualised**: link your data to other data to provide context

By reading the succession of these five principles, one can understand both the difference and the interaction between Linked Data and Open Data, which are often grouped together under the umbrella concept of LOD. Conceptually, the act of making data available in an open way is very different to providing links in between datasets. However, on the terrain both practices are intertwined and are presented together, which is also the case within this study.

Various international organisations and national governments have been making open data available online for more than a decade, which often encompasses data sets which do not correspond to the 4 or 5 star quality principle of Linked Data. Within this section, we will give an overview of the most successful projects and initiatives and see how the archival sector has positioned itself within this context. The potential value of publishing both structured, semi- and non-structured documents from archival institutions, regardless whether they contain Linked Data, which is addressed more in detail further on in the report.

The official open data portals from the American (Data.gov) and the UK (Data.gov.uk) government have been leading the way for other national governments and administrations such as the World Bank (<https://data.worldbank.org/>) or even cities (<https://opendata.cityofnewyork.us/>). The Publication Office of the European Commission has played an important role within Europe to catalyse open data through its own portal <https://data.europa.eu/>. These initiatives focused initially on providing simply data dumps of various statistical data sources, but have evolved over the last years to more complex applications with important additional features such as data cleaning tools and dashboards to navigate and browse through very big quantities of data. Data.gov does give an interesting overview of actual applications which have been created by making use of various open data, as for example apps to estimate residential energy use and plan home energy efficiency upgrades based on consumption efficiency statistics. Various initiatives have been launched to provide tutorials and tools to process open data in a better manner, such as the <http://opendatahandbook.org>.

Archives in particular have also initiated various open data projects. One of the flagship projects is without a doubt the open data effort of The National Archives to publish the UK's legislation on legislation.gov.uk. The application gives access to legislation as it is enacted by parliament, the changes it makes to existing legislation but also the historical documents, which have been digitised and are freely available for download. The Archives de France have launched https://francearchives.fr/fr/open_data, which is a part of the larger open data policy of the ministry of culture of France. On the portal <https://data.culturecommunication.gouv.fr/> an in-depth report can be found which details the various types of re-use.

After the initial enthusiasm in the early years of the open data movement, the actual uptake and promotion of the usage of the available open data remained relatively limited. The biggest hurdle for the actual usage of open data is their quality and documentation, which is also a big pain point for the uptake and re-use of open archival data, as will be demonstrated. All too often, big data dumps are provided with insufficient or incomplete metadata, which severely hinders their usage. To illustrate the extent of the problem, we can refer to the recent initiative of the publication as open data of DORIE, which stands for Documentation et Recherche sur les questions Institutionnelles Européennes, on <http://data.europa.eu/euodp/en/data/dataset/sg-dorie>. This database provides access to a collection of documents put together over the years by the unit in the Secretariat-General responsible for institutional issues. It is not an exhaustive collection, but it contains a selection of documents or excerpts of documents relating to general institutional issues and to the various intergovernmental conferences that have made changes to the founding treaties. However, when downloading and opening the documents, one sees various data inconsistency issues. By using the data cleaning tool OpenRefine, one can quickly understand to what extent the encoded values are consistent with their definition. Figure 7 illustrates for example the presence of inadequate entries for fields such as theme and document type, which contain to a large extent codes such as "nl". This value might make sense to a large number of end-users, who can interpret "nl" as the language code for Dutch. However, there is one row where descriptions have "shifted" and "nl" is instead in the "Category" field. This is typical example of how the uptake of open data is hindered by data quality issues and lack of consistent documentation. Open Data is too often perceived as a magical wand that improves quality. As this example demonstrates, data owners first need to ensure a reasonable level of quality and then it's justified to also think about the next steps. Now that we have given some examples of open data initiatives, we can see in more details what exactly Linked Data entail.

420 rows

Show as: rows records Show: 5 10 25 50 rows

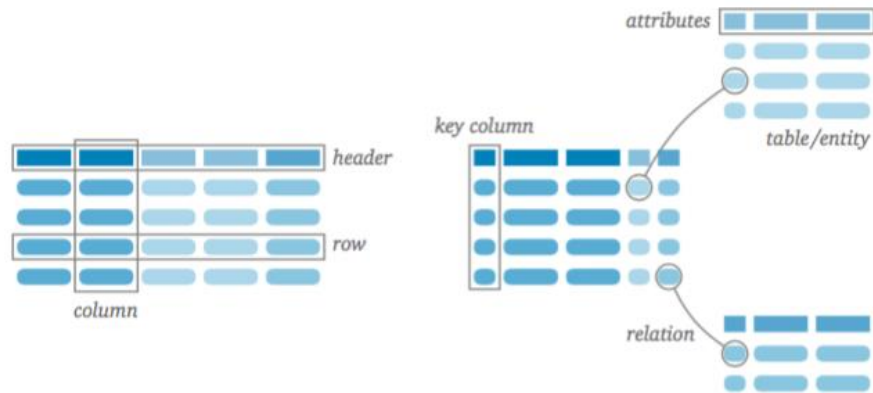
TITLE	LANGUAGE	CATEGORY_AR	CATEGORY_THI	CATEGORY_DO	AUTHOR	LINK
Compte rendu in extenso des séances (Excerpt)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=410615&locale=nl
VERSLAG AAN DE EUROPESE RAAD (Full text) - Serial: SN/1197/95(SPAAKI)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=411391&locale=nl
EERSTE BIJDRAGEN VAN DE COMMISSIE TOT DE INTERGOVERNEMENTELE CONFERENTIE "POLITIEKE UNIE" (Full text) - Serial: SEC(91)500	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=410417&locale=nl
Bericht (Dok. A3-209/92) von Herrn von der Wing im Namen des Nichtständigen Ausschusses „Von der Einheitlichen Europäischen Akte zu der Zeit nach Maastricht“ zu der Mitteilung der Kommission ¹ Von der Einheitlichen Europäischen Akte zu der Zeit nach Maastricht: Ausreichende Mittel für unsere ehrgeizigen Ziele" (KOM(92) 2000 - C3-61/92) - Erklärungen zur Abstimmung (Excerpt)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=410445&locale=nl
Programme d'activité de la Commission (Excerpt)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=410471&locale=nl
Van Rompuy begrijpt kritiek op zijn Europese Raad niet (Full text)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=930938&locale=nl
President zonder draaiboek (Full text)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=906778&locale=nl
EUROPA IS VERSTANDIG	EUROPA IS TAAI (Full text)	nl				
Mister Bling dépoussière les étoiles (Full text)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=476023&locale=nl
Second Chamber: go for European coalitions (Full text)	nl					http://ec.europa.eu/dorie/cardPrint.do?cardId=476415&locale=nl

Figure 7: Overview of some data quality issues within the data set SG DORIE data made available

To grasp the potential but also the limits of Linked Data, one needs to understand the four basic data models which have been used over the last decades to manage structured data. The advantages of RDF, the data model underlying the Linked Data vision, can only be fully understood in the context of previous data models. Also, the different data models presented do not supplant one another, but continue to coexist. The overview of the different models should make it clear that relational databases are here to stay, and will not be disposed of in favour of triple stores, for example. Technology vendors and IT experts have a tendency to overemphasise the role a new technology has to play. At the height of the popularity of XML, one sometimes got the impression that the back-end of any type of information system would become XML-based. A decade later, XML is often criticised, and new serialisation formats such as JSON are often preferred.

In order to get a high-level understanding of the four data models, Figure 8 provides a visual overview of how the different models can be differentiated. One can almost think of the four models as four different sets of glasses. Each pair of glasses presents a different worldview:

1. **Tabular data:** the world is managed in columns and rows, like one long list or excel spreadsheet
2. **Relational model:** the world is managed by identifying entities (families of objects characterised by attributes) which are related to one another, like in relational databases
3. **Meta-mark-up languages:** the world is managed with the help of a hierarchical tree, which has one root and parents and children, like in XML
4. **RDF:** the world is managed through a graph, in which each node can be connected to another one

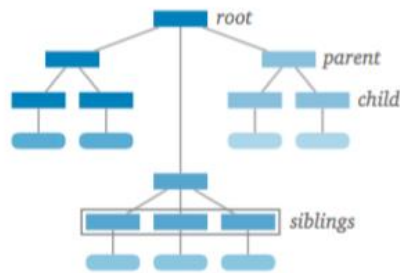


Tabular data

Each data item is structured as a line of field values. Fields are the same for all items; a header line can indicate their name.

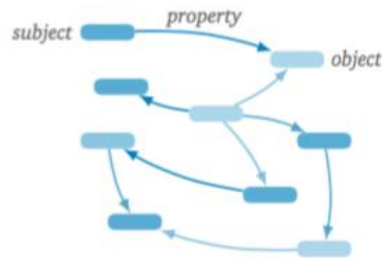
Relational model

Data are structured as tables, each of which has its own set of attributes. Records in one table can relate to others by referencing their key column.



Meta-markup languages

XML documents have a hierarchical structure, which gives them a tree-like appearance. Each element can have one or more children; there is exactly one root element.



RDF

Each fact about a data item is expressed as a triple, which connects a subject to an object through a precise relationship. This leads to graph-structured data that can take any shape.

Figure 8: Overview of data models (van Hooland & Verborgh, 2014)

Ever since the invention of writing, people have used lists to structure information. It is the most natural and intuitive way of organising data. From an IT perspective, this data model is serialised by using TSV or CSV file format. Figure 9: Serialisation formats (van Hooland & Verborgh, 2014) gives an overview of how flat files or tabular data (but also the three other models) are operationalised to be ingested into a computer.

data model	serialization formats
tabular data	CSV, TSV
relational model	<i>proprietary binary files</i>
meta-markup languages	XML, SGML
RDF	Turtle, N-Triples, RDF/XML

Figure 9: Serialisation formats (van Hooland & Verborgh, 2014)

Even if tabular data are very convenient for the ingestion and export processes, they have severe limitations. One cannot impose a structure and hierarchy, and above all, data values are encoded in an ad-hoc manner, which has a drastic impact on the quality and consistency of the data. Also, search and retrieval across large volumes of tabular data is ineffective and slow from a technical perspective. A tremendous step forward was taken in the 1970s with the development of the relational model. By making use of just three different concepts (entities, attributes and relations), relational databases allow to encode data in an extremely precise manner. By doing so, data consistency can be ensured, and search and retrieval across very large data volumes can be optimised. Relational databases can be considered as the biggest success story from the 20th century and, until today, they underpin most aspects of our information society.

However, with the increasing popularity of the Web in the latter half of the 1990s, organisations and people who invested decades of efforts in the creation and curation of large datasets within relational databases increasingly had the ambition to share and make available their data on the Web. This is the point where the disadvantage of relational databases came up due to their technical and semantic barriers for interoperability. First of all, databases are operationalised by making use of relational database software, such as MySQL or Oracle. If one wants to re-use the database of someone else, he/she will also need the same version of the software, as one cannot just copy/paste a binary file as it is the case with CSV or TSV files, which are non-binary text files. Individual tables can certainly be exported and downgraded to tabular data, but then comes the problem of semantics. Any database used to maintain a reasonably complex information system often contains hundreds of tables. If you want to understand the meaning of the data, you need to sit down and study in detail the database schema, which documents how the different tables are interrelated. Anyone who has to integrate databases or re-engineer a legacy database knows the complexity of this operation. Relational databases therefore also represent a semantic barrier to interoperability.

The arrival of XML after 2000, which is a meta-mark-up language, allowed bypassing the technical barriers of relational databases. As a serialisation format, XML is tremendously powerful: like tabular data, it is a non-binary text file format which can be opened and edited across platforms using a large variety of software tools. However, it can also impose a structure and very specific rules in regard to how data must be created. By providing an open standard to encode and exchange structured data, XML allowed to bypass the technical barrier towards interoperability. However, a complex information system based on XML also needs a schema, which defines how XML elements and attributes are used. This schema can then be used to validate the encoding of data. Interpreting an XSD schema, consisting sometimes of hundreds of definitions of elements and attributes, can be as challenging as reading a database schema. Therefore, even if XML allowed bypassing the technical barrier, the same semantic barrier towards interoperability would remain. In practice nowadays a combination of relational databases and XML data exchange is used. Archives still run their relational “silos” but have defined mappings between the relational structure and specific XML structures (and implemented these as service endpoints) which can be used to talk to / integrate with other “silos”, often implemented with the help of the OAI-PMH protocol.

So we see that both relational databases and XML offer wonderful possibilities to manage structured data, but they also have a big drawback: the need for understanding the schema describing the structure and interaction between the data. This is exactly where the last data model comes in.

Before getting into details regarding the data model, the need to bypass the problems associated with the reuse of locally defined semantics should be clarified. The vision behind the Semantic Web, as was introduced in the beginning of this section, was born out of the frustration of having only human-readable information on the Web, which restricts the ways in which software can help us find information. For instance, keyword-based search in a search engine like Google works well for terms such as “Jean Monnet”, merely based on string matching. Queries such as “publications by Jean Monnet” are already more difficult, since pages can use different wordings. This query for example would not retrieve HTML pages describing his memoirs. More complex queries such as “publications by politicians who worked with Jean Monnet” are utterly impossible to execute in Google. In the Semantic Web vision, the Web also becomes accessible for software agents instead of only containing human-readable information.

By adopting an extremely simple data model consisting of triples, data represented in Resource Description Framework (RDF) becomes *schema-neutral*. An RDF triple consists of a *subject*, a *predicate* and an *object*, as seen in Figure 10. This allows for maximum flexibility. Any resource in the world (the subject) can have a specific relationship (the predicate) to any resource in the world (the object). There is no limit on what can be connected to what. In order to implement this graph-based data model, URIs are used as key building blocks. For example, in order to express the fact that Jean Monnet had published “Memoirs”, the following triple can be published on a Web server and queried:

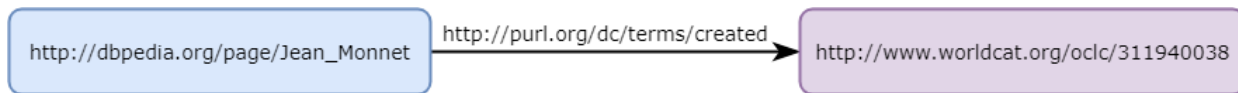


Figure 10. RDF triple example

The subject of this triple is Jean Monnet. Both a human and a software agent can look up the URI and obtain more information on Monnet. So, the URI serves both as a unique identifier and as a way to look up more information. The predicate is expressed by making use of one of the 15 Dublin Core metadata elements, allowing to express clearly (again, to both man and machine), what is meant by creating something. The object is expressed by pointing to the identification number of a specific work described in OCLC’s WorldCat database. The beauty of RDF as a data model exists in the fact that there is no need to look up and interpret locally defined semantics, but that there is a standardised and decentralised approach to define semantics so that both humans and machines can interpret the data. Any new type of information can be added without modification of the structure of the pre-existing data model.

At this stage it is essential, both from a conceptual and an operational point of view, to understand the differences of the *closed versus the open world assumption*. Relational databases and XML applications both excel at imposing rigid rules on what and how data can be encoded. These approaches are based on the closed-world assumption: only data which respects the schema and rules is allowed to exist in the information system. The Semantic Web and Linked Data vision is based on the open-world assumption, for which “*the absence of a particular statement means, in principle, that the statement has not been made explicitly yet, irrespective of whether it would be true or not, and irrespective of whether we believe that it would be true or not*”. The Open World assumption (Wikipedia, 2018) said in simpler terms is that, “*anyone can say anything about anything*”. Both paradigms come with their inherent strengths and weaknesses: the closed world assumption excels at rigorously maintaining coherence but represents a big bottleneck to create and disseminate information in a decentralised environment such as the Web. The open-world assumption provides the necessary flexibility to create data in a decentralised manner, but throws a lot of our traditional views about data quality and authority out of the window. Figure 11 summarises both the advantages and disadvantages of the four data models.

data model	(dis-)advantages	usage
tabular data	<ul style="list-style-type: none"> + intuitive approach + very portable + technology agnostic - prone to redundancy and leading to inconsistencies - inefficient search and retrieval 	import and export of data with a simple structure
relational model	<ul style="list-style-type: none"> + handling of complex data + optimized queries + mature software market - binary format - schema dependent 	management of complex data which require normalization
meta-markup	<ul style="list-style-type: none"> + platform-independent + both human and machine readable - complicated implementation for complex data - verbosity 	import and export of complex data
RDF	<ul style="list-style-type: none"> + schema-neutral approach + discovery of new knowledge - loss of normalization - immature software market 	making data available for linking

Figure 11: Comparison of four data models (van Hooland & Verborgh, 2014)

Although the large-scale adoption of the Linked Data paradigm has been problematic, there are unrecognisable advantages and possibilities. The best illustration of how everyone is consuming Linked Data on a daily basis is probably the Google search interface, which has had embedded data from Google's Knowledge Graph since a couple of years on the right hand side of its interface (Google Inside Search, 2018). As illustrated in Figure 12, the right hand side of the search results gives a good illustration of the RDF data model: the subject is the European Commission, which has a certain number of predicates, such as a headquarters and a foundation date. The values attached to the predicates as objects, such as Brussels (where the headquarters of the EC are based) can be considered as a subject itself by clicking on the link and obtaining more information about Brussels. The following section will point out the relevance of Linked and Open Data for the archival community.

Study on Standard-Based Archival Data Management, Exchange and Publication Final Report

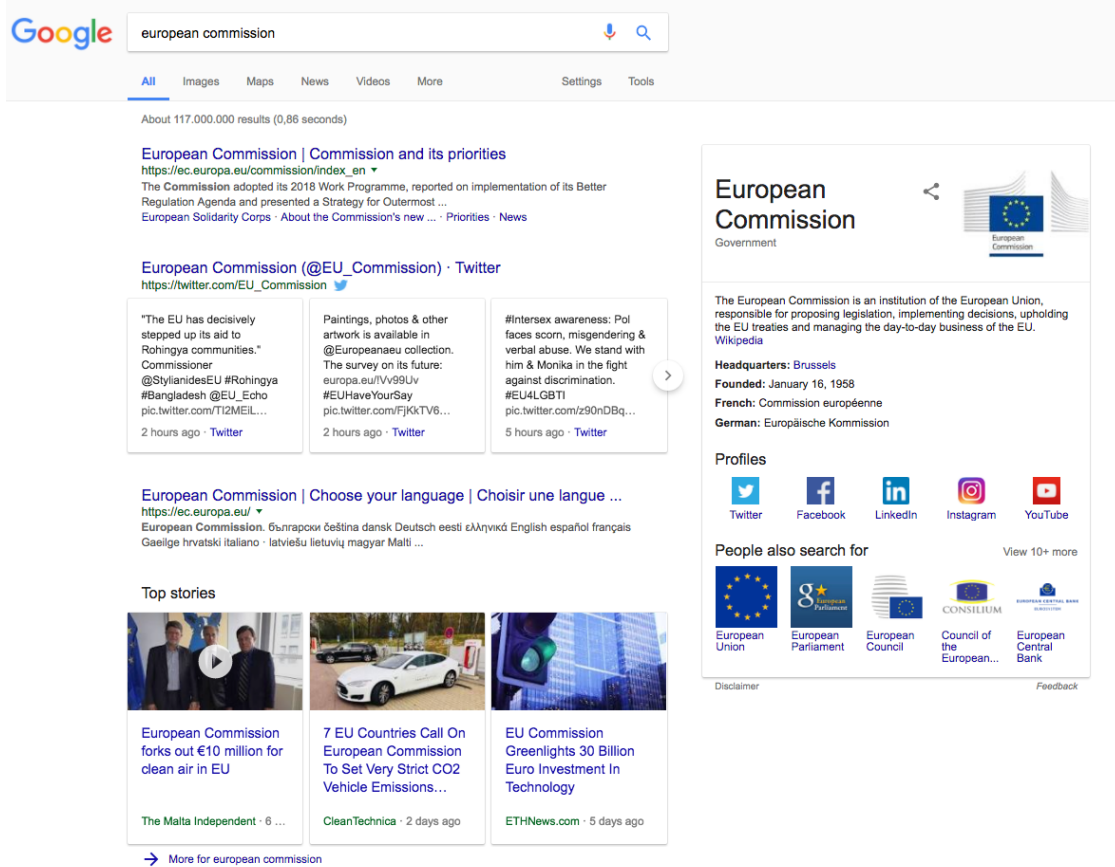


Figure 12: Example of the usage of Linked Data within the Google search interface

7.2.2 Relevance of Linked Data for the archival community

The promise of Linked Data to ensure technical and semantic interoperability across heterogeneous datasets sounds very promising to the archival community. Ever since the introduction of information technology within the archival domain, people have strived to consolidate finding aids and to offer end-users the possibility to query across aggregated finding aids with the help of portals. Before the study gives a more detailed overview of individual projects and initiatives which are working towards archival Linked and Open Data, this section will give a high-level overview of what the experiences and tendencies have been so far.

When going through the different data models, any archivist will have felt a natural preference with the third data model: XML. The hierarchical nature of XML provides a perfect fit with traditional descriptive practices in regard to paper-based archives, for which finding aids are developed on the principle of Russian nesting dolls. Primordial importance is given to the description of the fonds, broken down into series and files, in which the items ultimately reside. Metadata defined on a higher level are inherited on lower levels. This approach is in line with one of the biggest conceptual pillars of modern archival practice: provenance and original order, better known under the French term "respect du fonds". The focus of traditional descriptive practices does not lie in the description of individual items, as it is the case for example in other cultural heritage contexts (libraries and museums e.g.), but on how the fonds are constituted to respect the order of their components and add in this way important meaning and context. This approach has been fully embedded in the suite of descriptive standards developed by ICA, such as ISAD(G) and ISAAR(CDF), which have been adopted worldwide. When comparing with libraries and museums, it is also worth noting, that due to the high number of records which are usually submitted to an

archive at once, it has also been purely impossible to describe every individual item/record in full detail, justifying the hierarchical descriptive approach. However, in the current context in which line of business applications create digitally born records it has become possible to reuse original descriptions available from the source system. Therefore, partially reviewing the hierarchical description principles is justified.

So what does this all have to do with Linked Data? The current interest in graph-based data models renews the conceptual debate on the validity of the “*respect du fonds*”. The principle was first formally introduced by the French National Archives in 1841, moving away from previous approaches based on the content or the formal characteristics of the records. The notion that records should be kept in the original order in which the creator delivered them to the archive has been followed across the world. However, ever since the 1940s, individual archivists have questioned the validity of this principle. Iconic thinkers from the archival science and records management literature, such as Terry Cook, David Bearman or Luciana Duranti, have all criticised it for being a too literal interpretation of the importance of provenance. Peter Horsman expressed a synopsis of the critiques clearly in 2002 (Horsman, 2002): *“Archival methods centred on respect des fonds, therefore, serve custody and the convenience of the archivist in managing collections in tidy and well-defined groupings. They do not necessarily serve users or researchers. Of course archivists pretend – and they may actually believe – that their own administrative convenience also best serves users by protecting provenance. The user, however, has often been seriously misled by archivists and their fonds. The archival methods of arrangement and description, based on respect des fonds, present to the user a monolithic “grouping” of records that in reality never existed at any one point in time, outside the archives.”* Popovici provides the most recent and solid overview of the different critiques (Popovici, 2016). He concludes that *“it could be agreed that description from the general to the specific or non-repetition of information cannot offer a full support for describing contingent relationships, in a ‘non- logical’ arrangement. And, despite the fact they are pertinent in ISAD(G) mind-set, it has to be noticed, if relationships among archival entities would be more flexible, it would mean that each entity should be self-sufficient, meaning, furthermore, information would be in a certain amount, repetitive. A simple query retrieving an entity ‘file’ needs to give some information about upper levels also, in order to be contextualised”*.

Based on these critiques, the Expert Group on Archival Descriptions (EGAD) from ICA started the development of a conceptual model for archival description, named Records in Contexts Conceptual Model (RiC-CM), which should rework some of the basic assumptions behind archival descriptions. For example, no big changes are introduced on the level of the description entities as defined in the RiC-CM:

- RiC-E1: Record
- RiC-E2: Record Component
- RiC-E3: Record Set
- RiC-E4: Agent
- RiC-E5: Occupation
- RiC-E6: Position
- RiC-E7: Function
- RiC-E8: Function (Abstract)
- RiC-E9: Activity
- RiC-E10: Mandate
- RiC-E11: Documentary Form
- RiC-E12: Date

- RiC-E13: Place
- RiC-E14: Concept/Thing

These entities present the core archival objects, based on the analysis of the existing ICA standards, the FRBR, the CIDOC-CRM and models from Australia, New Zealand, Spain and Finland. An in-depth overview or analysis of RiC is outside the scope of this study (Dunia Llanes-Padrón, 2017), but one of the biggest changes in relation to ISAD(G) is the possibility of bypassing purely hierarchical relations. Popovici provides the following imaginary example:

Record1	is part of	Record set1 (title = file1)
Record1	is part of	Record Set2 (title = series1)
Record1	is part of	Record Set3 (title = fond1)
Record Set1	was part of	Record Set4 (title = series2)
Record1	is created by	Agent1
Record1	is collected by	Agent2
Record1	is evidence of	Function1
Record1	results from	Activity1
Activity1	Is performed to fulfil	Function1

Figure 13: Example provided by Popovici (2016)

This example illustrates understanding how the non-hierarchical view on archival descriptions finds a better home with RDF as a data model, compared to XML. Moreover, it underlines how the conversion of finding aids from a hierarchical to a graph-based model requires an exponential duplication of data, which substantially undermines the practical implementation of this approach. At this stage, it is important to mention that the RiC-CM work is mostly a theoretical exercise. No ontology that allows the implementation of the conceptual model has been developed. From a theoretical perspective, the approach is intellectually interesting, but how can it be implemented? The answer to this question is simple: no one really knows right away.

The debate in regards to the possibilities and limits of both XML and RDF sometimes leads to the impression that the approaches are exclusive, which is not the case. In practice, archival holders can augment their traditional hierarchical finding aids by inserting “snippets” of Linked Data. The changes implemented in EAD3 provide now more flexibility to include links to external data sources within the finding aid. Also, initiatives such as the W3C Architypes working group demonstrate how archival holders may leverage [Schema.org](https://www.schema.org/) to enrich finding aids when publishing them as HTML. As underlined by Kuldar Aas from the National Archives of Estonia (NAE), the functionalities linked to either the “management of archives” and the “access to archives” do not necessarily need to follow the same model. Parallel methods can be used for these significantly different processes. For example, the NAE still continue to manage their holdings according to the provenance principle, but in addition also offer users ways for accessing content from various vocabulary/ontology based viewpoints (i.e. search for persons, places and events).

From a European perspective, the NAE are probably one of the earliest adaptors of the Linked and Open Data approach. As early as in 2011, extensive work has been undertaken to replace the Archival Information System (AIS) used for the management of the finding aids, and a study of existing tools and the market space was launched. However, the conclusion was that no existing satisfactory solution (commercial or open source) was offered. The traditional solutions mainly lacked native support for the Linked Data vision of the NAE, for which they need features such as ontology management and the long-term management of persistent IDs. Also, software from the traditional archival market space, such as ScopeArchiv, have for example fully implemented EAD in all of its details but do not leave any room to

make modifications. This type of archival software is still too based on the scenario of the manual encoding of finding aids. After realising that no existing solution met their needs, the NAE started to develop its own software solution from scratch with the help of an outside contractor, based on the Symfony PHP framework. Seven years onwards, the NAE now has a beta release which represents a change of paradigm in regard to usability and search capabilities. The end-user interface offers access to the holdings via various facets such as persons, periods, places, topics and subject areas (high level functional thesaurus). For each record, the catalogue offers the possibility to download the metadata in RDF.

However, there have been various scalability and stability issues which have been troublesome and resource-demanding to solve. Most of the issues have been solved and the aim is to make the service public before the summer of 2018. The overall sentiment is that the project was too ambitious and too far ahead of its time. Another important lesson, relevant to EU member states which represent smaller language groups, is the fact that implementation of Linked Data is heavily dependent on language-specific annotation / semantic mark-up tools, which have been optimised for English (and to a lesser extent to French and German). Also, many annotation tools work with “widely-known current concepts” available in DBPedia or Wikidata. Archival content is often about long deceased and lesser known individuals (or renamed places) for which no reasonable references are available for automated annotation.

Even though it has not taken steps towards a large scale application as in the case of Estonia, The National Archives (TNA) of the UK has clearly expressed interest in the LOD paradigm and has actively promoted the interoperability and publication of archival data through projects such as <http://www.nationalarchives.gov.uk/about/our-role/plans-policies-performance-and-projects/our-projects/traces-through-time/>. The “Traces through time” project has focused on the usage of semantic markup to automatically detect person names within semi- or non-structured text and connect them probabilistically. However, as Matt Hillyard explained, vast amounts of work on data normalisation had to be performed within the process, which resulted in the creation of one million published links between records, limiting the scope of the project to that of a relatively big demonstrator. The approach is promising and hopefully will be picked up by other archival institutions and research projects, but it underlines the complexity of automated annotation, even in English.

Within the debate in regards to the Semantic Web and ontology engineering, John Sheridan, the Digital Director at TNA, highlights the importance to reflect on the “archival exceptionalism”. As a community, what is really specific about our data modelling and what existing schemas, metadata elements and ontologies can be reused? On the level of provenance for example, the work on the modelling of provenance metadata by W3C could definitely be of use (W3C, 2013). Also, confronted with the problem zone of the “open-world assumption”, underlying the LOD paradigm, archivists should be able to handle data quality issues with automatically generated metadata. It is therefore interesting that the concept of *probabilistic description* is forwarded by TNA, which refers to “*acknowledging in a transparent manner that data is imperfect and embracing uncertainty. This confidence rating might be a combination of computational and curatorial scores.*” (UK National Archives, 2017)

In the context of the interviews carried out for the report, we also spoke with the coordinator of data.bnf.fr, which is the LOD platform maintained by the Bibliothèque Nationale de France (BNF). Traditionally, the library community has been an early adaptor of technologies, and archivists have often taken over or adapted methods and tools from librarians within their own context. BNF has extensive experience with the hosting of very large LOD datasets, and has successfully published various controlled vocabularies such as RAMEAU. However, the interview also underlined the technical complexity of indexing and maintaining very large sets of data as RDF triples. The different business processes (catalogue, management of digital assets, etc.) of the BNF are still managed in relational databases, but the metadata are exposed to the outside world as triples. The updating process between the SPARQL endpoint and the various lines of business applications is particularly troublesome. It is also worth mentioning that the BNF

decided not to adopt BIBFRAME, which is the follow-up format to migrate from MARC to a format which builds natively on the LOD approach. BIBFRAME is developed and actively promoted by the Library of Congress (LoC). In parallel with the development of the format, LoC financed the development of cataloguing software, which should be natively based on RDF, but no working software has been published, nor is there a known deadline (Library of Congress, 2018). A more pragmatic initiative is the development of specific tags to describe library metadata with Schema.org, which proposes a set of “extensible schemas that enables webmasters to embed structured data on their web pages for use by search engines and other applications”. The W3C BibExtend Community has worked on a limited set of specific tags, allowing libraries to make their metadata available in a structured format accessible to search engines (W3C, 2015). Based on this work, the archival community has launched a similar undertaking with the W3C Archetypes Working Group, which is currently discussing what specific new elements need to be developed in order to expose archival metadata in a more meaningful manner to search engines. (W3C, 2018). This is a good example of how traditionally managed finding aids can be exposed as Linked Data towards end-users.

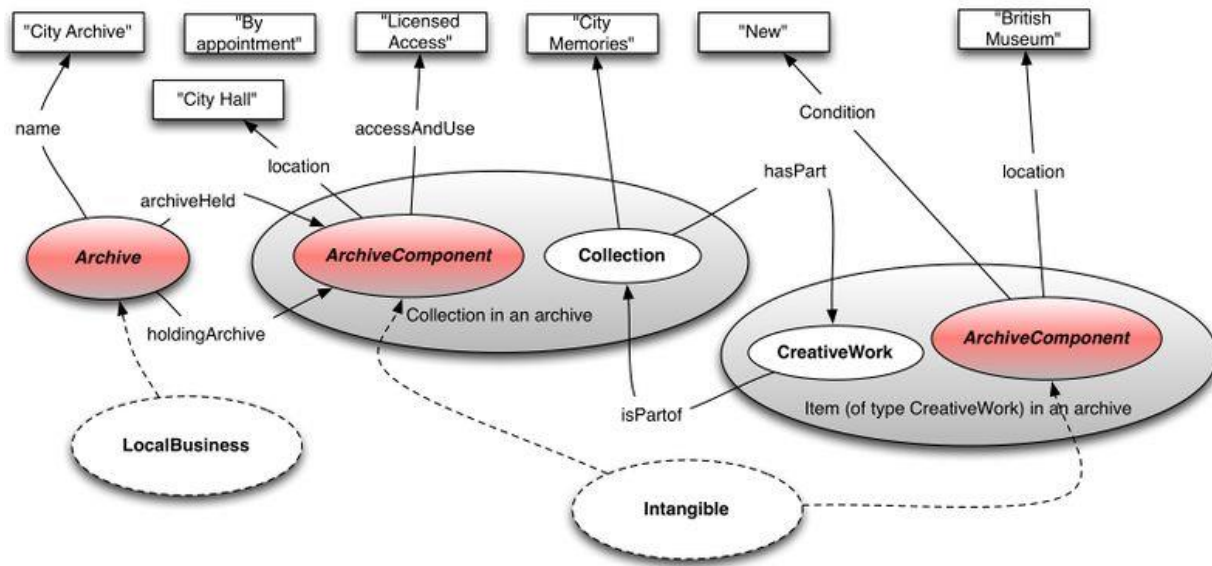


Figure 14: Proposal of an extension for Schema.org by the W3C Archetypes working group

Another refreshingly pragmatic approach to the debate around the usage of LOD within an archival context comes from Ross Spencer of the Archives New Zealand, who has been working on the automated creation of links in between records (Spencer, 2016). Spencer proposes a radically simplified approach, compared to the Records in Context (RiC) archival standard which describes 73 potential record-to-record relations, by focusing on eight relations:

- Relationship one: identical records
- Relationship two: similar records
- Relationship three: contains hyperlink
- Relationship four: contains enterprise content management system (ECMS) reference
- Relationship five: contains embedded digital objects
- Relationship six: contains intra-item relationships
- Relationship seven: contains object references

- Relationship eight: contains item mentions

Within the paper, Spencer gives an outline on how to extract this set of relationships by using free, and open-source, tools and toolchains. This approach also allows us to make the connection between LOD and the automation of archival and records management.

7.2.3 Relation to the automation of Records and archival practices

Regardless the conceptual discussion surrounding a hierarchical versus a graph-based data model, there is a more urgent question: how to appraise the excessively large volumes of electronic records which are currently being produced? NARA and TNA, arguably two of the leading archival institutions worldwide, have both underlined in recent years the radical impact of the digital deluge on appraisal, selection and sensitivity review processes for documents when transferring records to historical archives. The current evolutions introduced by the implementation of the General Data Protection Regulation (GDPR) on a European level underline the challenges records managers and archivists are currently facing. Striving towards compliance with both the protection of personal data and at the same time providing access to documents can be challenging. Most national archives have moved on to a pre-appraisal approach where instead of individual records they appraise agencies' business process – should a process be regarded as “of archival value”, all records created within this process get archival value automatically. This approach does not care about the number of records/data created and is well manageable. For example, in Estonia, more than 90% of agencies have been pre-appraised, meaning that to a large extent the government knows about archival value even before a piece of data is created or received. However, archives who have not implemented this pre-appraisal approach still rely on manual selection workflows, resulting in delays and incoherent decision-making. The 2015 judgment of the Court of Justice in case C-615/13P “Client Earth and Pan Europe versus EFSA” (InfoCuria - Case-law of the Court of Justice, 2015) has amply demonstrated the complexity of respecting Regulation (EC) No 45/2001 on the protection of personal data in parallel with access to document requests submitted under Regulation (EC) No 1049/2001 (European Law Blog, 2015).

However, this is not a recent observation. In his classic paper *Electronic Records, Paper Minds* from 1994 on the impact of the *digital turn* for the archival profession, Terry Cook's main thesis is that archivists and records managers can “*no longer afford to be, nor be perceived to be, custodians in an electronic world*” (Cook, 1994). Throughout the last two decades, the responsibility for declaring and organising records has shifted down towards end-users. This approach has often proved to be highly problematic, as underlined by (Vellino & Alberts, 2016). In parallel, some recent initiatives have demonstrated how civil society can successfully participate in the appraisal process. For example, the Swiss Federal Archives started up a program to pro-actively reach out to citizens. In November 2017, they organised a workshop for which they invited the public to debate on the appraisal of records in relation to public road infrastructure.¹⁴

To what extent can algorithms assist with the appraisal process? Commercial vendors and computer scientists are waving the magic wand of statistics and machine learning to make sense of large volumes of structured and non-structured content. More and more data scientists are being hired to tap into content and metadata scattered across shared drives and lines of business applications to discover trends and outliers for business intelligence. In this context, records managers can “*function as a partner in the analytic process, providing information about data's location, and improving the visual analyst's understanding and trust of data through explaining their context of creation, the history of their structure and semantics and their chain of custody*” (Integrity in Government through Records Management: Essays in Honour of Anne Thurston).

¹⁴ See <https://www.bar.admin.ch/bar/de/home/informationsmanagement/archivwuerdigkeit/diskutieren-sie-mit-.html>

However, a lot of misunderstandings and false hope circulates among the archival community. The 2014 NARA directive “Automated electronic records management report/plan” (Government, 2014) has been a landmark document, acknowledging the necessity to embed automation as an essential aspect within a records management strategy. The report distinguishes five different approaches of automation: no automation (manual management), rule-based automation, business process and workflow oriented automation, modular re-usable records management tools, and auto-categorisation. The report provides a much needed overview of the urgency of appraisal automation. However, NARA’s typology mixes methods (manual, rules-based), implementation (modular re-usable records management tools) and functionalities (auto- categorisation).

In 2016 TNA provided a more practical follow-up report entitled “The application of technology assisted review to born-digital records transfer, inquiries and beyond” (The National Archives, 2016).

Both reports give an overview of different methods and tools, but it is essential to distinguish two different approaches: automation can either be based on *rules* or on *statistics*:

- **Rules:** based on an abstract model of the content and its application domain, decisions on content can be automated. NARA’s Capstone approach to email is a simple example of this approach: from the moment, someone has a certain position within the hierarchy of an organisation, his or her email is automatically captured for example.
- **Statistics:** based on an analysis of the content itself, making use of either supervised or non-supervised machine learning techniques. Auto-classification tools to categorise email as having business value or not, based on a training set, is a typical example of supervised machine learning.

Both approaches have their advantages and limitations, which will be pointed out. On the terrain, both rules and statistics can be combined, as will be discussed.

Ever since the 1960s, the artificial intelligence community developed methods to represent knowledge and algorithms which can infer new knowledge from the pre-defined set of rules. Rules-based systems require the user to define rules, so that software can infer what to do in a certain situation. The danger of this approach is that if the rules miss a scenario, noise is generated as output, requiring ever more rules to be able to describe every possible scenario. In the 1980s, this strand of research culminated in the creation of the then-called expert systems. This type of software consisted of knowledge bases or ontologies containing a large amount of facts and statements connected by making use of formal logic. The drawback of this approach is the lack of adaptability: the system can only function based on the information it has. This implies that these systems can only be operational within well-delimited specialised application domains, such as a specific medical discipline. Also, the cost of creating and maintaining the rules tends to be prohibitive. (Research in the Archival Multiverse, 2017)

In the last two decades, we have seen a rise in not only the amount of data available and the volume of documents, but also in the variety of data types, complexity of sources and lack of structure. Hence, we see a shift from knowledge-driven methods to data-driven methods, which means that traditional rules are in general left behind, leaving room for statistical systems trying to find structure in the wealth of information available today. The tremendous advantage compared to the previous rules-based approach is that there is no need to develop an a priori model of an application domain, which is then used to apply the rules. Chris Andersons framed this change of paradigm boldly by stating that “*with enough data, the numbers speak for themselves*”. (Wired, 2008)

When introducing machine learning algorithms, an important distinction has to be made between so-called supervised and unsupervised methods. Over the last few years, the archives and records management community has almost exclusively experimented with supervised machine learning methods, which are based on humans feeding test data to the algorithm. Since a couple of years big software vendors, such

as OpenText for example, offer auto-classification tools allowing to automatically put documents into predefined categories. The software offers easy-to-use interfaces allowing records managers to select a test corpus, perform the manual classification of documents into a limited number of categories and then to check the quality of the auto-classification based on sampling. However, vendors do not provide any benchmarking studies or clear methods to assess the quality of their tools in an objective manner. Vellino and Alberts recently published a very detailed study on the possibilities and limits of automatically appraising email. The article underlines the need to formalise the organisational context by conducting semi-structured interviews and cognitive inquiries, followed by a data analysis (Vellino & Alberts, 2016). Based on this input, an abstract classification model was built, consisting of two top-level categories: emails with and without business value, further divided into 13 sub-categories. This study makes it very clear that the application of auto-classification requires substantial efforts and is not as straightforward as vendors suggest. In this context it is worth repeating that auto-classification is also dependent on the level of technological support available for a language, as such it does perform even worse for smaller languages (which is relevant in the multilingual EC context).

As the application of supervised machine learning is not as straightforward as many believe, this study wants to point out the possibilities of two unsupervised machine learning methods for archival holdings: Topic Modelling (TM) and Word Embeddings (WE). The term unsupervised is used as the process does not involve any pre-trained corpus. Let us first introduce Topic Modelling (TM), which has gained momentum over the last few years within the digital humanities to explore and interpret very large corpora of full-text documents.

In this context the mass digitisation of the EC's archives offers new exciting possibilities to query and analyse in an automated manner an archival corpus. However, there is a large gap between the promises made by big data advocates, who rely on statistics to discover patterns and trends in large volumes of non-structured data, and how historians can actually derive value from automatically generated metadata to explore archives and find answers to their research questions. Currently, researchers can only perform full-text queries in order to make sense of this massive corpus, as illustrated in Figure 15. In the context of a research collaboration, researchers from ULB applied various machine learning methods and are currently preparing a set of publications on the research outcomes. (Hengchen, Coeckelbergs, Hooland, Verborgh, & Steiner, 2016)

The screenshot displays the search interface of the European Commission's Historical Archives. At the top, there is a navigation bar with links for 'A-Z Index', 'Sitemap', 'About this site', 'FAQ', 'What's New', 'Legal notice', 'Cookies', 'Contact', and 'Search'. The language is set to 'English (EN)'. The main header features the European Commission logo and the text 'HISTORICAL ARCHIVES'. Below this, a breadcrumb trail reads 'European Commission > Historical Archives > COM files'. A sidebar on the left lists various categories: 'What we offer' (Archival fonds, Collections, Availability Rules), 'Searching' (Archival Fonds, Speeches, Minutes of meetings, COM files, ECSC inventory), and 'Visiting the archives' (FAQ). The main content area shows a search for 'ARCHISplus: searching COM files' with 69 results. The search criteria are 'Title contains : (%atomic%) and (%energy%)'. A 'Modify search' button is available. Below the search results, there is a pagination control showing page 1 of 7. The results table has columns for 'Ref', 'Title', and 'Date'. Three results are visible, all dated 05/02/1973, 07/02/1973, and 14/03/1973. The titles mention 'Atomic Energy Community' and 'MEMORANDUM OF THE COMMISSION AND DRAFT COUNCIL DECISION APPROVING THE 'SUBSIDIARY ARRANGEMENT''. On the right side, there are several informational boxes: 'Live web support', 'Who are we?' with the quote 'Remembering the past, to build a better future', 'Links' (Archives — EU institutions, Historical Archives of the European Union in Florence (HAEU), Archives Portal Europe, Other links), and 'U.E.' (The Estonian presidency of the Council of the EU: 1 July - 31 December 2017, Join the European debate).

Figure 15: Search interface of the COM files of the EC archives¹⁵

The dataset, spanning a period ranging from 1958 to 1982, is multilingual: it contains documents in French, Dutch, German, Italian, Danish, English and Greek, as those were the then official languages of the now-called European Union. For this experimental case-study, only the English corpus was taken into account, which represents a total number of 11.868 documents. In the context of the first exploratory study, Latent Dirichlet Allocation (LDA), which is the most popular TM algorithm, was applied on the corpus. As was already mentioned, the dataset presents close to no metadata on a document level: apart from an XML file corresponding to each PDF and containing basic information such as a unique identifier, a creation date, the number of a reference volume and the language and title of the document, no information on the actual content of the digitised documents is given. There is no insight as to what the documents encompass in terms of topics and themes, which makes the dataset difficult to use for historians. In the context of the first experiments with the dataset, researchers from ULB manually

¹⁵ http://ec.europa.eu/historical_archives/archisplus

interpreted the topics, in order to attach a descriptor from the EUROVOC thesaurus¹⁶. Figure 16 gives three examples from topics and the EUROVOC descriptors which were manually attached to the topics.

URI	label	tokens			
http://eurovoc.europa.eu/2965	agricultural aid	agricultural premium farms	areas directive production	aid number	measures eec
http://eurovoc.europa.eu/852	ECSC aid	coal industry measures	steel production community	ecsc iron	aid decision
http://eurovoc.europa.eu/1418	textile industry	fabrics crocheted products	textile fibres yarn	woven community	knitted agreement

Figure 16: Manual labelling of TM results with Eurovoc

It is important to underline that the authors within this first exploratory study were unable to attach a label to around 30% of the topics, due to either the very general nature of the terms (e.g. *agreement community parties negotiations*) or the fact that the authors were unable to find a semantic link between the terms (e.g. *lights bmw brazil eec coffee*). For some topics, OCR noise resulting in terms such as *cf, ii or ir*, was the main cause.

However, the manual labelling of topics with descriptors from the EUROVOC thesaurus is of course suboptimal. One of the key problems is the interpretation of the clusters of terms which form a topic. Throughout the examples, one can sense that in the majority of cases, topics do not point out to one clear concept, but often are a combination of concepts. This aspect makes the manual labelling process inherently subjective and troublesome. Ideally, one would also want to perform an automated reconciliation process (Van Hooland, 2013). Unfortunately, the semantic heterogeneity of topics also constitutes a stumbling block for this process, as there is no way to indicate within the reconciliation process how the different concepts within a topic should be tackled separately.

TM can be viewed as a method to learn more about the topics addressed in a large corpus of documents, whereas Word Embeddings (WE) are considered as a general, vectorial representation of language itself, allowing us to understand the distance between words. In the context of his doctoral research, Mathias Coeckelberghs designed an original methodology which brings together both sources of information.¹⁷ As WE allow the production of vectorial representations for language as a whole, this then allows us to estimate for example the semantic relatedness of terms found in the same topic. In other words, WE can be leveraged to automate the identification of different concepts present in one topic.

In order to make things more concrete, we can refer to an example from the EC archival corpus. As explained, TM creates a probabilistic model which clusters a determined number of keywords extracted from a document collection together in so-called topics. An example of a topic based on the archival holdings of the EC could be the following cluster of ten terms:

Gas fuel energy electricity coal power nuclear supply industry production

Upon reading the cluster of keywords, we understand that the subset of documents from the archival corpus with this topic probably address how the EC dealt with the usage and supply of energetic

¹⁶EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU. It contains terms in 23 EU languages and in the three languages of countries which are candidate for EU accession. It is managed by the Publications Office and it is mainly used by the European Union Institutions, the Publications Office of the EU, national and regional parliaments in Europe, plus national governments and private users around the world (<http://eurovoc.europa.eu/>).

¹⁷ Mathias Coeckelberghs is currently preparing an in-depth paper to present the usage of WE to interpret the results of TM.

resources. This example allows to point out the possibilities, but also one of the problematic aspects of TM, namely the interpretation of the topics. The combination of terms is often difficult to interpret as they cannot be mapped easily to one single concept, but rather consist of a combination of two or more concepts. In the case of the above example, WE for example help to split out the topic over two concepts:

- gas fuel energy electricity coal power nuclear
- supply industry production

The full set of results can be analysed on Github (Coeckelbergs, 2018). Within the data set, three different colour codes are used, which help to visualise the following different outcomes of WE on the TM results:

- Terms in orange indicate a topic which represents one single concept
- Blue and red are used to indicate the first and the second concept in a topic consisting of two different concepts
- Terms in light-blue are terms which do not indicate a clear link with the terms from the topic which surround them

Based on the examples analysed, there are clearly cases where WE does deliver a clear added-value to interpret the outcomes from TM. How can this help archivist? In future work, we plan to experiment with a reconciliation process between the terms from the topics and the EUROVOC thesaurus. The fact that we can automatically divide one topic into two different concepts, will allow us to increase the relevance of the reconciliation results, as we will not be forced to automatically assign one label to a topic which actually represents two different concepts. This is also one of the foundational concepts in the Records in Context effort – that a specific record (or even piece of information) can indeed be part of multiple business processes and thus be classified into multiple places! Any classification method which assumes a single classification is therefore almost unusable in a current business context.

7.2.4 State of play within the EC on Interoperability

As introduced in 2.1 Context, the main challenge of the digital management of the European Commission archives is the information exchange between different European institutions. At the European level, the complexity related to the organisations themselves, the types of documents they produce and the kind of metadata required to describe and preserve these documents is very high and could be an important barrier for the data archiving process. For this reason, the functional model developed in this study has to take into account these levels of complexity to enhance the potential synergies between the institutions, to represent the different domains involved in the democratic life (tenders, research, laws, economics, etc.). For this reason, the generic interoperability models are fundamental to cover the data exchange aspect between the institutions of the European Commission.

In order to promote the interoperability in Europe, the ISA² programme of the European Commission developed the European Interoperability Framework (EIF) (European Interoperability Framework, 2017). The EIF contains the main principles to be satisfied in order to expose trans-European services to data exchange between some of the institutions of the European Commission. The EIF principles and the related best practices are relevant for the design and implementation of the typical Open Data platforms that have to integrate, publish and share public information such as archives.

As data archiving follows the same requirements of any public infrastructure to manage digital information between several public organisations, here is a basic description of the 12 principles of the EIF that have to be considered:

Interoperability principles	Objectives
Subsidiarity and proportionality	Ensure that national interoperability frameworks and interoperability strategies are aligned with the EIF and, if needed, tailor and extend them to address the national context and needs.
Openness	Publish owned data as (Linked) Open Data unless certain restrictions apply.
	Ensure a level playing field for open source software and demonstrate active and fair consideration of using open source software, taking into account the total cost of ownership of the solution.
	Give preference to open specifications, taking due account of the coverage of functional needs, maturity, and market support and innovation.
Transparency	Ensure internal visibility and provide external interfaces for European public services.
Reusability	Reuse and share solutions, and cooperate in the development of joint solutions when implementing European public services.
	Reuse and share information and data when implementing European public services, unless certain privacy or confidentiality restrictions apply.
Technological neutrality and data portability	Do not impose any technological solutions on citizens, businesses and other administrations that are technology-specific or disproportionate to their real needs.
	Ensure data portability, namely that data is easily transferable between systems and applications supporting the implementation and evolution of European public services without unjustified restrictions, if legally possible.
User-centricity	Use multiple channels to provide the European public service, to ensure that users can select the channel that best suits their needs.
	Provide a single point of contact in order to hide internal administrative complexity and facilitate user access to European public services.
	Put in place mechanisms to involve users in analysis, design, assessment and further development of European public services.
	As far as possible under the legislation in force, ask users of European public services once-only and relevant-only information.
Inclusion and accessibility	Ensure that all European public services are accessible to all citizens, including persons with disabilities, the elderly and other disadvantaged groups. For digital public services, public administrations should comply with e-accessibility specifications that are widely recognised at European or international level.
Security and privacy	Define a common security and privacy framework and establish processes for public services to ensure secure and trustworthy data exchange between public administrations and in interactions with citizens and businesses.

Multilingualism	Use information systems and technical architectures that support multilingualism when establishing a European public service. Decide on the level of multilingualism based on the needs of the expected users.
Administrative simplification	Simplify processes and use digital channels whenever appropriate for the delivery of European public services, to respond promptly and with high quality to user requests and reduce the administrative burden on public administrations, businesses and citizens.
Preservation of information	Formulate a long-term preservation policy for information related to European public services and especially for information that is exchanged across borders.
Efficiency assessment	Evaluate the effectiveness and efficiency of different interoperability solutions and technological options considering user needs, proportionality and balance between costs and benefits.

Table 8: Principles of the EIF

These aspects will be considered when the initiatives and methodologies are analysed and assessed and positioned in the Magic Quadrant (see section 7.3.1 for further information).

One of the important attributes for interoperability is Openness (Cf. previously). This attribute, when applied to the data model specification, offers the possibility to use standardised models with a high level of generalisation. In this way, the normalisation of the data based on the common data formats (XML-based formats) or reference data highly facilitates the data exchange. More specific schemas such as the ontologies promoted by the Semantic Web initiative provide a better expressiveness of the data representation to control semantic ambiguities. More details are provided in the Linked Open Data section.

In the domain of Linked Open Data, a good example of the data governance plan is provided by Linked Open Government Data (LOGD). They propose to focus the effort on three different principles to enable data consumption:

- Open: share Open Data in a structured format such as Linked Data
- Link: structure and contextualise automatically the data with advanced vocabularies and ontologies
- Reuse: consume the data easily by using browsing and visualisation tools or by automatic importation through SPARQL endpoint

In the figure below, a generic roadmap depicts these aspects. Interactions between the data, agents and functionalities are represented to explain the process to be followed in order to produce Open Metadata, contextualise them (linked with existing vocabularies) and reuse them through specific portals or publication services.

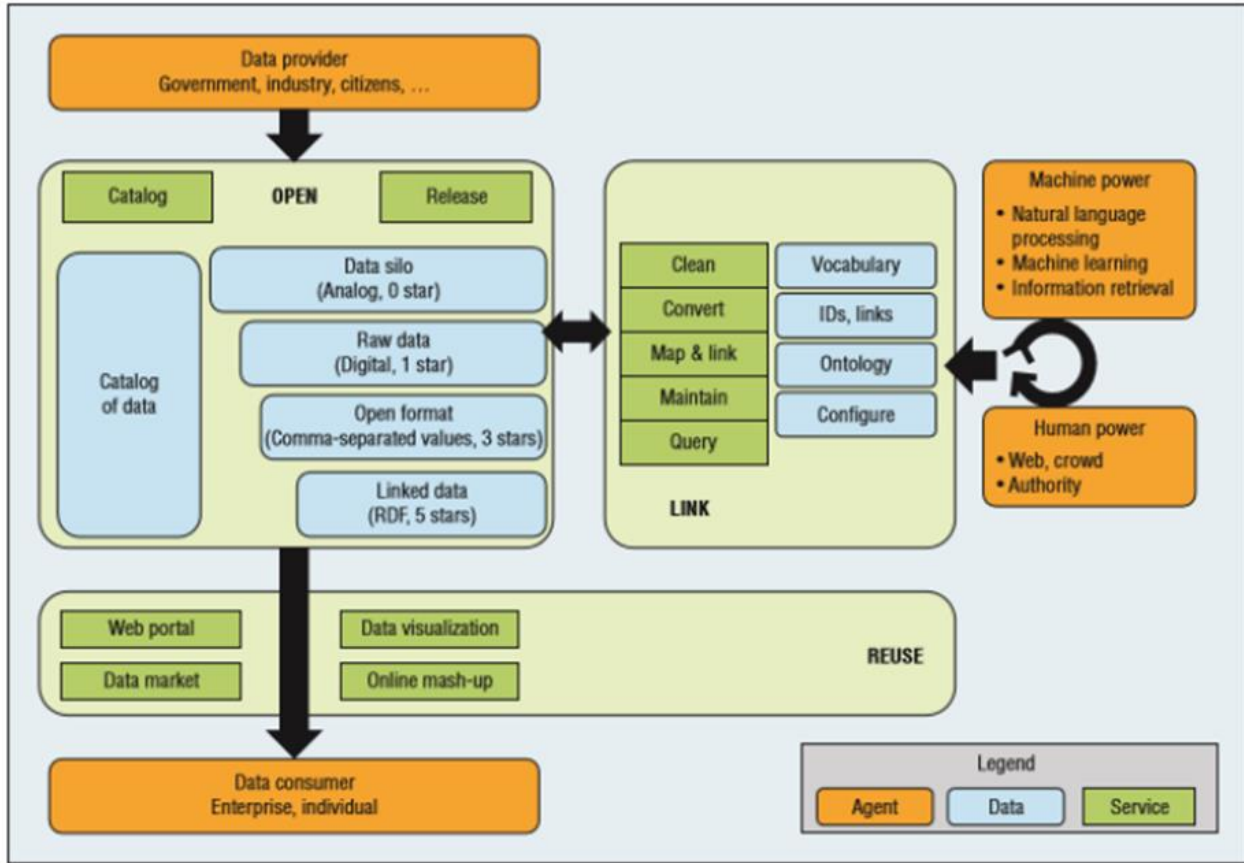


Figure 17: Roadmap of linked open government data (Ding, Peristeras, & Hausenblas, 2012)

In the table below, the formalised workflow to structure the data is presented. The linked data workflow defines a set of activities needed to transform structured data into linked open data. The workflow is conceived as a sequence of data processing services (NLP, data transformation, etc.) that collaborate to transform the data into LOD to enable the linkage and extraction of knowledge from this data in order to satisfy the aforementioned functional aspects (Open, Link and Reuse).

Linked Data workflow	Steps and description
Design	Prior to all phases of the Data Curation cycle, it requires the detailed analysis of the current data structure and the development of improvements: <ul style="list-style-type: none"> • <u>Data analysis</u>: Implies an analysis of the existing schemas of the structured documents and the possible structures to identify in the non-structured documents. • <u>Data modelling (context, structure, content)</u>: The main entities identified and their corresponding properties. This is the basic material required to build the core ontology, extended with other domain ontologies and reference data.
Collect Data collection, transformation	Retrieval of the relevant information fields to be enriched or curated (integrated in the validation workflow; described in the Data Management Plan): <ul style="list-style-type: none"> • <u>Data collection</u>: All information from external and internal sources is collected

and validation	and validated against the expected source format.
Normalise	<ul style="list-style-type: none"> • <u>Data cleaning</u>: The data is processed to strip text fields from unnecessary or meaningless information. • <u>Data transformation</u>: Transformation of source formats to fit in the generic structure.
Enrich	<p>Enrich / Complete: During this step, the data will be enriched inside the knowledge extraction pipeline through different semantic techniques:</p> <ul style="list-style-type: none"> • <u>Text processing</u>: Involves indexing content with NLP techniques (Tokenisation, Lemmatisation, etc.), and extracting the relevant keywords and acronyms. • <u>Automatic classification</u>: Consists on mapping the relevant keywords extracted from the content with the categories of any relevant taxonomy (e.g. EuroVoc.) • <u>Knowledge extraction</u>: Is an identification of the specific entities (e.g. acronyms, concepts or Named Entities such as Person, Organisation, Location, Event, etc.) and an automatic annotation of the content.
Link	<p>This phase is based on the concept of finding connections with validated reference datasets and improve the semantic coherence of the corpus by defining ambiguous elements with precision:</p> <ul style="list-style-type: none"> • <u>Semantic Annotation and Entity Linking steps</u>: Discover and associate the potential reference of the entities available in the LOD repositories • <u>Reasoning and semantic disambiguation</u>: Replacement with the reference entity, merging of records and update missing or incomplete data. Enriched annotations with inferred data
Validate	<ul style="list-style-type: none"> • <u>Data validation</u>: Application of business, semantic and technical validation rules;
Publish	<p>This step will complete the Data Curation workflow with the persistence of newly generated data:</p> <ul style="list-style-type: none"> • <u>Data publication</u>: Involves publishing the content in an advanced data portal (e.g. such as LOD through a semantic repository and exposed through a SPARQL endpoint).
Preserve	<p>This step defines what and how must be preserved, including both content and metadata. In the Linked Open Data approach, the preservation operation consists of the extension of the conceptual model with archiving metadata and to link it to the digital object.</p>
Reuse	<p>The final step will apply known semantic techniques to promote the dissemination, exploitation and re-use of the curated data and includes following steps:</p> <ul style="list-style-type: none"> • <u>Data search</u>: Stands for: (i) Term-based search: users can search for content by writing the query with keywords; (ii) Semantic search: possibility to look for specific entities by selecting the type of entity and their relative keywords; (iii) Hybrid search: combines the semantic and the term-based search. • <u>Data consumption</u>: Is about: (i) collecting specific datasets through the SPARQL endpoint: the associated content available in LOD can be used to

	enrich the description; (ii) triggering federated queries: the big advantage of LOD is the full compliance between the datasets in terms of interoperability.
--	---

Table 9: Linked Data workflow

An important aspect of the data archiving process is metadata management. Three different layers of metadata that are important for the archiving purpose were identified:

- Contextual information: to describe the context of production and publication (author, owner, creation and/or publication date, validity, etc.)
- Archiving metadata: to describe the conditions of archiving (physical and/or digital location of the document)
- Document description: to describe the content of the document (topics, structure, references, named entities, etc.)

So, the data archiving management can be considered as a typical data structuration process. All the metadata layers can be represented by concepts already defined in existing ontologies.

7.3 INVENTORY OF CURRENT SOLUTIONS

The approach consists of enlisting the relevant initiatives in the domain of Open Data and Linked Data as well as the methods to transform the metadata in Linked Open Data format. In order to develop this inventory, the steps below were followed:

1. Desk research and interviews: to identify the relevant initiatives and methods. For some of them, interviews or informal discussions during official events (conferences) were carried out.
2. Inventory of the Open Data and Linked Data initiatives: description of the most relevant Open Data initiatives and Linked Data initiatives for the data archiving.
3. Inventory of data structuration approaches and tools: description of the most relevant approaches, methods and tools that can be used to structure and transform the metadata of the data archiving domain in Linked Open Data format.
4. Cartography of the initiatives (magic quadrant): based on the previous inventories, the different solutions are located in the magic quadrant in function of their relation to the Open Data/Linked Data dimension and the proximity to the data archiving domain.
5. Selection of the most relevant solutions: from the cartography of solutions, the initiatives and the methodologies that are closer to the Linked Data technologies and that are related to the data archiving domain are selected to be elaborated in the recommendations of the study.

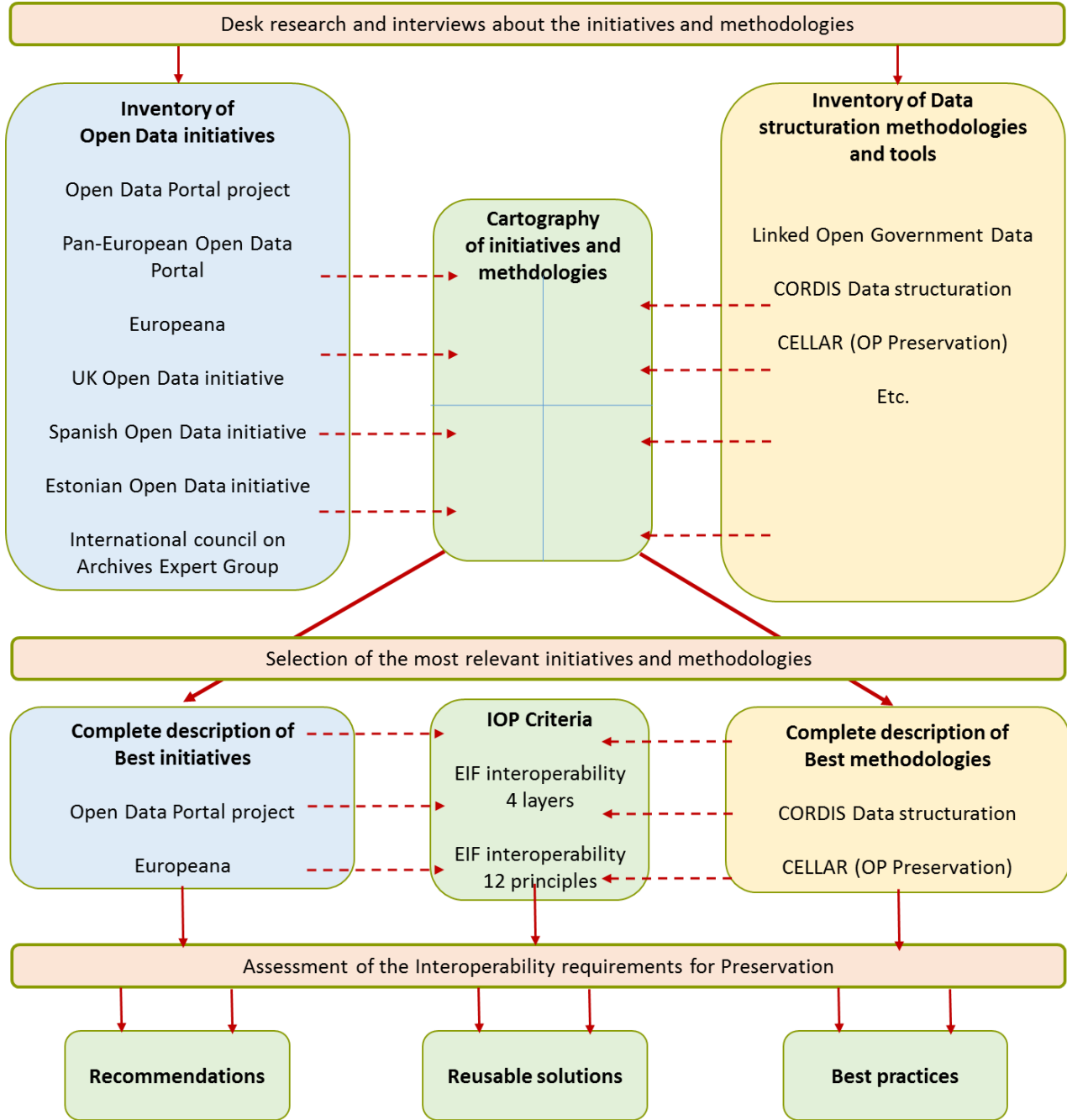


Figure 18: General approach to analyse the Open Data and Linked Data initiatives

The following table presents the template that is used to describe all the identified initiatives:

Category	Evaluation criteria	Description of the evaluation criteria
Basic description	Initiative ID	Identifier of the initiative / project that is arbitrary for this study.
	Initiative name	Official name of the initiative / project
	URL	Website
	Basic description	General description of the initiative
	Owner	Organisation or company that initiated the initiative
	Contact	Email of the person responsible for the initiative
	Type (sub-type or initiative specificity)	Project, tool, platform, etc.
	IPR	License conditions
Technical description	Technologies	List of the technical details (programming language, protocols, infrastructure, etc.)
	Features	List of functionalities
	Data models, formats	List of supported formats (input, output)
	Archival Data models, formats	List of supported formats that are specific to archival data.
	Other relevant technical aspects	Description of technical specificities of the initiative/method.
Relevant aspects for Data Archiving Management	Open Data functionalities	List of functionalities that are related to Open Data aspects
	Open Data formats	List of Open Data compliant formats
	Linked Data functionalities	List of functionalities that are related to Linked Data aspects
	Linked Data formats	List of Linked Data compliant formats
	Data transformation methods and tools	List of methods used to transform the data in other formats
	Data access / consumption	List of methods provided to access or consume the published data
	Data reusability	List of methods provided to reuse automatically the published data

Table 10: Template for the description of the initiatives

The table below lists the Open Data and Linked Data initiatives and methodologies that are later described in more details in Annex E: Linked and Open Data Initiatives. The projects and initiatives related to data structuration are also listed and are described in this annex:

Category	Name	Basic description
Open Data initiatives	Open Data Portal project of the European Union managed by the Publications Office http://data.europa.eu/euodp/en/home/	This portal provides access to Open Data published by the European institutions and bodies.
	Pan-European Open Data Portal project of the European Union managed by DG CNECT https://www.europeandataportal.eu/	This portal gathers metadata of Public Sector information available on portals across European countries.
	LOGD (Linked Open Government Data) https://joinup.ec.europa.eu/document/story-business-models-linked-open-government-data-bm4logd	Linked Open Government Data (LOGD) is a set of principles for publishing, linking and accessing open government data as a service on the Web.
	Estonian - Open Data portal http://opendata.ra.ee/	This page describes the archives of the National Archives as open records. Archival descriptions are information about the contents of records, regardless of their physical medium (paper, electronic media, film tape, etc.).
Linked Data initiatives	Europeana https://www.europeana.eu/portal/en	Europeana.eu is the EU digital platform for cultural heritage. Aggregator of data sources coming from the Archives Portal Europe Foundation.
	Archives Portal Europe Foundation http://www.archivesportaleuropefoundation.eu/	The Archives Portal Europe is an online research tool providing free 24/7 access to high-quality information held in archival institutions throughout Europe. The Archives Portal Europe is the result of the APENet project (2009-2012) and the APEx project (2012-2015), both funded by the European Commission. To guarantee the sustainability of this unique aggregation and publication platform and to ensure that contributing content to the Archives Portal Europe will always be free of charge, the national archives of the participating countries – the driving forces behind the former projects – have established the Archives Portal Europe Foundation.
	H2020 Holocaust archiving https://www.ehri-project.eu/	EHRI is a European research project (H2020) that provides online access to information about dispersed sources relating to the Holocaust

		through its online portal, and tools and methods that enable researchers and archivists to collaboratively work with such sources.
	Swiss – Federal Archives http://www.alod.ch/	The project aLOD ¹⁸ is the entry point of the project description. It is a prototype whose purpose is to serve as a test bench for evaluating the potential of Linked Data (LD) technology in the field of archives.
Data archiving initiatives	Cellar (Publication Office) https://data.europa.eu/euodp/data/datas-et/sparql-cellar-of-the-publications-office	The CELLAR is the central content and metadata repository of the Publications Office of the European Union.
	Archives Hub (UK) https://archiveshub.jisc.ac.uk/ http://data.archiveshub.ac.uk/	The Archives Hub brings together descriptions of thousands of the UK's archive collections. Representing nearly 300 institutions across the country, the Archives Hub is an effective way to discover unique and often little-known sources to support research.
	Social Archive - SNAC project http://socialarchive.iath.virginia.edu/	The SNAC project aggregates authority data in regard to persons and organisations from all sorts of heterogeneous sources, in order to facilitate the re-use of authority data. The project involves a complex methodology for merging and matching very large volumes of semi-structured data and can be an important source of inspiration on how the EU institutions may exchange authority data.
	Bibliothèque Nationale de France (BNF) http://www.bnf.fr/fr/acc/x.accueil.htm	The Bibliothèque Nationale de France has designed a new project in order to make its data more useful on the Web. It involves transforming existing data, enriching and interlinking the dataset with internal and external resources, and publishing HTML pages for browsing by users and search engines. The raw data is also available in RDF following the principles of Linked Open Data architecture.
Tools for data transformation	CIIM http://www.k-int.com/products/CIIM	A modular suite of software which sits between institutional data sources (such as collections management systems, library systems, archives and DAMS) and a range of publication end points (such as the institution's online web presence or an in-gallery display).

¹⁸ aLOD project: <http://www.alod.ch>

	<p>Silk http://silkframework.org/</p>	<p>Silk is an Open Source framework for integrating heterogeneous data sources. The primary uses cases of Silk include:</p> <ul style="list-style-type: none"> - Generating links between related data items within different Linked Data sources. - Linked Data publishers can use Silk to set RDF links from their data sources to other data sources on the Web. - Applying data transformations to structured data sources. <p>Silk is based on the Linked Data paradigm, which is built on two simple ideas: First, RDF provides an expressive data model for representing structured information. Second, RDF links are set between entities in different data sources.</p>
	<p>LinkedPipes https://etl.linkedpipes.com/</p>	<p>LinkedPipes ETL is an RDF based, lightweight Extract, Transform and Load (ETL) tool.</p>
<p>Methodologies for Semantic Data management</p>	<p>Publication Office Preservation (Cellar approach)</p>	<p>The Publications Office of the European Union is responsible for making available and disseminating the official publications and bibliographic resources produced by the institutions of the European Union. The central component of its information system is the CELLAR repository, providing semantic indexing, advanced search and data retrieval for multilingual resources.</p>
	<p>CORDIS Data structuration http://cordis.europa.eu/</p>	<p>CORDIS is the Publication Office portal to publish information about the European research projects funded by the EC. It publishes all the different scientific documents such as project description (FP7, H2020, etc.), news and scientific reports. Recently, CORDIS started a data curation process to extract the knowledge from the content, validate it and offer this content as Linked Open Data.</p>
	<p>Project Open Data https://project-open-data.cio.gov/</p>	<p>Project Open Data (POD) is a project initiated by the US government; however it is open to public for participation (Office of Management and Budget; Office of Science and Technology Policy, 201?). Its goal is to develop an OGD methodology which would support the public sector bodies in publishing data in line with the Open Data Policy (Executive Office of the President, 2013). This methodology provides basic definition of Open Data, recommendations</p>

		for implementation of the Open Data Policy, overview of the suitable software tools and recommended readings and resources. A set of case studies is also available at project website.
	<p>Guidelines on Open Government Data for Citizen Engagement http://workspace.unpan.org/sites/Internet/Documents/Guidelines%20on%20OGDCE%20May17%202013.pdf</p>	<p>The Guidelines on Open Government Data for Citizen Engagement (the Guidelines on OGDCE, or simply the Guidelines) is a practical, easy-to-understand and easy-to-use set of guidelines for everyone, especially policy-makers and technologists. They show what open government data is, why it is important and how it can be of great help for citizen engagement. It will also provide detailed advice on how to assess a country's readiness and how to successfully design, implement, evaluate, and sustain an OGD initiative for citizen engagement in development management.</p>
	<p>TopBraid Enterprise Data Governance https://www.topquadrant.com/products/topbraid-enterprise-data-governance/</p>	<p>An agile data governance (modular) solution for today's dynamic enterprises. Using standards-based graph technologies, TopBraid EDG supports integrated data governance across the ever- growing numbers and types of data assets and governance needs.</p>
	<p>GraphDB https://ontotext.com/products/graphdb/</p>	<p>GraphDB™ Enterprise is an enterprise-level triple store which has proven to scale in production environments where simultaneous loading, querying, and inferring of graph data statements occur in real time. It features a new data transformation functionality that makes it easier to leverage legacy data, establish interlinked enterprise master data and ultimately build 360-degree data view.</p>

Table 11: Inventory of solutions

The complete inventory is provided in Annex E: Linked and Open Data Initiatives while the next section provides a cartography of those initiatives.

7.3.1 Cartography of the Linked Open Data initiatives

Figure 19 shows the cartography of the initiatives and methodologies according to their focus areas and dimensions in order to highlight the most relevant initiatives to consider in the specific context of Linked Open Data for archives. These initiatives are in the Magic Quadrant in the top right corner of the figure.

For information, the generic tools (e.g. GraphDB, TopBraid) are not considered in the landscape because they are not initiatives.

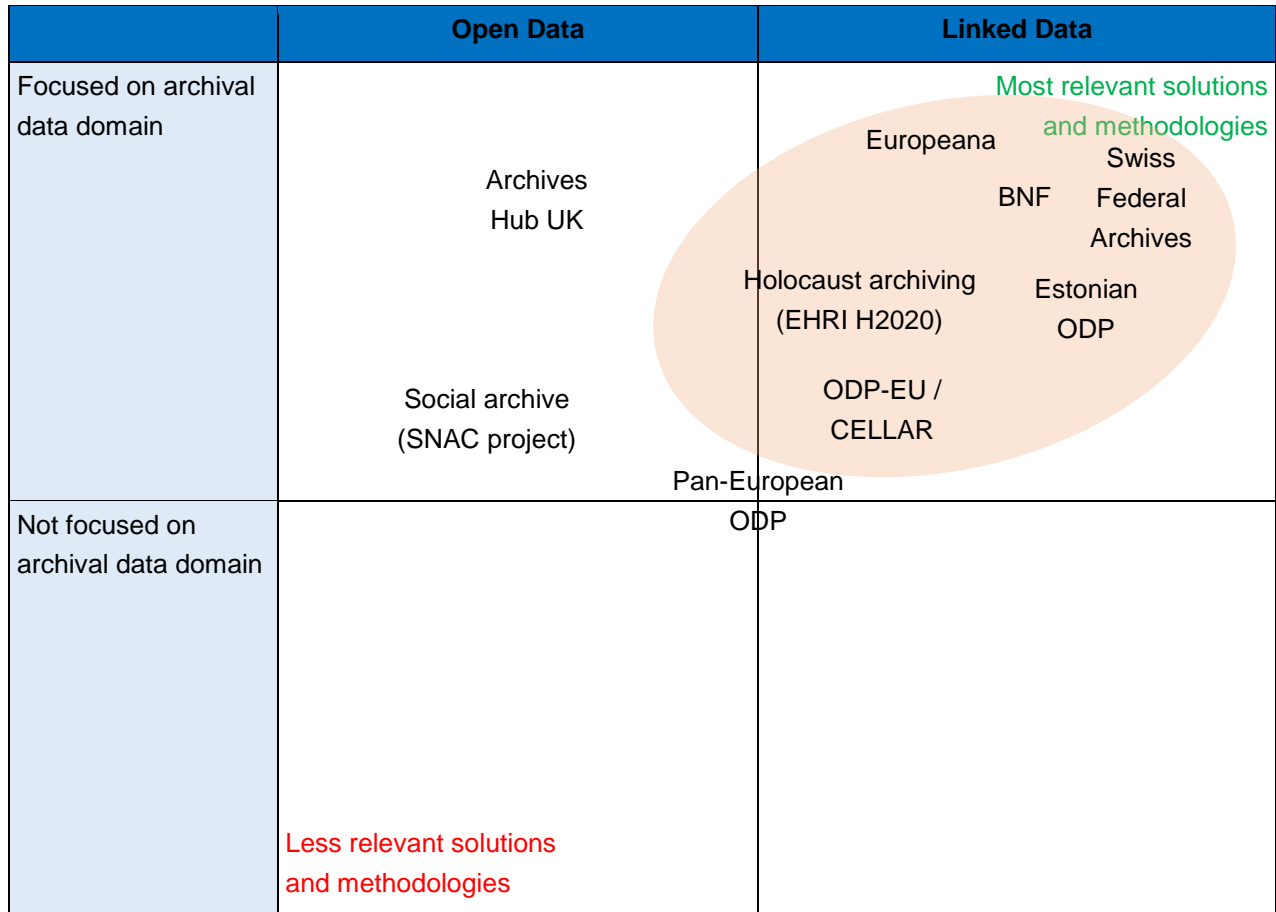


Figure 19: Magic Quadrant for Linked Open Data initiatives

We can distinguish three different areas of the Magic Quadrant:

- Open Data + No Digital Archive domain = “**Not relevant initiatives**”

In this area, we identified some initiatives that are focused on sharing data in Open Data format. They provide useful information about the general approach to share data as Open Data (especially the format of data, e.g. in Pan-European Open Data Portal). But as they are too generic approaches, they do not provide any specific input for the digital archiving domain.

- Open Data + Digital Archive domain = “**Relevant initiatives**”

In this area, the initiatives are also focused on Open Data aspects but at least they take into account the digital archive domain. For example, the “Archives Hub in UK” and the “Archives Portal Europe Foundation” are initiatives that use specific data formats to publish digital

archives as Open Data. For some of them they also use techniques to integrate and transform data in the Open Data formats.

- Linked Data + Digital Archive domain = “**Very relevant initiatives**”

Finally, the last area groups the initiatives that produce and publish digital archives in Linked Open Data which is the highest level of data quality. Some of these initiatives are still research project (EHRI project) or are in an experimental phase (Swiss Federal Archives).

These initiatives are the most relevant ones because they use:

- Shared data models,
- Techniques to extract, transform and load (ETL) data in Linked Data format,
- Centralised approach to integrate data with standardised format: data providers use predefined data models and local tools to upload their data in the central repository. Platforms act as aggregators of unified data,
- Data are published in Linked Open Data,
- Data are shared and reused through a SPARQL endpoint.

7.3.2 Potential for Linked Open data at the Archives of the EC

This section aims to identify existing data and metadata from the EC archives which could potentially be published as Linked Open Data. Within the document "Overview of Metadata and Digital Objects Managed by the Historical Archives Service of the European Commission" (HAS, 2017), a high-level overview is given of the information architecture which has been developed across decades to manage both physical, digitised and digital born archival holdings and their metadata. As new needs emerged and technology evolved, the archives service added features which were either integrated in the core database or developed as extra modules and services. As within many archives across the world, this situation has resulted in a complex information architecture, where it is challenging to ensure data normalisation and keep reference and authority data consistent across data sets. Figure 20 allows to understand how the key applications interact with each other.

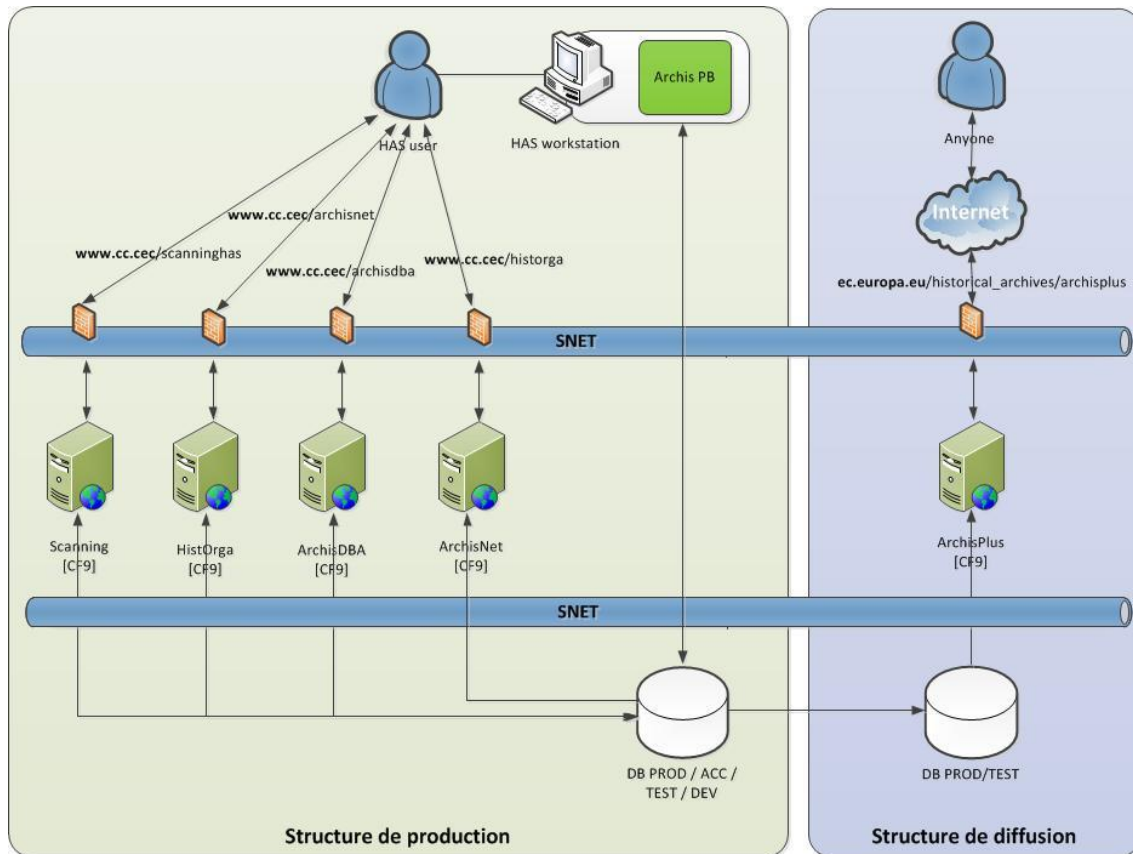


Figure 20: High-level overview of data architecture, from EC internal documentation

ArchisPB is the name of the archives management system and the core application used for the creation of finding aids and to manage most of the other business processes. Throughout the years, a very large number of tables have been created to manage a wide variety of data and metadata. Before migrating to a new application, a thorough analysis should be performed in order to see whether the number of tables could potentially be rationalised, without impacting the data integrity and performance of the application.

Within this diversity of data sources, certain subsets should be analysed in more detail in order to evaluate the potential for re-use. First of all, the so-called transmission forms and the basic file lists¹⁹ play a fundamental role within the communication between a specific service or Directorate-General and the EC archives, as both allow to document the transfer of both physical and electronic files. Within their metadata, the title of the files holds the most semantics and could potentially be of value to researchers. Within the basic file list, the field “Analysis” also gives quite detailed information. An example of a file title would be “Inspections concernant la sécurité nucléaire” and the field “Analysis” adds the following information “Note sur la récupération de 3 kilogrammes d'uranium enrichi par les autorités italiennes. Notes sur l'exportation des matières nucléaires vers l'Irak.” No structured information on the subject matter is given. As within most archival descriptive practices, little to no metadata are available on the level of individual documents. Figure 21 gives an overview of the metadata encoded on the level of the file.

¹⁹ Documents created by the archives producer and documenting the accession process.

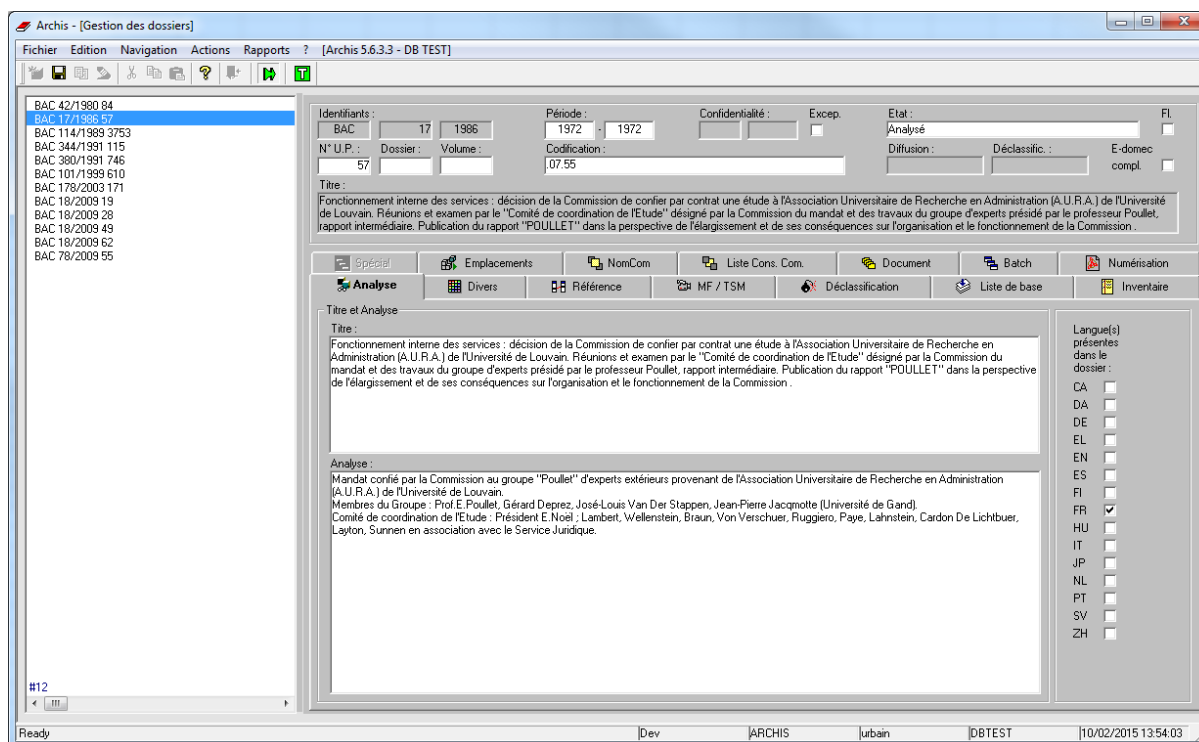


Figure 21: Metadata encoded on file level, from EC internal documentation

Files are opened to the public via a web interface called “ArchisPlus”. The software gives the possibility to encode metadata according to the ISAD(G) rules, but this has only been done for a very limited number of files and remains a work in progress. A big step forward was taken in the context of the “digiCOM project”, which allowed digitising the content of a large volume of files and providing end-users with digital surrogates of the files and their documents. Minimal metadata encoded in EAD are available for this particular collection.

A specific module has been developed within ArchisPB to describe archival producers according to the ISAAR standard. However, it should be noted that the current application is not able to export these metadata into the EAC format. It is mainly here on the level of the identification of archival producers that the service could make investments in an open data policy which would be highly relevant both for internal use, for partner institutions but also for external entities and researchers. This point is also relevant in the context of the HistOrga project, which is presented below.

A multitude of other metadata are available in other databases maintained by HAS, which mainly relate to practical elements allowing the management of the physical entities. Within the variety of these metadata, an underexploited element might be the references to NOMCOM, which is the filing plan used for the records management processes of the EC. End-users might find it valuable to explore the headings of a filing plan in more direct and user-friendly manner to navigate throughout archival holdings. Another interesting opportunity for open data would lie in an initiative to present in a user-friendly way both the filing plan specifically developed by the archives service and to try to map it to the current version of the corporate filing plan (NOMCOM) as it is used within the different Directorate-General. NOMCOM illustrates in a good way how metadata could be looked at as an archival corpus itself, as the changes made in NOMCOM also help to understand how the activities and functions of the EC have evolved throughout the years, when facing new challenges such as increasing environmental concerns or more recently, the immigration or financial sector crisis.

In 2017, a digital preservation infrastructure based on the Preservica platform has been implemented (a-REP). At the end of the Administrative Retention Period (ARP) as defined by the EC's retention schedule, files from the records management system (called HAN) destined for permanent preservation are transferred to a-REP... This step forward of course broadens the horizon significantly to deliver new services towards end-users. The key difference, in particular in regards to an open data strategy, is the fact that a-REP, in a parallel to manage the digital preservation of digitised files, ingests digital born material. Also, integration is foreseen with a SIP creator tool to ingest content from shared drives. Ingested data will be packaged in the following structure: file > sub-file (if any) > document > translation (if any) > digital object. The system then also uses this structure to create exports into a DIP. Figure 22 gives an overview of the archival filing plan used to provide access to these files. The same 30 year rule of accessibility that was established for the paper archives of the EU institutions applies to these files, meaning that the files which are currently ingested will not be available in the immediate future, but it would be important to take the potential of this evolution into account when reflecting about an open data strategy, when issuing for example identifiers for files and entities (persons or organisational structures) related to the files.

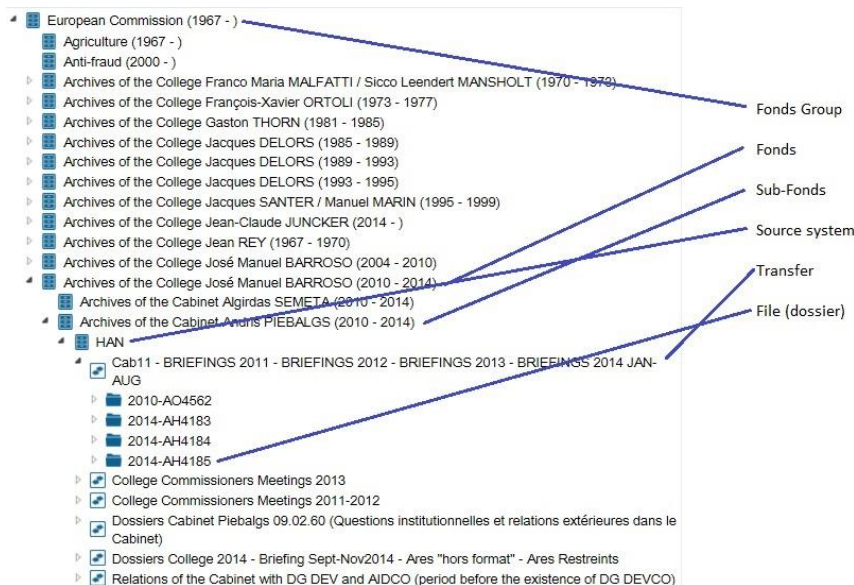


Figure 22: Overview of the archival filing plan

However, two existing projects of the archives service are of strategic immediate importance to underline the potential of open data: the digitisation of the so-called COM files and documents (digiCOM projects mentioned above) and the HistOrga project. Since a few years, a specific type of official documents, COM documents, have been digitised by an external contractor in collaboration with the Publications Office. The output of these projects is currently stored on the storage infrastructure of the Publications Office. At a later stage the content should be transferred to a-REP. It was decided to store a particular part of the metadata also in the ArchisPB in order to be able to search in the descriptions and to perform the light EAD export as mentioned previously. The collection of COM documents is in a rather particular way composed. Bound together in volumes, all documents can be found organised per number and per year. For each particular number, many documents may be available: different preliminary versions, final version, language versions, administrative documents, notes etc. All PDF files are currently available in CELLAR, the public repository of the Publications Office. The PDF files can be reached by means of permanent URI, which is used to integrate the content in the public website of the HAS and the public

website of the HAEU. However, the actual usage of this unique collection, already available as open data, remains rather limited. Within the section of the usage of machine-learning techniques to explore full-text, this report has developed some examples of how more value from these files could be derived by extracting semantics in an automated manner.

Apart from the COM files, the HistOrga project holds a lot of potential in the context of open data. This application aims at gathering basic information regarding the organisational entities (e.g. general directions, directions, units, services, departments, units, divisions etc.) that have existed at the EC since the very beginning (1952 and the creation of the High Authority of the ECSC) until nowadays by keeping track of their historical evolution through time. For each organisational entity the following information is recorded:

- Basic metadata
- Hierarchical position
- Historical relationships through time with other organisational entities
- Responsible people with their names, function and activity dates
- Missions
- Source documents where the information has been found

It allows users the downloading of reports regarding the organisation charts and their historical relationships. The archivists use HistOrga not only to answer questions about the institution's history but mostly to keep track of the archives creator and the context in which the archives were created. Currently, the database is only available for internal use but anyone studying the history and evolution of the EC would be interested in having access to these data. Within the archival descriptions of partner institutions such as the European Council or the Parliament, identifiers issued by HistOrga could be of potential great use in order to deliver more precise information on who and which entity exactly from the EC intervened in a particular case. It should be noted that since the beginning of the years 2000 DG HR and DIGIT are maintaining good recording of organisational entities, persons, jobs etc. in a data warehouse called COMREF that can be reused by applications using people and services related information. Neither Sysper2 (application for personal file management and organisation charts management) nor COMREF are managing the historical evolution of the organisational entities. For that reason, a new project has been launched aiming at integrating HistOrga with COMREF in order to collect relevant data in real time, enrich them and making them reusable in the HAS context.

This first analysis has identified various data sources which would be both of value internally, to partner institutions and to end-users if exposed as open data. Based on the discussions held during the November workshop, the development of a common approach to create and share authority files for person names in the EAC standard seems very relevant. In parallel, the existing work on the HistOrga should also be valorised and given priority within an open data agenda of the archives service. More in depth work on these data sets could take place in the form of case-studies during the next phase of the project.

8 CONCLUSIONS

The study focused on identifying existing standards and business processes in archives management and data exchange domains; and explored supporting IT tools and services.

During the execution of the study several actions were carried out:

- Task-1.1 - Discovery of archiving standards: the identification and categorisation of commonly used and accepted national and international standards related to archives management and archival data exchange. The output of this activity is summarised in D01: Inventory of standards.
- Task-1.2 - Discovery of business processes: the identification of common business processes – and their corresponding tasks and activities- in archives management. The output of this activity is summarised in D02: Inventory of Business Processes.
- Task-1.3 - Discovery of archiving IT tools and services: the identification and analysis of standard-based IT tools and services used to manage archives. The output of this activity is summarised in D03: Inventory of IT Tools.
- Task-1.4 - Design of the evaluation framework: creation of a comparative assessment tool that could easily be used by public administrations to assess the best tools for their needs. It should enable the archival community to define minimal requirements for archives management systems and archival data exchange. The output of this activity is summarised in D04: Assessment Tool.
- Task-2.1 - Discovery of Linked and Open Data initiatives and methodologies: this task consists in the identification and the analysis of the Open Data and Linked Data initiatives and the tools used to transform the data in Linked Open Data. The output of this activity is summarised in D06: Inventory of existing open data initiatives in the domain of archives and records management.

8.1 INVENTORY OF BUSINESS PROCESSES (D02)

The inventory of business processes provides an overview of the main business processes identified for archives management and preservation. The information on these business processes has been retrieved from both European and non-European sources, including archival institutions and other related entities.

The first conclusion with regards to the business processes in the field of contemporary archives management is that the business processes need to cover management of both digital archives and non-digital archives. Non-digital archives can be split further into two categories depending on whether the related metadata are available in electronic format. Nowadays most records managed by archival institutions belong to the group of non-digital archives with digital metadata. Nonetheless, given the evolution of technology and the nature of archiving itself, all archival institutions have to manage both digital and non-digital archives simultaneously. This creates a complex situation because while some processes are similar, others differ and require a distinct implementation.

The second conclusion is that while there is a general set of processes being used (i.e. appraisal, description, etc.), there is no existing general agreement or common framework, which is being applied to all the institutions. This means that every archival institution develops its processes in an ad-hoc way, resulting in a multitude of models adapted to their own particular context and structure. Furthermore, this situation hinders the opportunities resulting from a more structured approach towards data exchange between archival institutions. Therefore, a set of predefined functional blocks and business processes have been proposed to create a basic framework for the roles and activities carried out by any archival institution. The workshop held on 24th of November 2017 with several European Institutions has confirmed the potential benefits of better alignment on several aspects and their willingness to work towards a common solution.

Another important conclusion is that, although the processes to manage traditional archives are mature and there is a sort of consensus within the archivists' community about their meaning and behaviour, digital archiving is a new practice that is currently object of studies and research. The latter entails that the processes related to digital archives are yet under definition and, therefore, there is a lack of reference documentation that can be used to attain a consensus on their meaning and behaviour. The E-ARK project can be considered as a first step towards the standardisation of the business processes needed to manage digital archives. Hence, the deliverables produced by E-ARK can be considered as guidelines to further consolidate these processes.

Regarding digital preservation, which has become one of the main concerns of archival institutions due to the appearance of (born-)digital assets, the OAIS framework has been found to be the broadest one, including the development of standards, and therefore the most suitable for its implementation.

Finally, archives need to put the focus on adding value for all their users, internal and external. As such, two areas are of particular interest:

- Improving the interactions between the archives management system and the preservation system to ensure better synchronisation between those two core systems at pre-ingest and ingest time;
- Facilitating the publication of information (metadata) on various platforms, considering both traditional approaches – such as the existing web portals – and new technologies – such as Linked Open Data – that could allow an exponential dissemination.

8.2 INVENTORY OF STANDARDS (D01)

The conclusions of the study are that there is a large set of generally adopted national and international standards defined for archival description. For other traditional archival processes such as appraisal, selection, elimination, ingest, storage and management of physical space, control of access and security, no widely adopted standard at international or European level exists at this moment. This gap is covered by each institution through the development of its own guidelines, custom-made procedures and the use of the methodologies and definitions from standards belonging to other areas, like records management, long term preservation, or recently new as IT security.

With regards to archival description, another relevant fact is the need of normalisation and the progressive use of standards developed and implemented traditionally in the scope of libraries and museums for improving archival description and information discovery. Examples of these standards would be the ones used for bibliographic description and cataloguing; or value standards as vocabulary list of authorities, subjects in the form of controlled lists, thesauri or ontologies.

The four principles of Linked Open Data are (1) the use of URIs to identify things; (2) the use of HTTP URIs to find content; (3) the use of structured data using controlled vocabularies and dataset definitions expressed in open standards to provide useful information about what a name identifies and (4) the use of HTTP-URI based names when publishing data on the Web. The usage of Linked Open Data and the semantic web entail the usage of controlled vocabulary and demonstrates that it will enrich archival description by making it interoperable, granular and more accessible to users.

With regards to new regulations such as the GDPR, although it adds new constraints on organisations, it does not fundamentally affect archival institutions as data protection and access control were already one of their main priorities. Taking into account digital archives will merely result into an adaptation of the existing rules for non-digital archives. Furthermore the new regulation explicitly defines exceptions for the purpose of archiving.

Finally, the raise of new techniques for massive data and content management, such as content analysis and automatic classification based on methodologies imported from Artificial Intelligence (AI) and

statistics, might accelerate the acquisition and processing of the archives. The business processes “Ingest” and “Description / Indexation / Enrichment” as well as “Accession” to some extent might benefit from those new technologies. Although no recognised standards can be identified yet in this promising field, there are already methodologies under development that could be adopted by archives.

8.3 INVENTORY OF IT TOOLS (D03)

The tools included into the Inventory of IT Tools cover the following domains:

- Archives Management Systems
- Preservation Systems
- Connectivity Tools
- Library and Collection Management Software
- Content Analysis and Automatic Classification
- Secure Deletion and Content Masking

The current inventory focuses on the research about IT tools that give support to business processes identified during the study, providing an overview of a large selection of tools in each category. For Archives management systems, a more in-depth description of their key features and architecture is also provided.

The study comes to the following conclusions on the key topics:

- There is a variety of open and proprietary archives management and preservation systems available on the market which, when implemented together, can manage all internal archival processes performed by archival institutions for digital archives, with the exception of the areas of data exchange and access by Consumers. It should be noted that most of these IT tools are composed of a platform of interconnected components that were developed or added over time. As such, there is no one-size-fits-all single software.
- Archives management software supports most of the main archival description metadata as well as transformations from content standards (ISAD(G), DACS, and RAD) to structural standards (EAD, MODS, MARC21, and DC).
- Some IT tools are more oriented towards serving as collection management tools and offer additional capabilities for the management of multiple types of digital archives (traditional documents, images, and audio and videos files).
- Regarding the support for Linked Data and Linked Open Data, the group of tools of library and collection management software offers an interesting degree of compliance and/or support. However, it decreases significantly for classical archives management system and preservation system.
- A list of new tools (coming from the GDPR environment) in the scope of automatic classification, content analysis and security such as content masking tools and secure deletion are included in the inventory. However, no direct connection exists between them and archives management and preservation systems at the moment of the study.

8.4 INVENTORY OF LINKED DATA AND LINKED OPEN DATA INITIATIVES (D06)

Although the hype surrounding Linked Data has catalysed a renewed debate on the validity of fundamental conceptual pillars of the archival community, such as the “*Respect du fonds*” and the hierarchical nature of archival descriptions, which both have been under discussion ever since post-modern thought entered the scientific archival literature throughout the 1980s and 1990s, Linked Data can coexist with traditional approach to archives management rather than replace it. Linked Data offers interesting benefits in regard to publishing and providing access to archival holdings.

This study cannot give a complete overview of these conceptual discussions, but the essential discussion points and references to the most relevant literature should help bridge the gap between archivists and IT specialists, in finding a solution. Some of the conceptual discussions can have tremendous operational consequences. It is essential that both archivists and IT experts can discuss the advantages and drawbacks of using either a hierarchical or a graph-based data model for the creation of finding aids. The success of a renewed strategy of the EC and EU archives depends to a large extent on having a better dialogue between archivists and IT experts, using the same language and understanding how the specific set of methods and tools from the LOD paradigm impacts the archival profession, and how it can benefit from them.

The key outcome of this deliverable is not to over-focus on the discussion, but to focus on the expression “a little semantics go a long way”. Its main conclusion is that the complexity of setting up an archiving data management platform is related to the quality of the data to be managed in terms of context, structure and content. However, as was underlined during the November 2017 workshop, the current archival descriptions contain too much non-structured content. Refining existing standards to make the format and content of fields more precise would be of great importance. In parallel, methods and tools should be promoted to automatically enrich both existing metadata and full-text content from digitised holdings with identifiers from outside sources of information, such as EUROVOC, GEONAMES or VIAF for example. In the context of the ISA² programme, special attention should be given to support in all EU languages to determine whether a set of reference data must be made available in all languages or only a subset of it. In order to automatically create metadata on the level of the content, EU institutions could start to experiment with unsupervised machine learning methods such as Topic Modelling and Word Embedding to make sense of very large collections of digitised content which cannot be described manually.

Different best practices about the application of interoperability in the domain of digital archives have been identified and structured around three main components: the definition of a governance plan, a central infrastructure that can integrate different data sources in a common repository, and the functionalities required to support linked open data.

- **Governance plan:** To ensure the interoperability of the entire platform and the right evolution of the metadata, it is highly recommended to adopt a governance plan that covers the four layers of interoperability defined by the European Interoperability Framework:
 - Legal layer, focused on the identification and sharing of similar policies and principles.
 - Organisational layer: focused on the definition of the governance model, which will define the different bodies to put in place, its members, roles and responsibilities.
 - Semantic layer: with the use of conceptual models – including standards, vocabularies, taxonomies, etc. and its relations - that will help to represent and share a common understanding of the data.
 - Technical layer, focused on the implementation of interoperable APIs to consume and share Linked Open Data as well as on the implementation of the exchange protocols defined by the semantic layer.

- **Centralised infrastructure:** Ontologies – i.e. shared data models – has become a successful option to deal with the management of heterogeneous data spread in different databases, to build an integrated solution that hinges on the federation of an ecosystem of SPARQL endpoints to retrieve data in parallel. Ontologies hence allow the integration of different systems and organisations with a minor impact on the metadata production processes.

The archiving data management platform could therefore consume, enrich and publish metadata already produced and standardised in the form, for example, of Linked Open Data. Europeana is a clear example that demonstrates the technical feasibility and shows the benefits of this approach.

- **The implementation of LOD** involves four main processes: data aggregation, data transformation, data access and search, and data consumption and reuse. The following best practices by process have been identified during the study:
 - Data aggregation: data should be collected/aggregated in an open and standard format
 - Data transformation: data should be validated and converted in Linked Data format to reach the highest level of interoperability
 - Data structuration and validation: data can be processed through a knowledge extraction pipeline in which new entities/metadata will be identified and disambiguated. To satisfy the qualitative aspects, a human validation can be required.
 - Data access and search: data should be published in the Linked Open Data format to facilitate the access and the search of the information

As the study has revealed, uptake of Linked Open Data within an archival context is still at an early stage. A Proof of Concept will help to explore possibilities, constraints and limitations. It will consist of several steps whereby it will initially focus on the preparation and publication of Open Data for a rapid return on investment.

8.5 COMPARATIVE ASSESSMENT TOOL (D04)

Based on the findings of D01, D02 and D03, a comparative assessment tool has been developed by the everis team.

The main purpose of this comparative assessment tool is allowing the benchmarking between the selected set of IT tools taking into account the business processes coverage as well as the supported standards and additional technical requirements.

Therefore, the main input for creating this assessment tool are the outcome of the D01, D02 and D03 deliverables, whose inventories can be found in Annexes A, B, and C. Notwithstanding those criteria, additional elements have been taken into consideration, such high level technical requirements and interoperability requirements. In order to develop this tool, several existing assessments tools in this area were a source of inspiration, such as the EIRA's Interoperability Quick Assessment Excel tool and CAMSS self-assessment tools.

The comparative assessment tool's criteria are organised in the following sections:

- Questions related to the support of business process and standards implemented by the analysed IT tool: the coverage of high level requirements detailing the business processes, the standards supported by the tool, the degree of customisation, etc.
- General questions related to the interoperability aspects
- Specific questions related to exchange of data supported by the tool;

- And finally questions about general technical capabilities.

The assessment tool can be found in Annex D: Self-Assessment Tool.

9 RECOMMENDATIONS FOR THE PROOF OF CONCEPT

After the identification, analysis and understanding of the existing standards, tools and processes for managing archives while ensuring their interoperability, the following set of recommendations are proposed in order to set up a future Proof of Concept (PoC) to be implemented in the next phase of this ISA² action.

This PoC has multiple objectives considering the different stakeholders interested in its outcome:

- With regards to the specific objectives of HAS, it should contribute to the preparation of the actual implementation of an archives management system foreseen for 2019 by:
 - confirming its feasibility;
 - making sure that the integration with the current preservation system (Preservica) is possible at a reasonable cost;
 - identifying gaps where custom development might be needed;
 - identifying the main risks and defining mitigation actions;
 - assessing the scalability of the proposed solutions.
- With regards to the broader needs and expectations of other archival institutions, it should:
 - provide guidance on how to approach the implementation of an archives management system;
 - identify needs and opportunities for more standardisation and definition of interoperability layers to facilitate the interactions between the main components of the solution.

To this end, an architecture model for the whole solution has been defined based on a set of interrelated layers and components that support the standards and business processes identified in the previous phase of the project. This architecture is presented in the section 9.1 Architecture of the overall solution. Then the approach, scope and roadmap of the PoC are outlined in section 9.2 Proposition for a Proof of Concept.

The PoC should be evaluated on a set of criteria, including (but not limited to):

- Ability to perform the tasks outlined in the business models,
- Technical feasibility, with a focus on the following elements:
 - interface between the archives management system and the preservation system,
 - publication of content as (Linked) Open Data,
 - establishment of data exchange mechanism with other archival institutions.
- Effective support for the required standards and best practices,
- Learning curve and support needs from the supplier (or the community of users in case of open source software),
- Real costs for implementing and operating the solution (compared to the initial estimated budget),
- Ease of use and satisfaction of users (archival institution, producers, consumers and interoperability (IO) stakeholders).

During the PoC, feedback should be continually collected and analysed so that the solution can be improved and refined.

9.1 ARCHITECTURE OF THE OVERALL SOLUTION

The proposed architecture is a multi-layer modular model outlining how the business processes defined in Chapter 4 Business Processes and Annex A: Business Processes could be implemented. As such, it follows the breakdown into the functional modules and allows the mapping of these functional modules and their business processes with the proposed services and the underlying applications implementing them. It should be noted that the overall solution combines the archives management system, the preservation system and other supporting applications, i.e. for data exchange and authority lists management.

The architecture is composed of the following layers:

- **Business processes:** this corresponds to the functional model defined in section 4.4 Functional Model;
- **High level Applications Services:** this layer identifies the key services required to deliver the business processes. These services can be provided by interfacing multiple applications through the layer “Applications Services”;
- **Applications Services:** this layer displays the services delivered by the underlying applications;
- **Applications:** this layer presents the applications envisaged in the context of HAS
 - **Preservica:** existing preservation system
 - **Archives management system:** new system based on an existing archives management software to support the business processes within the archival institution;
 - **Reference data database:** application that consists of a database of reference data tables and a simple user interface. It shall also implement an interface with the interoperable module supporting the management of authority lists;
 - **Web portal:** application provided to the consumers for search, request and consultation of the archives managed and published by the archival institution;
 - **Interoperable module:** standalone application that provides interfaces to manage authority lists and exchange data in a structured way with other stakeholders.

Figure 23 below outlines the relationships between the different layers. Furthermore it allows the identification of the key elements to be implemented and tested in the PoC. This aspect is developed in the next section.

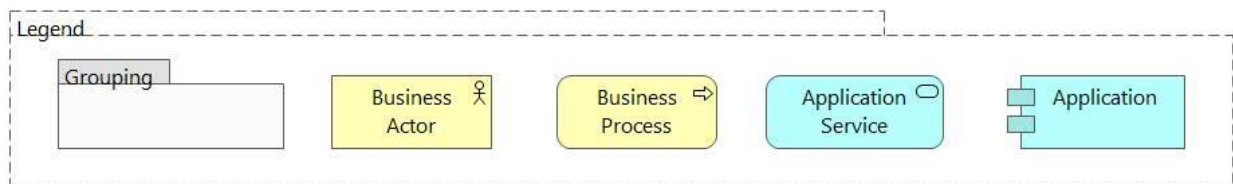


Table 12: Archimate standard elements legend

Legend		
Business Processes		Business process applicable to digital and non-digital archives
		Business process only applicable to digital archives
		Business process only applicable to non-digital archives

Table 13: Archimate specific archiving elements legend

Study on Standard-Based Archival Data Management, Exchange and Publication Final Report

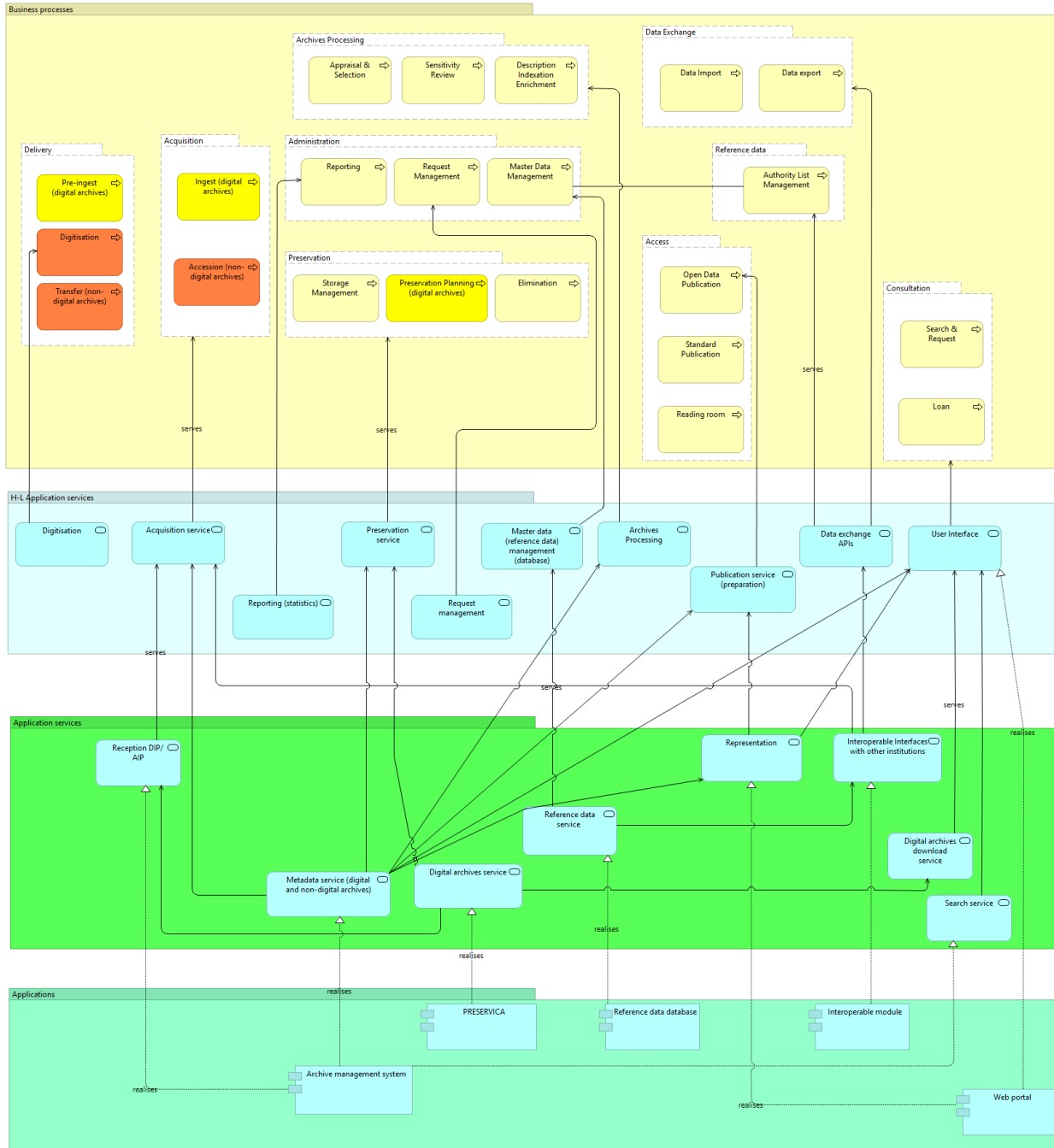


Figure 23: Technical Architecture

Data Exchange

The Data Exchange features are brand-new elements where only basic elements are provided out-of-the-box by archives management systems. Therefore, customisation and development will be required. Figure 24 provides an overview of the services to put in place in order to exchange data. It is instantiated with the Historical Archives of the European Union as the counterpart to HAS but this approach is applicable to any other archival institution as long as standardised interfaces (API) are defined. A phase for functional and technical specifications is foreseen at the start of the PoC to define the way those systems will communicate with each other.

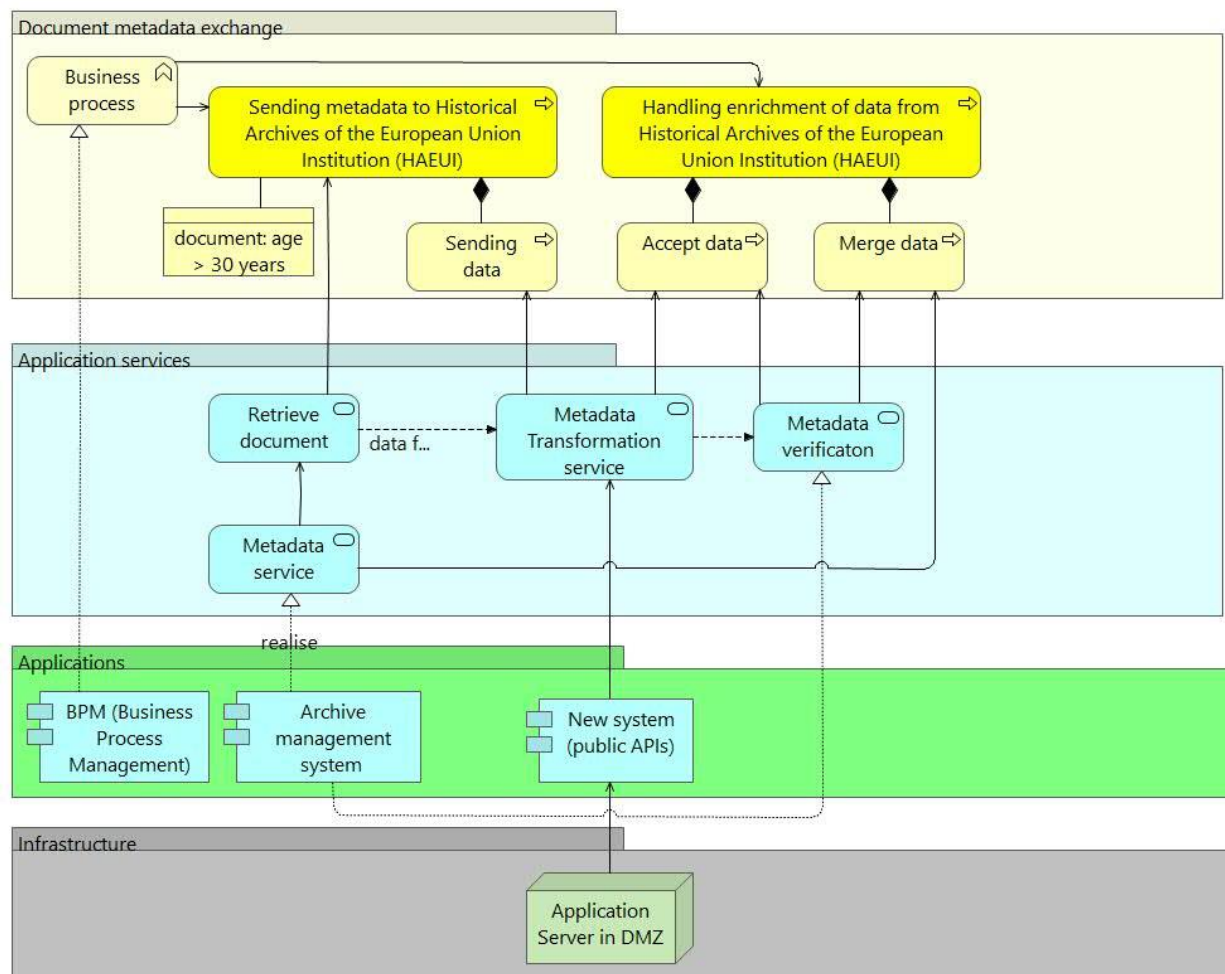


Figure 24: Technical Architecture

General technical recommendations

In addition to the functional requirements expressed by the functional model, additional general requirements should be considered such as costs and license model, scalability and performance, support and training, perspective of evolution and long term availability of the software, and migration roadmap from the current archives management system to the new one

- **Technical support:** services in charge of archives management should assess the type of support, response time and customer service. This may influence the selection of the solution towards an open source or a commercial software.
- **Data migration:** services in charge of archives management should investigate about the migration and ingestion tools provided by the IT Tool itself and its degree of customisation according to their content characteristics. Furthermore they should check if enhanced migration such as validation and error treatment functionalities are provided.
- **Long-term evolution:** the capacity of the supplier to provide improvements and new functionalities and customise the IT tool to the specific needs of the customer.

The involvement, advice and support from the HAS IT team and from the Directorate-General for Informatics of the European Commission (DIGIT) is highly advisable when analysing the requirements about:

- Storage capabilities,
- Network capabilities,
- Hosting capabilities in the cloud or on premise,
- Performance, response time, and scalability,
- Adherence to technical standards, best practices and methodology of the Commission,
- Respect of security processes, standards and rules
- Disaster recovery, backup, and fault tolerance requirements.

9.2 PROPOSITION FOR A PROOF OF CONCEPT

9.2.1 Approach

The Proof of Concept (PoC) will cover all the important internal processes of archives management and implement key aspects of the exchange of information with other stakeholders. The objective of the PoC consists in deploying an archives management system together with a preservation system in order to assess the technical feasibility of the whole solution, test the interactions between the two systems, and demonstrate the effectiveness of the interoperability principles. It is proposed to implement the PoC in three steps:

- Step 1: Deploy and configure an archives management system and a preservation system to demonstrate the support of the internal business processes (acquisition, archives processing, preservation and) and the interactions between the two systems
- Step 2: Implement and configure a shared tool for the management, synchronisation and use of authority lists
- Step 3: Implement the processes to enable and test data exchange with other stakeholders

The three figures below illustrate this approach to highlight the focus of each step and indicate the key areas where specific aspects of the whole process should be tested. These elements are documented as Use Cases later in this section and form the proposed scope of the PoC. The suggested software stacks are presented hereunder and the roadmap is presented in section 9.2.2.

Study on Standard-Based Archival Data Management, Exchange and Publication
Final Report

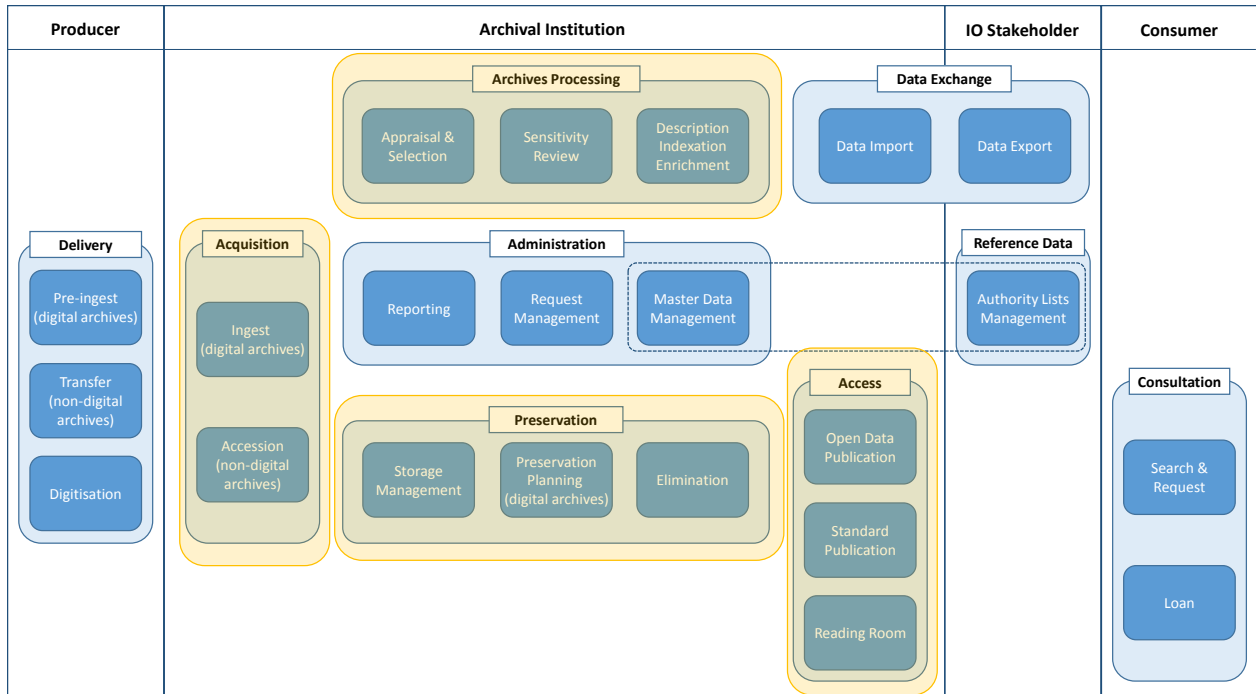


Figure 25: Step 1 – Internal Processes

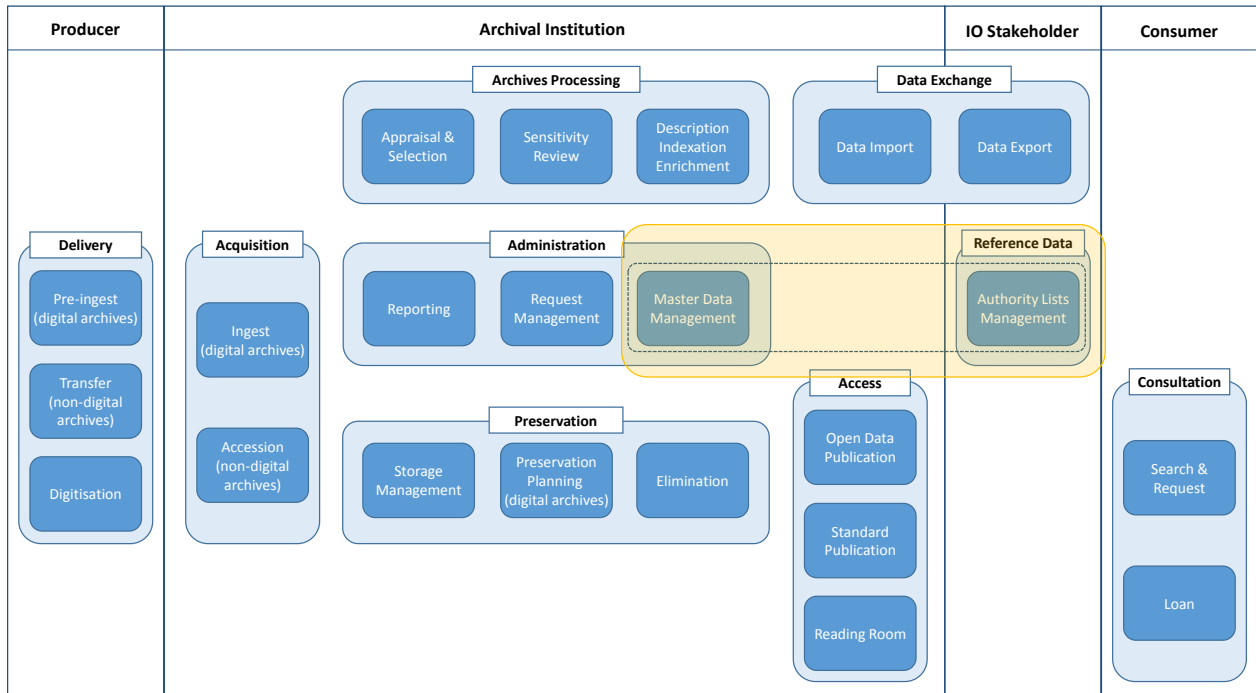


Figure 26: Step 2 – Authority Lists Management

Study on Standard-Based Archival Data Management, Exchange and Publication
Final Report

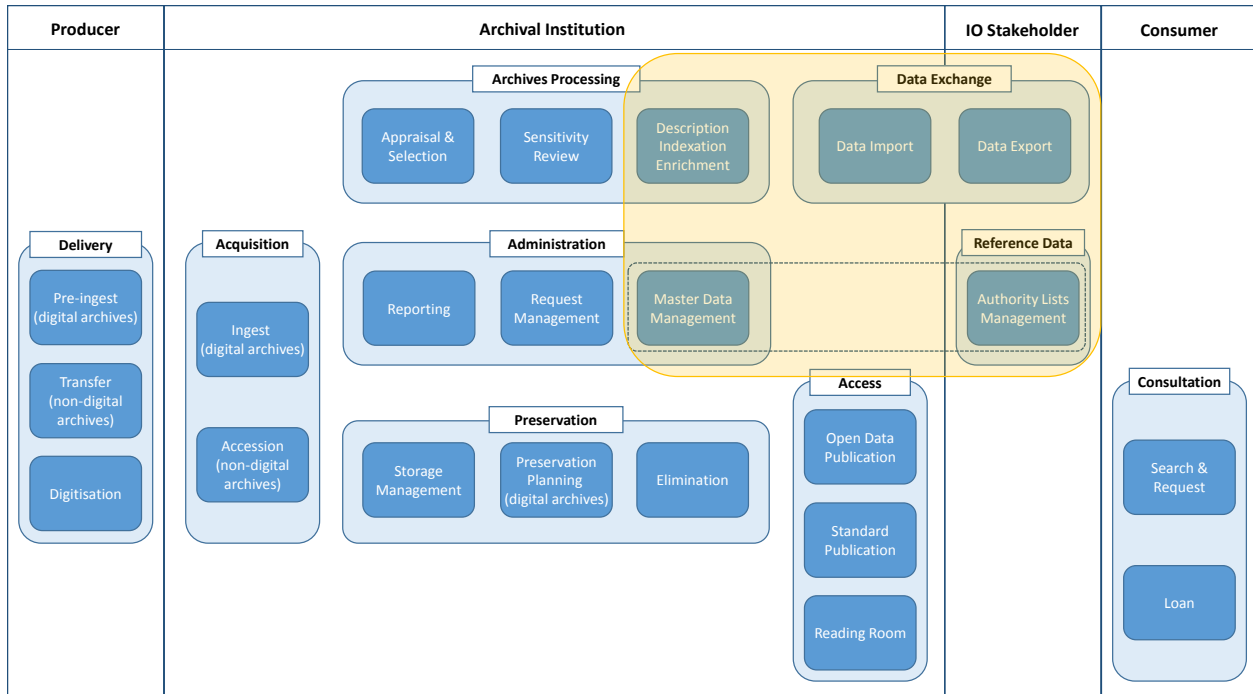


Figure 27: Step 3 – Data Exchange

Proposed Use Cases

The following table provides the main recommended Use Cases. They aim to cover the handling of both digital and non-digital archives. The focus is on key processes but additional features such as loan management or physical asset location management could be implemented depending on the time and budget available in the next phase.

Use Case	Description
Acquisition	<ul style="list-style-type: none"> Reception of metadata coming from a transfer request within the archives management system, including acceptance and rejection of the transfer request Second review process (integration with Preservica)
Archives Processing	<ul style="list-style-type: none"> Archival description using ISAD/ISAAR standards: <ul style="list-style-type: none"> For digitised or digital-born archives, implement and test the processes to automatically acquire the metadata, transform the content and validate the structure Implement the processes to encode the metadata for paper-based archives
Migration / data extraction from legacy system	<ul style="list-style-type: none"> Data extraction from Archis to the new archives management systems, addressing the main challenges (location of transfer, data export related to physical inventory) allowing to test scalability and performance.
Interaction with Preservation Systems	<ul style="list-style-type: none"> Identify the data that needs to be synchronised between the archives management system and the preservation system Implement or configure the interfaces between both systems aiming for a standardised approach independent from the selected tools

	<ul style="list-style-type: none"> • Carry out digital [and physical] transfers towards the long-term storage
Publication as Linked Open Data	<ul style="list-style-type: none"> • Implement any required transformation towards a LOD format: entity extraction, entity linking, and knowledge validation • Store archival metadata in a semantic repository • Verify the ease of metadata and semantic repository maintenance • Publish the LOD archives through a SPARQL endpoint • Implement a user interface to test the publication as LOD
Authority List Management	<ul style="list-style-type: none"> • Define the process for creating, managing and sharing new authority lists, allowing co-creation and co-management of those lists • Implement the interface with external sources: EuroVoc, GeoNames
Data Exchange with other institutions	<ul style="list-style-type: none"> • Implement two-way transfer and data synchronisation processes <ul style="list-style-type: none"> ◦ Transfer metadata (or give access to the metadata repository in the other institutions) ◦ Enrich metadata with additional information provided by HAEU • Identify and define required data transformation to guarantee digital continuity (e.g. semantic coherence) between the different organisations involved

For the first step, some other institutions may be involved in different ways:

- Other European and national archives services may be requested to provide existing archives that will be ingested by the archives management system.
- The digitisation process may be outsourced to another stakeholders, such as the Publications Office.

The second and third steps may also involve other stakeholders as this was discussed in the workshop on 24th November 2017.

The potential stakeholders and their involvement should be discussed in the early stage of the PoC.

Proposed Software Stacks

The selection of the software stack relies on an assessment tool that has been produced during the execution of the project. The assessment tool can be customised to emulate the context of a given institution in terms of standards to be supported and processes to be implemented. It also takes into account if specific components are compulsory, optional or not-needed.

Hence, the context of the HAS has been configured in the assessment tool producing the results shown in Figure 28 and Figure 29. The assessment tool concludes that three software stacks have a similar score in this specific context:

- Stack 1, combining Preservica (commercial) with ArchivesSpace (open source);
- Stack 2, a full open source solution combining Archivematica and AtoM as the main tools. The combination of the two tools covers a broad framework of business process identified in this study at high level of percentage.
- Stack 3: a mixed stack, combining RODA (open source) and Archeevo (commercial). As the tools selected in Stack 2, this combination of tools of the same vendor covers a broad framework of business processes identified and is an alternative of analysis of the previous option

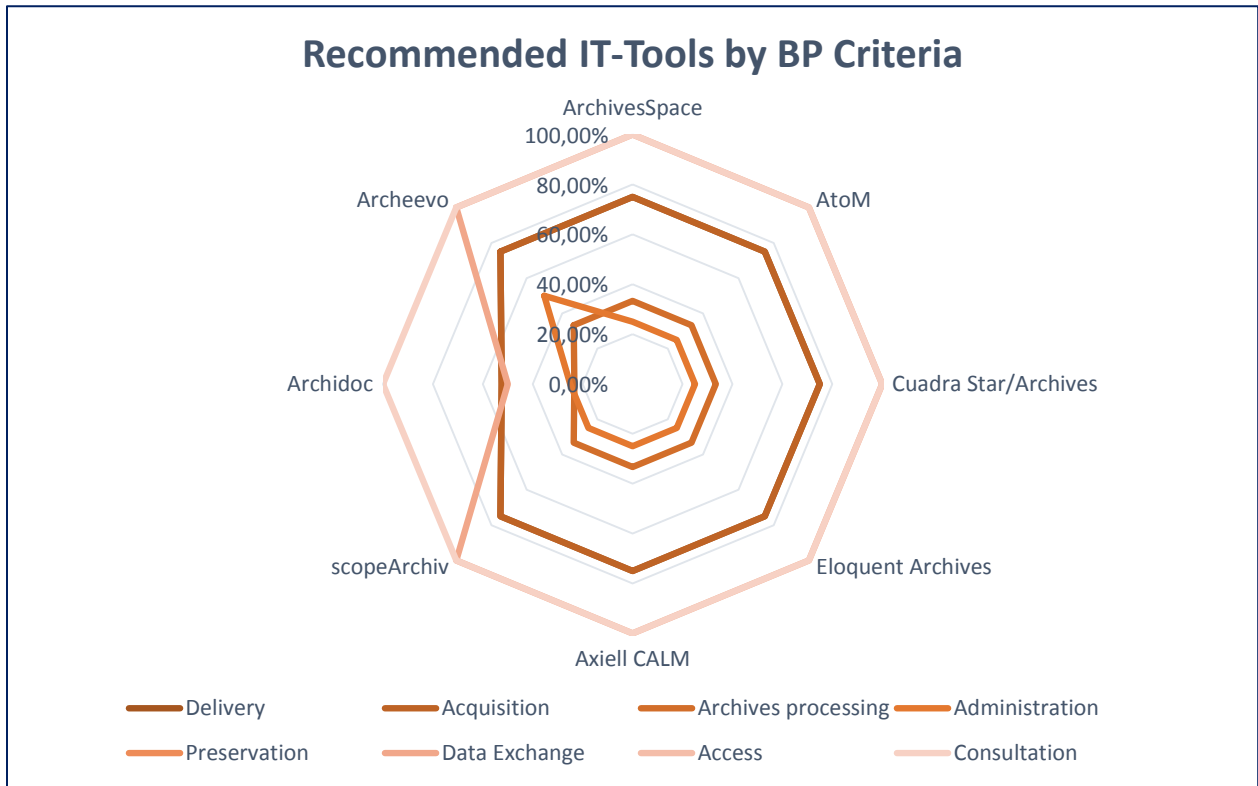


Figure 28: Ranking of archives management tools by business process criteria

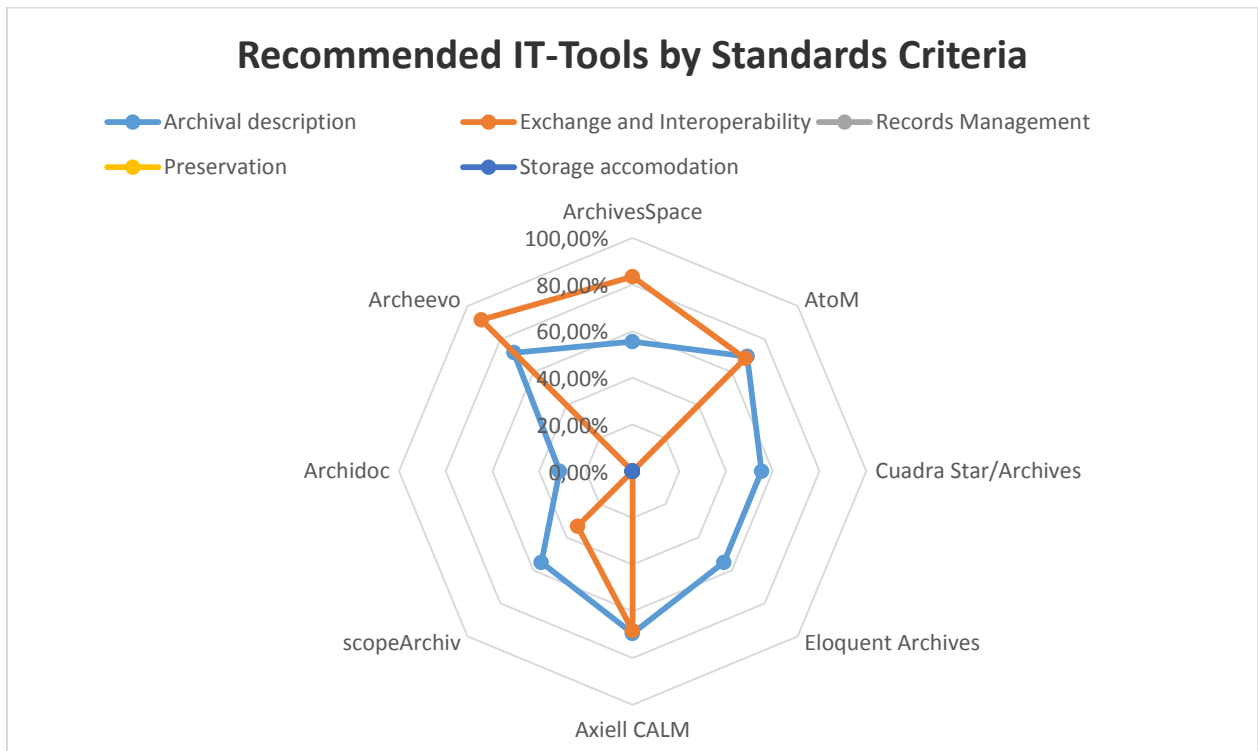
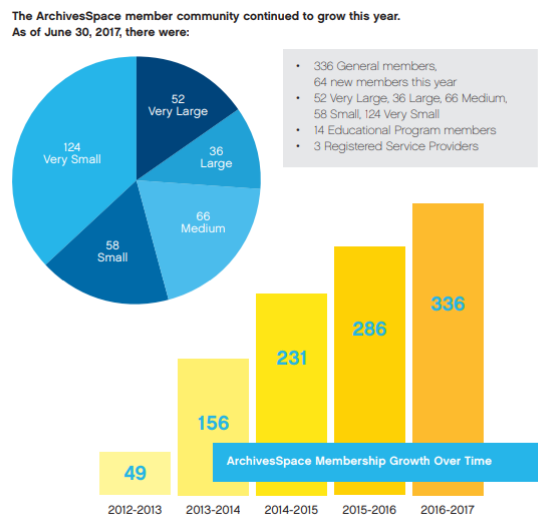


Figure 29: Ranking of archives management tools by standards criteria

Id	Category	Product/Tool Name	Functional requirements coverage	Non-Functional requirements coverage	Average coverage
1	AMS	ArchivesSpace	35.40%	25.64%	15.26%
2	AMS	AtoM	35.34%	20.86%	14.05%
3	AMS	Cuadra Star/Archives	30.39%	22.48%	13.22%
4	AMS	Eloquent Archives	30.39%	14.61%	11.25%
5	AMS	Axiell CALM	35.34%	20.86%	14.05%
6	AMS	scopeArchiv	32.39%	23.21%	13.90%
7	AMS	Archidoc	25.22%	20.58%	11.45%
8	AMS	Archeevo	38.38%	25.15%	15.88%

The technical stack is based on trustful and well-known solutions within the archivists' community, what strengthens the results of the assessment tool. These tools are the following:

- Preservica is a commercial solution for preservation systems with a very broad user base, including the European Commission who relies on Preservica to preserve its electronic archives.
- ArchivesSpace is a well-known open source solution to manage archives that has already reached a broad user community;



⁹ ArchivesSpace: Annual Report 2017

Figure 30: Report on users' community of Archives Space²⁰

- Archivematica is an open source solution for preservation systems with many clients around the world;

²⁰ <http://archivesspace.org/about/reports-presentations>



Figure 31: Clients of Archivematica²¹

- AtoM is the other leading open source solution for archives management. It integrates with Archivematica and it has features to support LOD publication;
- Archeevo, the best scoring commercial software, is a product provided by Keep Solutions for archives management, a Portuguese IT company that has been involved in the E-ARK project²².
- RODA is also a product developed by Keep Solutions for preservation management and integrates with Archeevo.

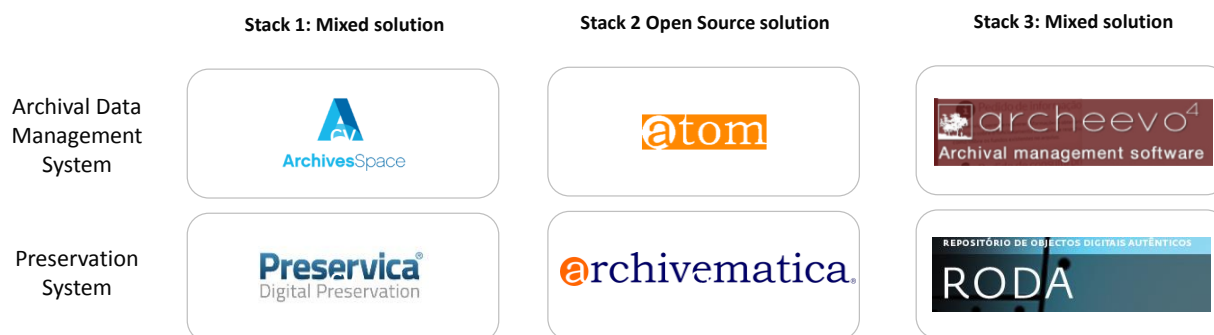


Figure 32: Software stack

9.2.2 Roadmap

In order to implement the PoC, preliminary activities have to be performed:

²¹ <https://www.artefactual.com/clients/>

²² The recommendations for the PoC focus on open source solutions; hence only one commercial solution has been selected. However it has to be noted that Axiell Calm is the second best scoring commercial solution and could also be tested especially in the specific context of the European Commission where Preservica is already used.

1. The data model to support the needs of the HAS should be defined and designed based on the standards and formats identified during the study. This data model will be the basis of the conceptual framework (network of ontologies that represents all the dimensions of the data archiving domain) that will be used to represent and store the information that will be managed by the different tools.;
2. Based on the study related to the IT tools and data transformation, the functional requirements will be defined at general level to identify the required functionalities for the development of LOD platform for data archiving.

Once this analysis phase has been completed, the implementation phase can begin. It should be noted that Step 1 and 2 can be performed in parallel while Step 3 depends on the completion of the two others steps.

The table below provides only an indication of the timeline. It is purely indicative at this stage.

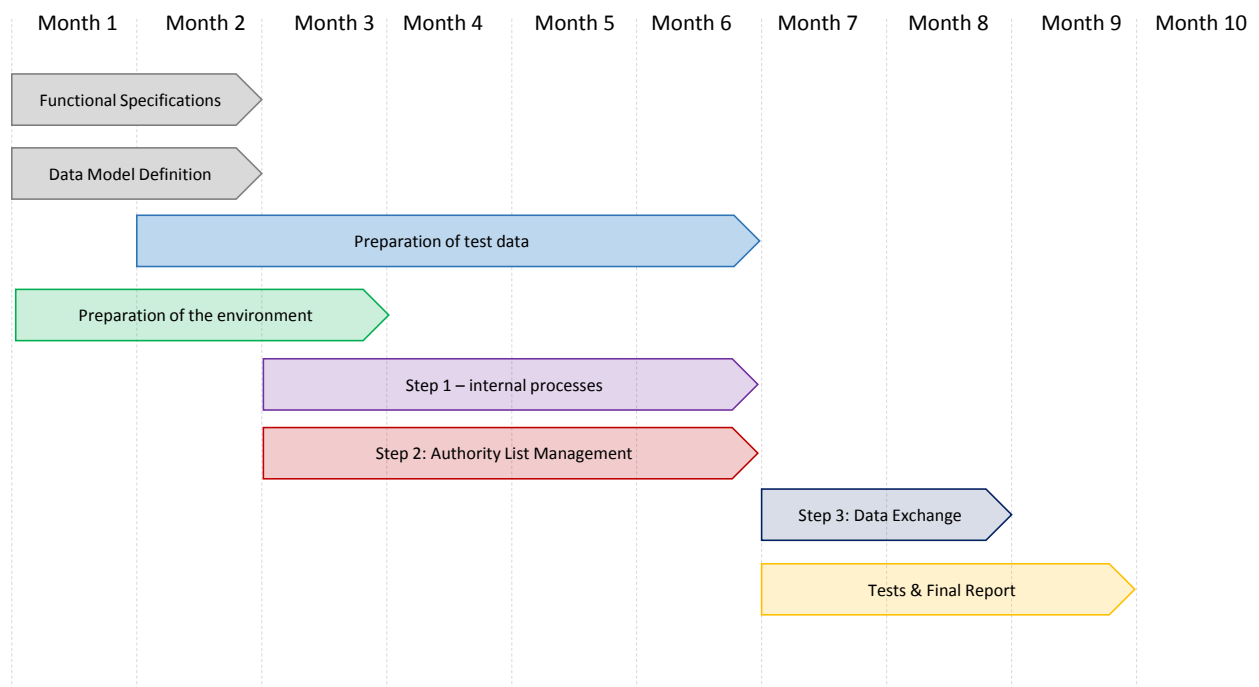


Figure 33: PoC Roadmap

List of PoC activities

The table below describes the activities that have to be performed during the PoC.

Tasks	Description
Phase 1: Analysis and specification	
Functional Specifications	Use Cases will be elaborated to define test scenarios. Functional specifications and technical design will be detailed with the aim of defining generic, reusable and standardised interfaces that will constitute the basic components for a future production-ready system.
Data Model Definition	A generic data model will be defined based on the identified standards of data

	<p>model and the existing data models used by the stakeholders. This data model could be implemented as an ontology in OWL format to enable LOD and test publication as LOD.</p>
Preparation of test data	<p>Test data will be gathered and prepared to feed the different test scenarios. This phase should not be underestimated and will require coordination between HAS and the implementation team.</p>
Phase 2: Development	
Preparation of the environment	<p>The different software stacks will be installed and their connections to the repositories containing the archives will be configured.</p>
Step 1: Internal Processes	<p>The interactions processes between archives management and preservation systems will be implemented and tested for the different software stacks. This is a key point to ensure interoperability for between different types of systems. With regards to the “Access” process, the digital archives could be published as LOD through a SPARQL endpoint. Examples of semantic queries and data reusability will be provided.</p>
Step 2: Authority List Management	<p>Two separate streams are envisaged:</p> <ul style="list-style-type: none"> • Use of existing external sources • Co-creation of new common authority lists
Step 3: Data Exchange	<p>This includes setting up proper channels, defining interfaces and implementing any necessary data transformation to ensure that new information or updates provided by another organisation can be processed and stored.</p>

Table 14: List of PoC activities

10 ANNEX A: BUSINESS PROCESSES

This section of the document provides descriptions of all the processes and activities that conform the Functional Model, including information about the stakeholders in charge of each of the processes and their corresponding inputs and outputs.

10.1 DELIVERY

10.1.1 Pre-ingest (digital archives)

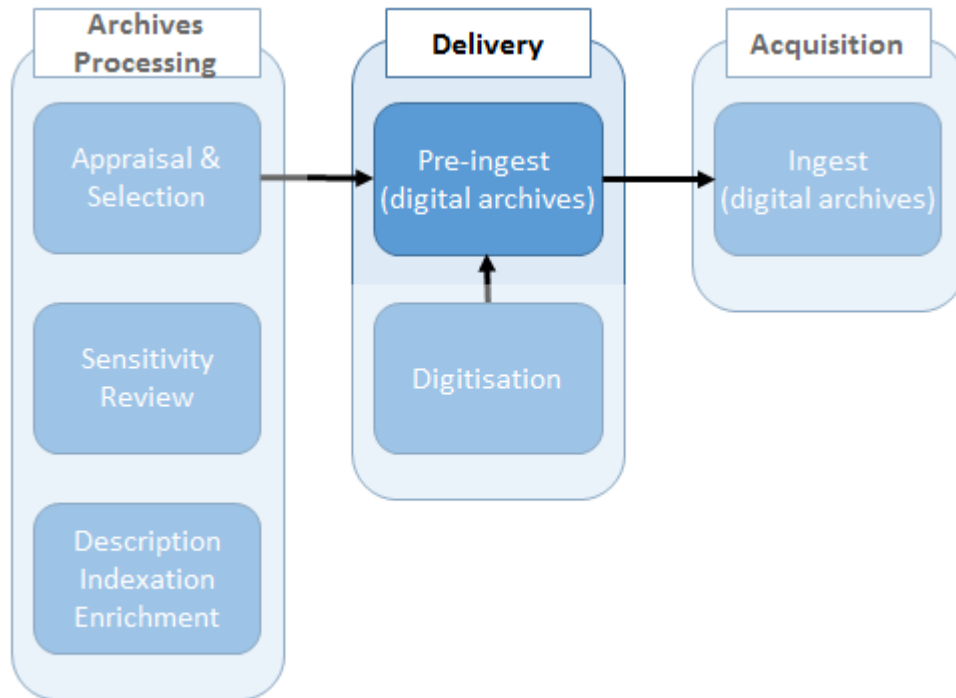


Figure 34: Pre-Ingest

Pre-ingest (digital archives)	
Description	Responsible
<p>This process only affects digital archives.</p> <p>Process of data negotiation/acquisition that includes a check of the rights and access criteria, licensing and data submission.</p> <p>Following the OAIS model, the pre-ingest process covers the producer's and archivist's activities of creating the Submission Information Packages (SIP). According to the OAIS task partitioning, all the activities related to data selection, preparation and extraction from the producer data sources belong to the Pre-Ingest (E-ARK Project, 2017).</p>	<ul style="list-style-type: none"> ▪ Producer

<p>The elaboration of the SIP will consist in formatting a dataset according to the rules defined between the producer and the Archival Institution and may include various steps and activities:</p> <ul style="list-style-type: none"> - Archives selection. Depending on the archival practices within organisations or at state or national level, a set of activities that are described under "Archives Processing" may already be carried out by the Producer before the archives are delivered. - Metadata mapping. The mapping of automatically and manually added system metadata with selected metadata schemas such as METS or ISAD(G) in order to have sufficient elements to describe the archives and their producer. - The descriptive and technical metadata of the transfer itself. - Additional format validation. - A final quality check before the delivery of the SIP to the Archival Institution. - Confirmation of the proper reception of the SIP "draft" in the pre-ingest repository and of the integrity of its content. 	
<p>Inputs</p>	
<ul style="list-style-type: none"> ▪ Digital archives generated by the Producer 	
<p>Outputs</p>	
<ul style="list-style-type: none"> ▪ SIP (digital records that will – after having processed them – be ingested and preserved by the Archival Institution) 	

Table 15: Pre-Ingest (digital archives)

10.1.2 Transfer (non-digital archives)

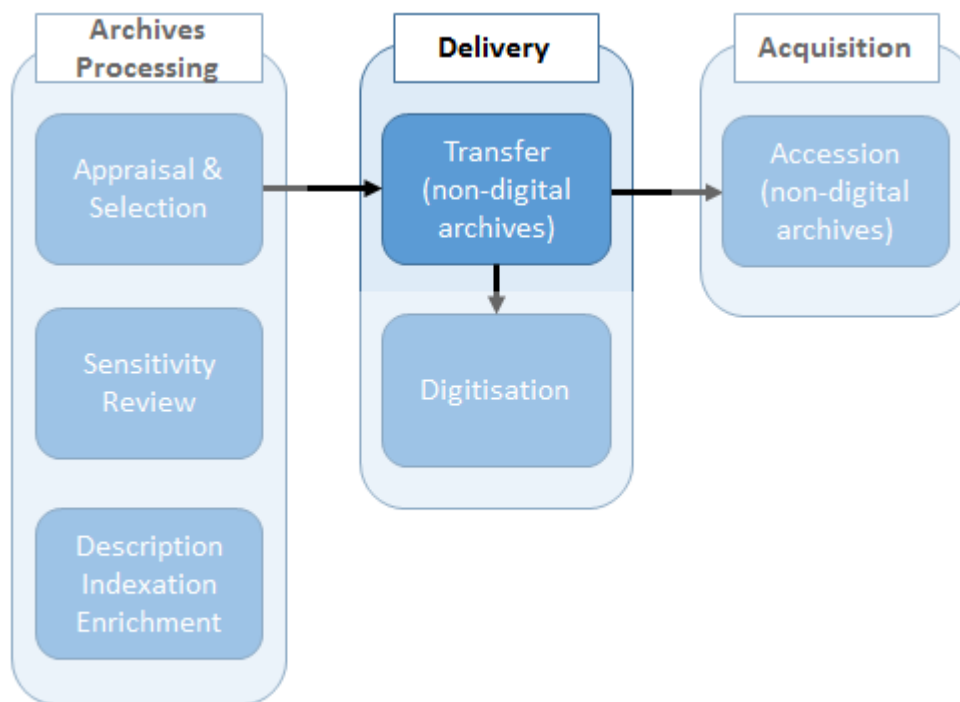


Figure 35: Transfer (non-digital records)

Transfer (non-digital archives)	
Description	Responsible
<p>This process only affects non-digital archives.</p> <p>The process of moving records as part of their scheduled disposition, especially from an office to a records centre, or from a records centre to an archival institution (Society of American Archivists, 2017).</p> <p>Once the documents have been properly selected and evaluated they are set to be sent to the Archival Institution.</p> <p>Once the archives have been properly selected by the producer, they are prepared to be sent to the Archival Institution.</p> <p>This process includes the following activities:</p> <ul style="list-style-type: none"> • Archives selection • The description of the archives to be transferred and their producers. • The descriptive and technical metadata of the transfer itself. • The validation of the transfer, that consists of checking that all the parts of the transfer have taken place according to the planning. 	<ul style="list-style-type: none"> ▪ Producer

<ul style="list-style-type: none"> The actual transfer of the non-digital archives. 	
Inputs	
<ul style="list-style-type: none"> Non-digital archives generated by the Producer 	
Outputs	
<ul style="list-style-type: none"> Non-digital archives that will – after having processed them – be accessioned and preserved by the Archival Institution. 	

Table 16: Transfer (non-digital archives)

10.1.3 Digitisation

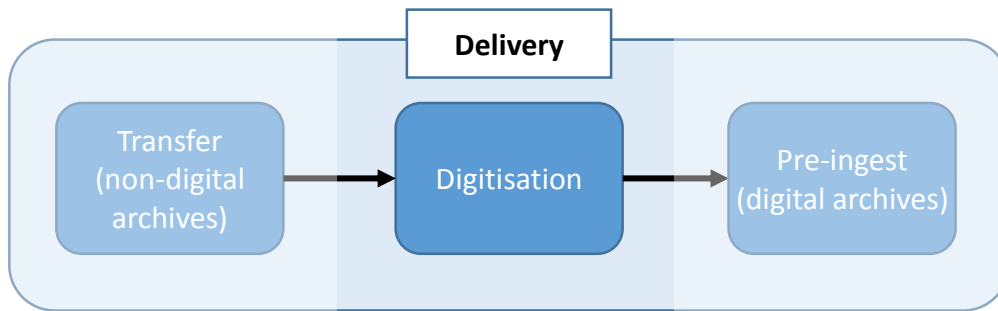


Figure 36: Digitisation

Digitisation	
Description	Responsible
<p>This process consists of using non-digital archives as a basis to produce a digital reproduction. This results into digitised archives whose characteristics will be identical to the original non-digital version with the exception of the format.</p> <p>This process could be carried out by various roles (an external contractor, the Producer or the Archival Institution) and several activities may be involved like metadata management, interactions with the AMS, digital object conversions and quality control. The activity is in any case always finalised by a pre-ingest and ingest activity.</p>	<ul style="list-style-type: none"> Producer²³
Inputs	
<ul style="list-style-type: none"> Non-digital archives 	
Outputs	
<ul style="list-style-type: none"> Digital archives 	

Table 17: Digitisation (non-digital archives)

²³ In case the Archival Institution is in charge of the digitisation, it is considered that it plays the role of Producer.

10.2 ACQUISITION

10.2.1 Ingest (digital archives)

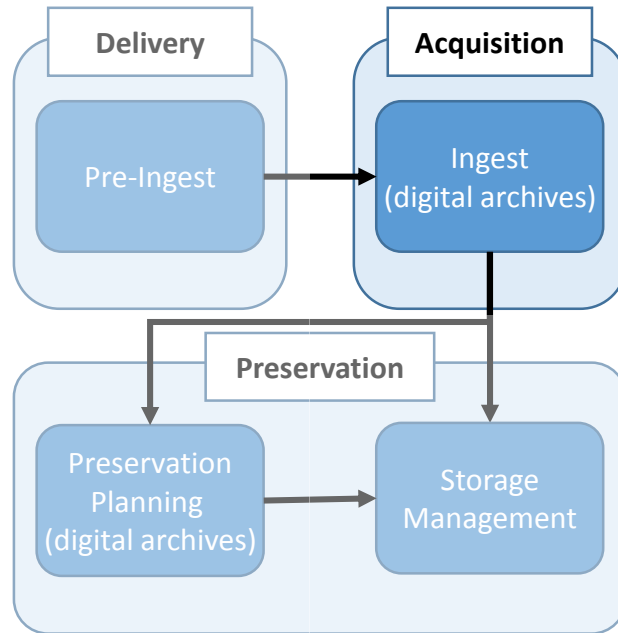


Figure 37: Ingest (digital archives)

Ingest	
Description	Responsible
<p>This process is responsible for the incorporation of the new archives into the permanent repository and it involves every task necessary to this end.</p> <p>This encompasses the services and functionalities needed to accept Submission Information Packages (SIPs) from various sources and prepare the contents for storage and management within the archive.</p> <p>Following the OAIS framework, Ingest functions include:</p> <ul style="list-style-type: none"> • The reception of the SIPs. • The performance of a quality assurance on SIPs. • The generation of an Archival Information Package (AIP) compliant with the Archival Institution's data formatting standards. • The extraction of descriptive information from the SIPs for its later inclusion in the Archival Institution's AMS. • The coordination of the updates regarding Archival Storage and Metadata Management. <p>Additional Ingest functions may be:</p>	<ul style="list-style-type: none"> ▪ Archival Institution

<ul style="list-style-type: none"> • The transformation of submitted digital objects to normalised formats validated for long-term preservation. • The addition of descriptive metadata. • The storage of the original bit-stream of the digital object together with the normalised version. 	
Inputs	
▪ SIPs	
Outputs	
▪ AIPs	

Table 18: Ingest

10.2.2 Accession (non-digital archives)

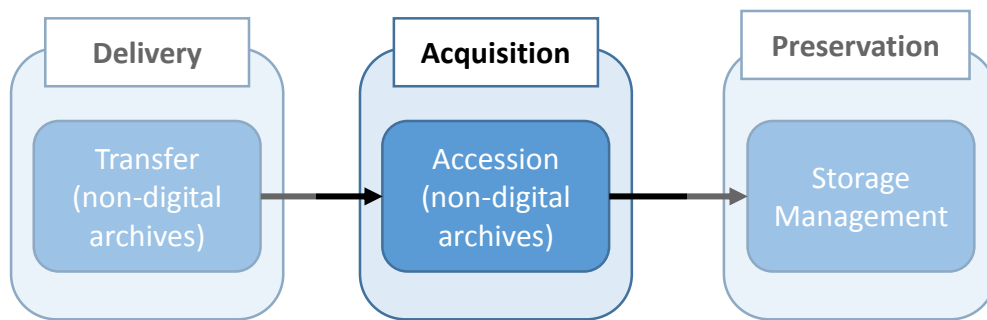


Figure 38: Accession (non-digital archives)

Accession	
Description	Responsible
<p>To take legal and physical custody of a group of records or other materials and to formally document their receipt (Society of American Archivists, 2017). The means through which the materials can be acquired are by gift, request, purchase, transfer, retention schedule, or statute.</p> <p>It is observed that the terms “accession” and “acquisition” cannot be used in the same way. Whether as nouns they are synonyms, as verbs “accession” implies a larger set of actions regarding the physical and intellectual control over the materials by entering brief information about these materials in a register, database, or other log of the repository's holdings.</p> <p>“Having made sure that new material has been legally transferred to your archives, the next, and vitally important, step is to gain control over it. This initial process is called “accessioning” which records information about origins, creator, contents, format and extent in such a way that documents cannot</p>	<ul style="list-style-type: none"> ▪ Archival Institution

<p>become intermingled with other materials held by the archives. Accessioning provides the basic level of physical and intellectual control over incoming material” (Pederson, 1987).</p> <p>Accessioning consists of a sequence of different activities. These include:</p> <p><u>Transfer request:</u></p> <p>The transfer request will be the document that will trigger the transfer of the records from the Producer to the Archival Institution.</p> <p><u>Finding aid:</u></p> <p>The finding aid is the element that will contain all the metadata regarding the transference or the records from the Producer to the Archive.</p> <p>A single transfer request and finding aid shall be associated to each series of records transferred.</p> <p>When reception by the Archival Institution, this will carry out a series of processes in order to ensure that the process has been carried out correctly. These activities are:</p> <ul style="list-style-type: none"> - Validation and encoding of transfer metadata: this activity consists on performing a check of the finding aid to verify its completeness regarding transfer metadata and the encoding of this metadata according to the Archival Institution standards. - Validation and encoding of basic inventories: this activity consists in checking that all the records that were supposed to be transferred to the Archival Institution have effectively been so. - Validation of extra information: this activity consists in checking if any information about the records is available on the level of appraisal, sensitivity review or archival description; and if there is one ensuring all the gathered information is validated and encoded. 	
Inputs	
<ul style="list-style-type: none"> ▪ Non-digital archives 	
Outputs	
<ul style="list-style-type: none"> ▪ Validated and encoded metadata, basic inventories and extra information 	

Table 19: Accession

10.3 ARCHIVES PROCESSING

10.3.1 Appraisal & Selection

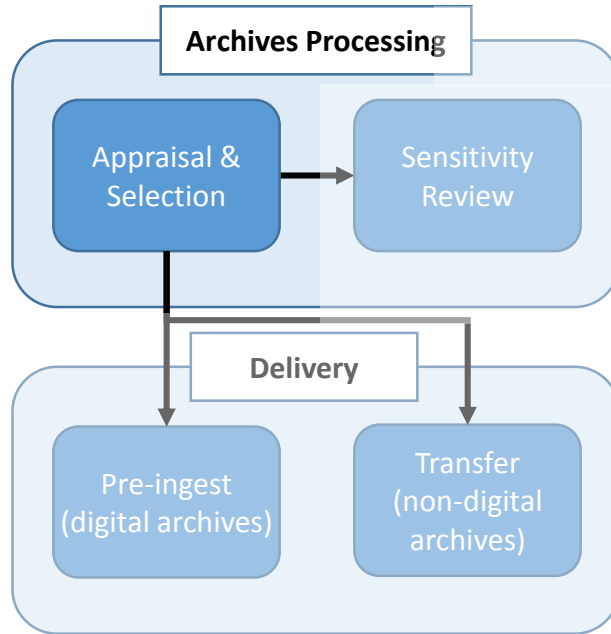


Figure 39: Appraisal & Selection

Appraisal & Selection	
Description	Responsible
<p>This is a core process for each and every Archival Institution as it determines which part of the archives will be preserved.</p> <p>“Appraisal involves measuring the drivers for retaining a dataset or record against the costs of doing so, and determining the point at which the costs outweigh the drivers. It requires assessing the data against a certain set of criteria.” (Harvey, 2008).</p> <p>It can include the following activities:</p> <ul style="list-style-type: none"> - The identification of archives in terms of historical and legal value. - The selection of the archives to be preserved. - Updating metadata (status changes) about the archives to be preserved. - The identification of the technical issues that may arise while managing the selected records. - (for appraisal carried out by the Producer) The preparation of the transfer package, including the format of records that will be transferred. <p>This process can be carried out by the Producer or by the Archival Institution. When appraisal is carried out by the Producer, a validation process by the Archival Institution will be necessary in order to approve the decisions taken.</p>	<ul style="list-style-type: none"> ▪ Archival Institution
Inputs	

<ul style="list-style-type: none"> Archives waiting to be analysed by the archivist(s)
Outputs
<ul style="list-style-type: none"> Appraised and selected archives

Table 20: Appraisal & Selection

10.3.2 Sensitivity Review

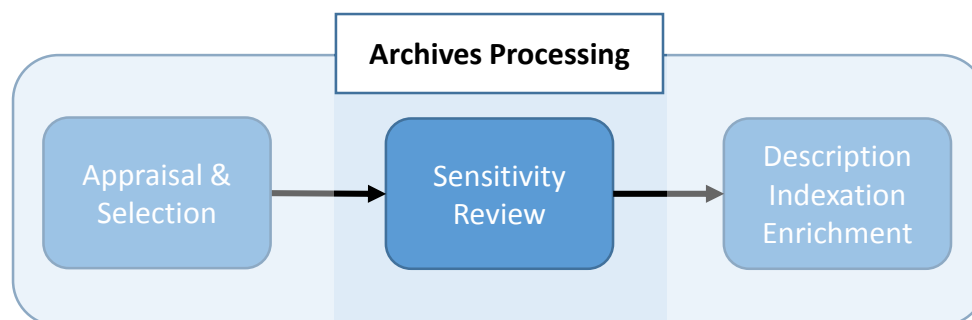


Figure 40: Sensitivity Review

Sensitivity Review	
Description	Responsible
<p>Process of determining the degree of sensitivity of the information present in archived files or documents in order to decide the most suitable degree of openness for it.</p> <p>This process will take into account the character and nature of the information as well as external factors affecting the context of it.</p> <p>This process can be carried out by the Producer or by the Archival Institution. In case it is performed by the Producer, a validation process by the Archival Institution will be necessary in order to approve the taken decisions.</p> <p>Review decisions may lead to the formal declassification of files or documents.</p>	<ul style="list-style-type: none"> Archival Institution
Inputs	
<ul style="list-style-type: none"> Non-reviewed archives 	
Outputs	
<ul style="list-style-type: none"> Reviewed archives 	

Table 21: Sensitivity Review

10.3.3 Description / Indexation / Enrichment

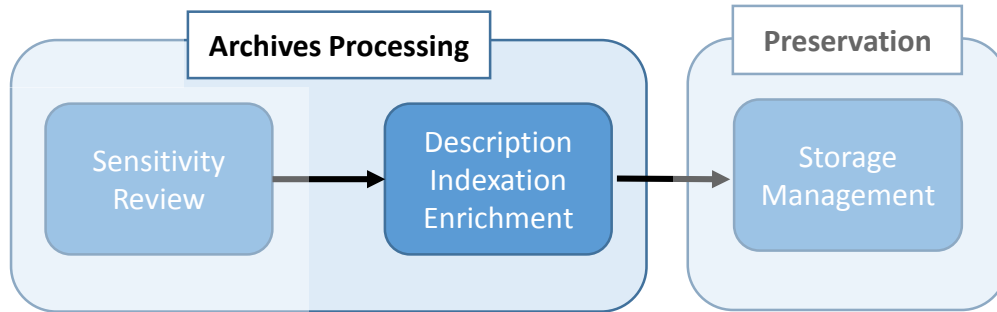


Figure 41: Description / Indexation / Enrichment

Description / Indexation / Enrichment	
Description	Responsible
<p>After the validation of ingested or accessioned archives, they can be further analysed, described and enriched with additional metadata (e.g. contextual information).</p> <p>When the metadata information is complete, all the records and the related metadata are processed and indexed according the most important aspects (e.g. source, series, topics, dates, etc.) in order to retrieve them easily.</p> <p>The corresponding index is stored in a database in order to facilitate access to the archives.</p>	<ul style="list-style-type: none"> ▪ Archival Institution
Inputs	
<ul style="list-style-type: none"> ▪ Undescribed archives 	
Outputs	
<ul style="list-style-type: none"> ▪ Described archives 	

Table 22: Description / Indexation / Enrichment

10.4 PRESERVATION

10.4.1 Storage Management

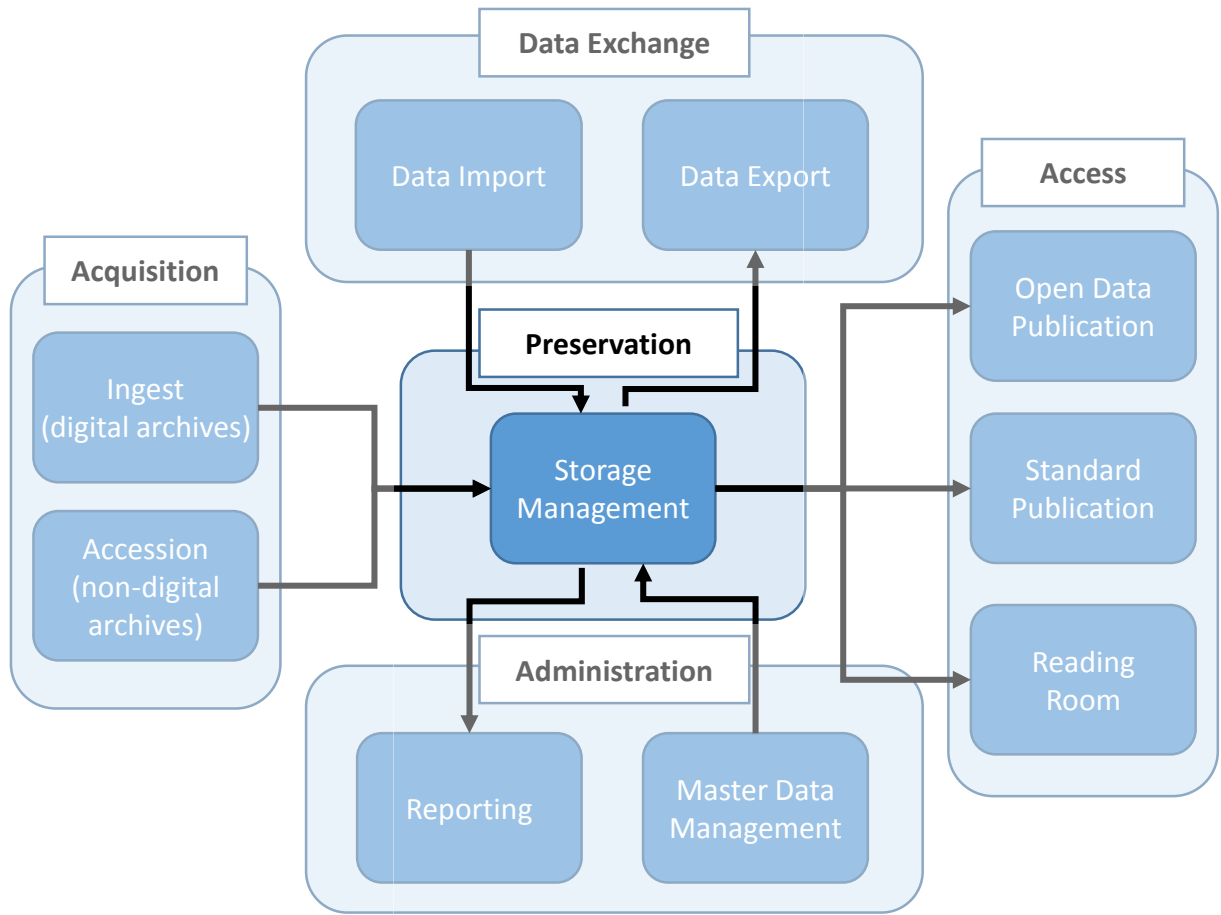


Figure 42: Storage Management

Non-digital archives

Storage Management (non-digital records)	
Description	Responsible
<p>The main activity performed in this sub-process is the storage of non-digital archives, which includes access management for the archival material stored in the physical facilities of the Archival institution.</p> <p>Regarding the sub-processes belonging to the Acquisition functional block, the Store Management sub-process controls both the physical storage in which the non-digital records will be stored.</p> <p>Regarding the Administration functional block, the Reporting sub-process will retrieve data about the physical storage controlled by the Store Management sub-process in order to elaborate its reports.</p> <p>Finally, regarding the Access functional block, its sub-process Reading Room has as its main objective the provision of access to the records stored in the physical storage controlled by the Store Management sub-process.</p>	<ul style="list-style-type: none"> Archival Institution

Inputs
▪ Non-digital archives
Outputs
▪ Stored non-digital archives

Table 23: Storage Management (Non-digital archives)

Digital archives

Storage Management (digital archives)	
Description	Responsible
<p>The main activity performed in this sub-process is the storage of digital archives, which includes the access management for the archival material stored in the databases of the Archival Institution.</p> <p>This sub-process interacts with the sub-process belonging to the Data Exchange functional block of the model as in both, data import and data export, the repository into which the data is imported or from which it is retrieved for its export is controlled by the Store Management sub-process.</p> <p>Regarding the sub-processes belonging to the Acquisition functional block, the Store Management sub-process controls the repository into which the digital-born records will be ingested.</p> <p>Regarding the Administration functional block, the Reporting sub-process will retrieve data about the repositories controlled by the Store Management sub-process in order to elaborate its reports; and the Master Data Management will perform changes to the data stored in those repositories.</p> <p>Finally, regarding the Access functional block, its sub-processes Open Data Publication and Standard Publication have as their main objective the provision of access to the data stored in the repositories controlled by the Store Management sub-process.</p>	▪ Archival Institution
Inputs	
▪ Digital archives	
Outputs	
▪ Stored digital archives	

Table 24: Storage Management (digital archives)

10.4.2 Preservation planning (digital archives)

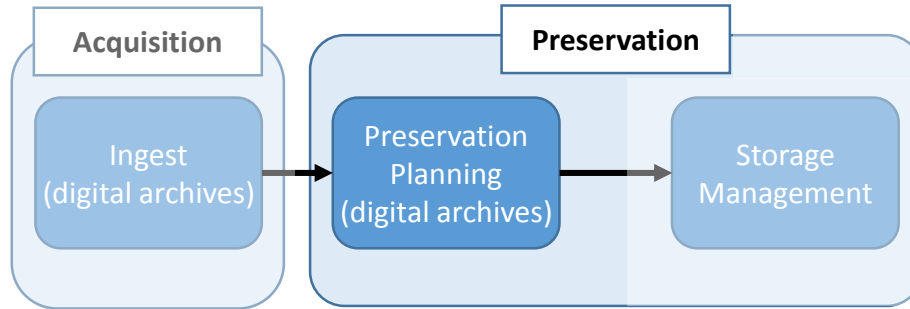


Figure 43: Preservation Planning

Preservation Planning (digital archives)	
Description	Responsible
<p>According to OAIS framework, the Preservation Planning Functional entity provides the services and means for monitoring the environment of the OAIS preservation model.</p> <p>This will be done through the provision of recommendations and preservation plans to ensure the accessibility and comprehensiveness of the information stored in the OAIS system.</p> <p>The target users of this information will be the Designated Community over the Long Term, even if the original computing environment becomes obsolete.</p> <p>Preservation Planning activities include:</p> <ul style="list-style-type: none"> - The review of content stored in the permanent storage; - The review of and recommendation on archival information updates, such as migration of data objects to valid formats. 	<ul style="list-style-type: none"> ▪ Archival Institution
Inputs	
<ul style="list-style-type: none"> ▪ Archives to be preserved 	
Outputs	
<ul style="list-style-type: none"> ▪ Preserved archives 	

Table 25: Preservation Planning

10.4.3 Elimination

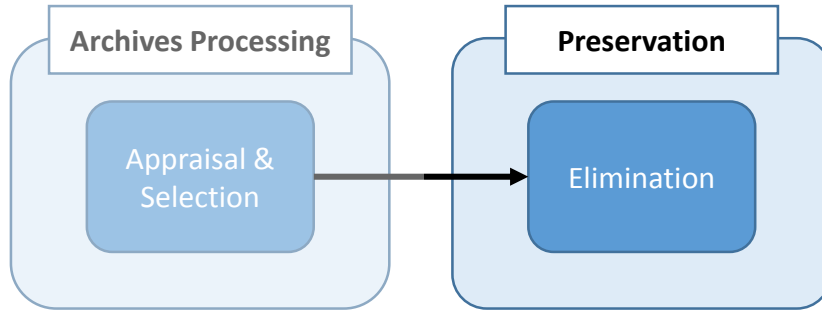


Figure 44: Elimination

Non-digital archives

Elimination (non-digital archives)	
Description	Responsible
This sub-process is a direct consequence of the “Appraisal & Selection” process and it will consist of the destroying of both the digital and non-digital archives that have been deemed to be not relevant enough to be preserved by the Archival Institution. In the latest case, this destruction will normally be made through means of an external contractor.	<ul style="list-style-type: none"> Archival Institution
Inputs	
<ul style="list-style-type: none"> Non-digital archives 	
Outputs	
<ul style="list-style-type: none"> Eliminated non-digital archives 	

Table 26: Elimination (non-digital archives)

Digital archives

Elimination (digital archives)	
Description	Responsible
This sub-process is a direct consequence of the “Appraisal & Selection” process and it will consist of the destroying of the digital archives that have been deemed to be not relevant enough to be preserved by the Archival Institution.	<ul style="list-style-type: none"> Archival Institution
Inputs	
<ul style="list-style-type: none"> Digital archives 	
Outputs	

- Eliminated digital archives

Table 27: Elimination (digital archives)

10.5 ADMINISTRATION

10.5.1 Reporting

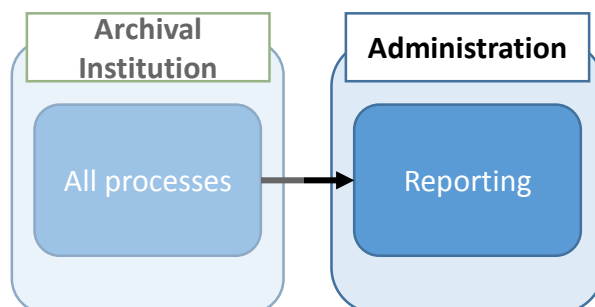


Figure 45: Reporting

Reporting	
Description	Responsible
Process of analysis and elaboration of reports including diverse characteristics of the data stored and the processes carried out by the Archival Institution. These reports will constitute the basis for the contextualisation of the current state of the records within the Archival Institution and, as such, the basis on which its administration will take the decisions concerning the preservation of the records and the management of the institution.	<ul style="list-style-type: none"> ▪ Archival Institution
Inputs	
<ul style="list-style-type: none"> ▪ Data collected regarding diverse aspects of the Archival Institution 	
Outputs	
<ul style="list-style-type: none"> ▪ Reports 	

Table 28: Reporting

10.5.2 Request Management

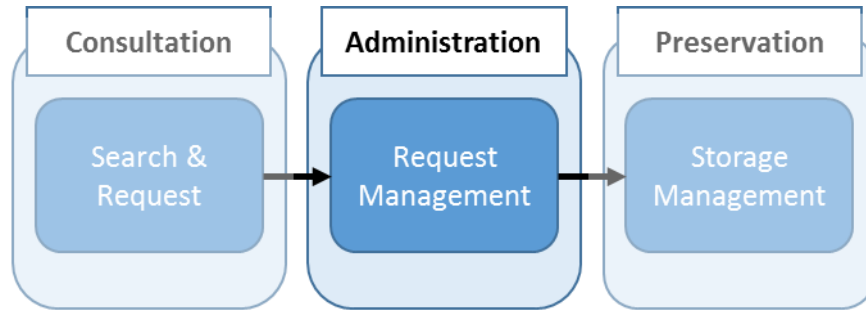


Figure 46: Request Management

Request Management	
Description	Responsible
<p>Administrative process focused on the control of the requests made by the Consumers and other IO Stakeholders for the retrieval of information from the storage.</p> <p>It will be composed of the following activities:</p> <ul style="list-style-type: none"> • Request reception • Request log • Request management • Request reply (through the provision of access to the Archival Institution's records or through the denial of the request) • Reply log • Timing control. This will only take place when the request made has a time constraint associated (i.e. loans) 	<ul style="list-style-type: none"> ▪ Archival institution
Inputs	
<ul style="list-style-type: none"> ▪ Request 	
Outputs	
<ul style="list-style-type: none"> ▪ Managed request 	

Table 29: request Management

10.5.3 Master data management

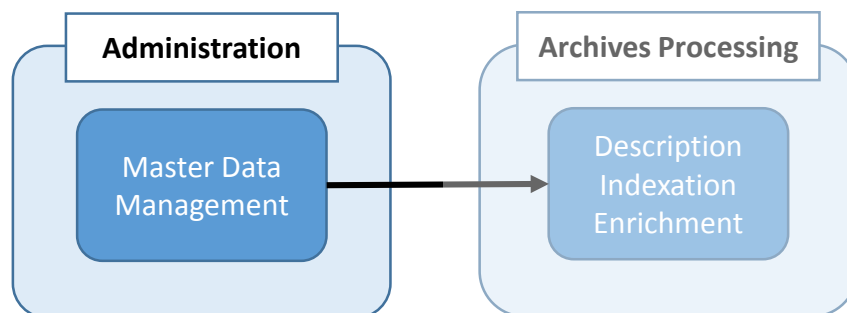


Figure 47: Master Data Management

Master Data Management	
Description	Responsible
<p>This process deals with the activities required to reach the high quality of the data. For this, the different master data (main entities of the data model) are unified and cleaned in order to obtain a set of consistent and uniformed identifiers.</p> <p>Master data management refers to the set of processes, tools and policies defined to manage the master data used in the different repositories. It has the objective of providing processes and policies to:</p> <ul style="list-style-type: none"> • Define, use, update, deprecate and delete master data. • Collect, aggregate, match and migrate existing/new master data. • Assure, control and audit the quality of the master data itself and of its usage in the different processes defined in the archives management system. • Curation and migration of existing data triggered by updates/changes of master data. • Dissemination and distribution of master data across the organisation and its external stakeholders. <p>Master data management plays a key role in the standardisation and use of agreed master data not only within the archives management system but also across multiple organisations. The policies defined thereof enable the implementation of the interoperability agreements²⁴ arranged by the organisation with external stakeholders. These agreements can be used afterwards for description and/or enrichment purposes.</p>	<ul style="list-style-type: none"> ▪ Archival Institution
<p>Inputs</p> <ul style="list-style-type: none"> ▪ Digital and non-digital archives 	

²⁴ Interoperability agreements are the formalisation of cooperation arrangements between organisations to promote interoperability and can be defined in any of the four interoperability layers defined by the European Interoperability Framework. While interoperability agreements at legal and organisational level will usually be very specific to the European public service concerned, interoperability agreements at technical level and, to a lesser extent, at semantic level can often be mapped onto existing formalised specifications.

<ul style="list-style-type: none"> ▪ Master data sources ▪ Interoperability agreements
Outputs
<ul style="list-style-type: none"> ▪ Consistent and uniformed digital and non-digital archives

Table 30: Master Data Management

10.5.4 Authority Lists Management

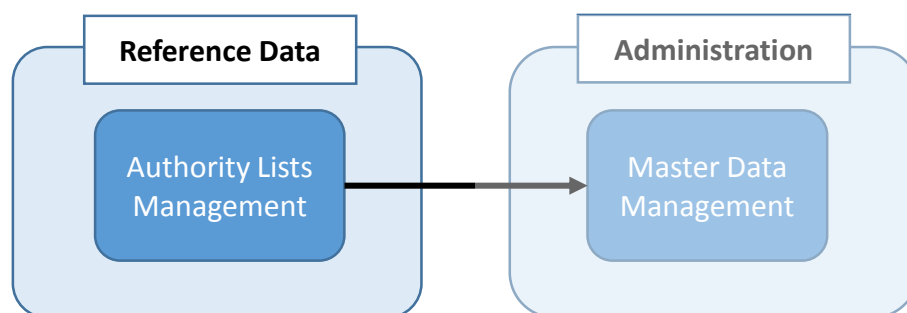


Figure 48: Authority Lists Management

Authority Lists Management	
Description	Responsible
<p>This process can be considered an extension of master data management focused on reference data. Before describing the different strategies to handle reference data, it is important to identify the three main typologies of reference data that can be used by an organisation:</p> <ul style="list-style-type: none"> - Common vocabularies, produced and maintained by authorities like EuroStat, ISO, etc. Examples of common vocabularies are NUTS²⁵ or ISO country codes²⁶. - Domain specific vocabularies, produced and maintained by a community in order to share a common terminology. An example can be the vocabularies defined by MDR²⁷, which offers definition data used by European Institutions involved in the legal decision-making process. - Application vocabularies, ad-hoc and shareable vocabularies defined and maintained by one organisation to support its internal needs. 	<ul style="list-style-type: none"> ▪ IO Stakeholder

²⁵ The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU. See <http://ec.europa.eu/eurostat/web/nuts/background> for further information.

²⁶ ISO standard that defines internationally recognised codes of letters and/or numbers that we can use when we refer to countries and subdivisions. See <https://www.iso.org/iso-3166-country-codes.html> for further information.

²⁷ The Metadata Registry (MDR) registers and maintains definition data (metadata elements, named authority lists, schemas, etc.) used by the different European Institutions involved in the legal decision-making process. See <http://publications.europa.eu/mdr/index.html> for further information.

<p>The management of reference data varies from one type to the other. Common vocabularies can be considered as master data, and therefore, managed by the master data management process.</p> <p>Domain specific vocabularies require the adoption of a governance model that defines processes to manage the vocabulary, roles and governance bodies, and interoperability agreements. From the perspective of a stakeholder, it entails the definition of processes to:</p> <ul style="list-style-type: none"> - Produce/contribute to the definition of the vocabulary. - Manage change requests/updates on the vocabulary and on releases (e.g. minor, major), steered by a change management board. - Plan the roll out of new versions, including the coexistence of different versions when needed. - Support the transition and mapping of different versions. - Uptake and publish the reference data. - Define/decide the terms of usage of the vocabulary. <p>Application vocabularies are produced by an organisation who is offering the vocabulary to others. Their management falls primarily on the organisation that has produced the vocabulary. The nature of this sharing can vary from a mere publication to a more sophisticated model like the one defined for domain specific vocabularies. An application vocabulary used in a domain by different stakeholders can turn into a domain specific vocabulary.</p>	
<p>Inputs</p>	
<ul style="list-style-type: none"> ▪ Data list ▪ Change requests 	
<p>Outputs</p>	
<ul style="list-style-type: none"> ▪ Reference Data (code lists, controlled vocabularies, taxonomies). ▪ Governance model (optional) ▪ Interoperability agreements (optional) 	

Table 31: Authority Lists Management

10.6 ACCESS

10.6.1 Reading Room

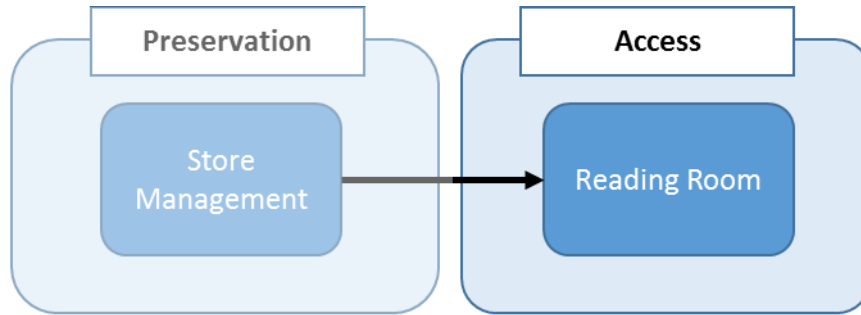


Figure 49: Reading Room

Reading Room	
Description	Responsible
This process consists of an Access modality based on the temporary loan of the archives preserved by the Archival institution to a Consumer. Certain restrictions apply to the loan, such as the location where it will be possible to access the records. In this case, the Archival Institution designates what the facilities are.	<ul style="list-style-type: none"> Archival Institution
Inputs	
<ul style="list-style-type: none"> Archives preserved by the Archival Institution 	
Outputs	
<ul style="list-style-type: none"> Loan of archives to a Consumer so he can access them at the facilities designated by the Archival Institution. 	

Table 32: Reading Room

10.6.2 Standard Publication

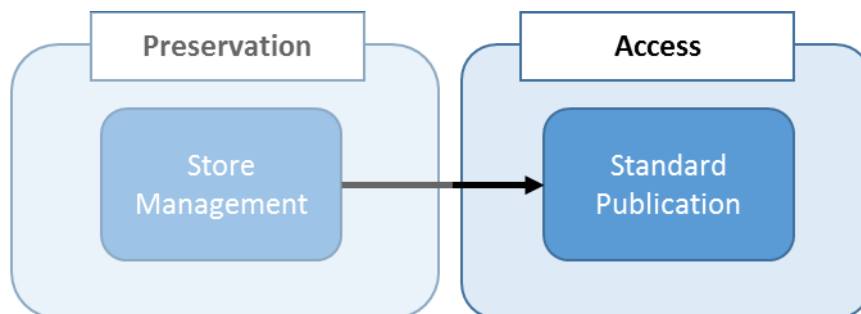


Figure 50: Standard Publication

Standard Publication

Description	Responsible
Process of data dissemination in a timely manner and in both human and machine-readable formats. This process will allow the massive dissemination of data through a wide variety of platforms (i.e. data portals, websites, etc.) and may be used by other machines through the use of an API.	▪ Archival Institution
Inputs	
▪ Records preserved by the Archival institution	
Outputs	
▪ Published records	

Table 33: Standard Publication

10.6.3 Open data publication

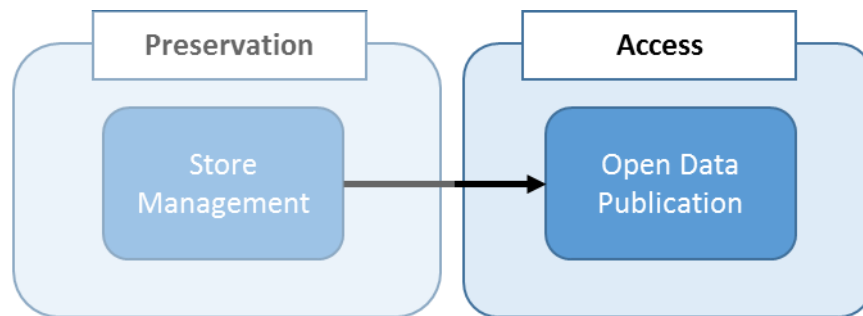


Figure 51: Open Data Publication

Open Data Publication	
Description	Responsible
<p>The W3C Linked Data Cookbook defines the publication in LOD as a group of processes to transform a simple dataset into a Linked Data Representation. This will enable better search, access and re-use of open information (Highland & Wood, 2011).</p> <p>Below, the activities conforming the Linked Open Data publication process will be further detailed in order to give a general overview of this process:</p> <ul style="list-style-type: none"> - Modelling the Data: in order to make data available within an organisation or on the public web, it first needs to be identified (which datasets will be used), structured, and modelled. Identifying the data (things) helps expressing how these data are related to other data. In this first phase, it is important to carry out a benchmarking investigation to see how others are modelling similar data. 	▪ Archival institution

<ul style="list-style-type: none"> - Name things with URIs: use Uniform Resource Identifiers (URI) to name the objects modelled in the first step. Each object must have a unique identifier. - Re-use vocabularies: LOD processes use an existing and used vocabulary whenever possible (i.e. Dublin Core, FOAF, vCard, GeoNames). This will promote interoperability and accessibility. - Publish human and machine-readable descriptions: The data schema is built as self-describing as possible. Self-describing data suggests that information about the encodings used for each representation is provided explicitly within the representation to promote the reusability. - Convert data: once the schema is finished, the next step is to convert this schema to RDF (a Linked Data representation) (i.e. Turtle, RDF/XML, N-triples). This step finishes with a validation to avoid unnecessary errors. - Specify an appropriate license: inform the users on how the data can be (re)used (i.e. Creative Commons). - Publish Linked Data: make your data fully available and accessible in the web so it can be used and reused by all users. This will allow the users to get and attach more content, enrich and give context to other data in order to exploit the existing relationships between it. - The data will be transformed into RDF format, stored in a semantic repository and published through a SPARQL endpoint 	
Inputs	
<ul style="list-style-type: none"> ▪ Unstructured data 	
Outputs	
<ul style="list-style-type: none"> ▪ Published Linked Open Data 	

Table 34: Open Data Publication

10.7 CONSULTATION

10.7.1 Search & Request

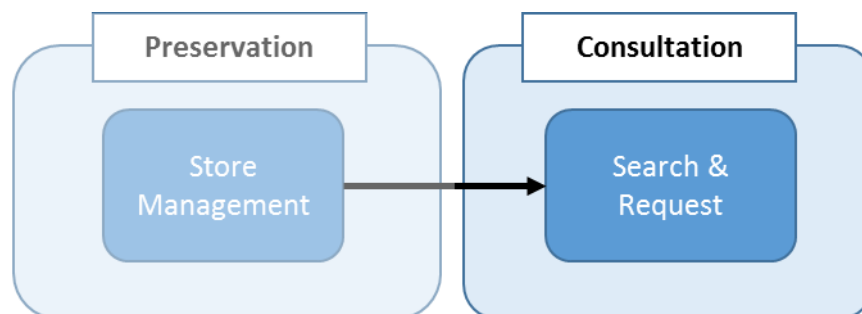


Figure 52: Search & Request

Search & Request	
Description	Responsible
<p>This process consists of an Access modality based on a request formulated by the Consumer in order to retrieve specific content form the Archival Institution.</p> <p>This request will be carried out in the shape of a query that will be sent to the Archival Institution through the means the Archival Institution has made available (i.e. portal of the archive).</p> <p>This process is format-agnostic, which means that it will take place in the same way whether the requested records are digital or non-digital.</p>	<ul style="list-style-type: none"> ▪ Consumer
Inputs	
<ul style="list-style-type: none"> ▪ Query 	
Outputs	
<ul style="list-style-type: none"> ▪ Retrieved records 	

Table 35: Search & Request

10.7.2 Loans

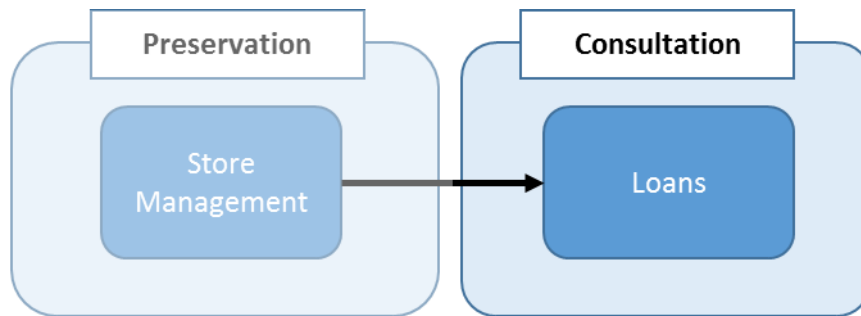


Figure 53: Loans

Loans	
Description	Responsible
<p>Loans are a specific type of request where records are borrowed to a Consumer and to which a due date applies. These will usually be requests for non-digital records, and they will have to be managed differently to other requests by the Request Management process since a schedule needs to be established that will enable the retrieval of the records borrowed from the Archival Institution.</p>	<ul style="list-style-type: none"> ▪ Consumer
Inputs	
<ul style="list-style-type: none"> ▪ Query 	

Outputs
▪ Retrieved records

Table 36: Loans

10.8 DATA EXCHANGE

10.8.1 Data Import

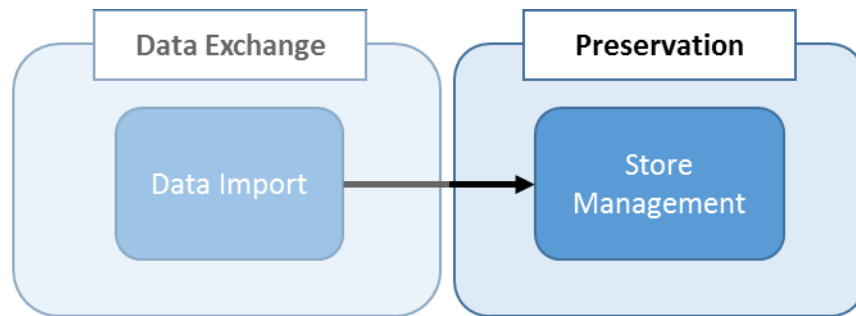


Figure 54: Data Import

Data Import	
Description	Responsible
<p>The process supports the synchronisation of metadata related to digital and non-digital archives that have been modified/updated/enriched by IO Stakeholders.</p> <p>This process requires the following activities:</p> <ul style="list-style-type: none"> - Match the archive to which the change refers to with an archive in the archives management system. The match relies on the identifier defined by the standard used to encode the dataset, e.g. EAD. - If the archive is not present in the archives management system, the archive is discarded, and a warning is included in the log file. - Otherwise, the archive metadata management systems will map the fields of the imported archive with the ones in the internal repository. <p>The update of the internal archive will be done as follows:</p> <ul style="list-style-type: none"> ○ Non-updated metadata fields will remain unaltered. ○ New metadata fields will automatically be incorporated to the archive. ○ Updated metadata fields will be merged or replaced depending on the nature of the change. <p>This process can be performed either manually or automatically. In both cases,</p>	<ul style="list-style-type: none"> ▪ Archival Institution / IO Stakeholder

<p>policies and approval workflows may be required. This process could be further extended by processing the content of the metadata fields if needed.</p>	
Inputs	
<ul style="list-style-type: none"> ▪ Datasets with metadata related to digital and non-digital archives. 	
Outputs	
<ul style="list-style-type: none"> ▪ Updated digital and non-digital archives ▪ Log with the result of the input. 	

Table 37: Data Import

10.8.2 Data export

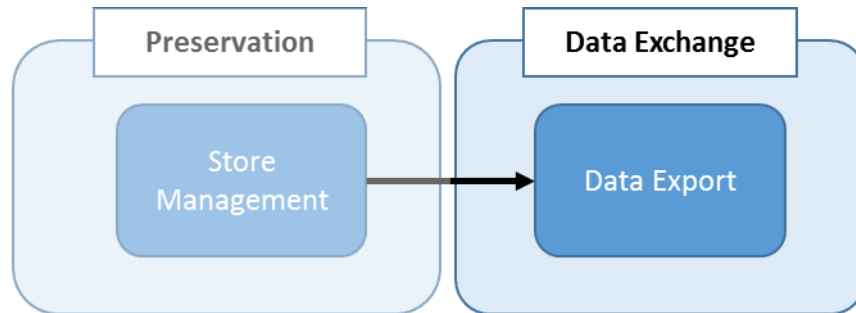


Figure 55: Data Export

Data Export	
Description	Responsible
<p>The export process is used to produce datasets with metadata related to digital and non-digital archives that will be exchanged with IO Stakeholders.</p> <p>This process requires the following activities:</p> <ul style="list-style-type: none"> • Selection of the digital and non-digital archives to be exported. • Selection of the metadata fields to be exported. • Selection of the format(s) to be exported. <p>Based on the selection above, the archives management system will produce a set of files that can be shared afterwards with the IO Stakeholder by a shared folder, an API, email, a storage device, etc.</p>	<ul style="list-style-type: none"> ▪ Archival Institution / IO Stakeholder
Inputs	
<ul style="list-style-type: none"> ▪ Digital and non-digital archives to be exported ▪ Metadata fields to be exported ▪ Format selection 	

Outputs
▪ Datasets in predefined formats

Table 38: Data Export

11 ANNEX B: STANDARDS AND BEST PRACTICES

11.1 INTRODUCTION

The structure of the inventory is aligned on the following sections referring:

- Archival description
- Archives interoperability
 - Access and Publication
 - Data Exchange
- Records management
- Preservation
- Security
- Storage Information
- Metrics and key performance indicators

Each section refers to a group of standards that supports activities identified in the scope of archives management and data exchange.

The criteria and description of each standard cover its general characteristics, identify its typology and describe the use of the standard into the archives management business processes. They also cover each standard's evolution based on information about general status and releases.

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	<i>Standard name</i>
	URL	<i>List of links interest about standard and its description</i>
	Basic description	<i>Main topics and brief definition</i>
	Owner	<i>Current proprietary of the standard</i>
	Group of standards	<i>Classification of standard based on description obtained in sources such as Data content, Data structure, Data value, Technical standard, ISO or National standard, Best practices, Guidelines, Technical specifications</i>
Use	Business area	<i>List of processes identified in this study²⁸</i>
	Features	<i>More detailed description of its characteristics and uses</i>
	Area of use	<i>Geographical area</i>
Evolution	Standard maturity	<i>Historical information and maintenance status</i>

²⁸ D02 Inventory of Business processes

Table 39: Inventory template²⁹

11.2 ARCHIVAL DESCRIPTION

Archival description is the process of analysing, organising, and recording details about the formal elements of a record or collection of records, such as creator, title, dates, extent, and contents, to facilitate the work's identification, management, and understanding. The purpose of archival description is to identify and explain the context and content of archival material in order to promote its accessibility.

Descriptive standards can facilitate the exchange of information between archives at local, national, and international levels.³⁰

There are three categories of standards for the archival description activity: data content standards, data structure standards and data value standards³¹.

An institution will implement content, structure and data value standards to describe its holdings. For instance, an archives service may follow ISAD(G) and ISAAR/CPF for the description of archival materials and their creators. It may generate multi-level archival description in EAD to exchange data and single-level description in MARC library cataloguing or Dublin Core standards to publish data into library catalogues or external repositories as Europeana. Each metadata of ISAD(G) might use controlled names, such as LCSH, for subject description, GeoNames and VIAF for authority names, or Library Classification Systems as UDC.

Specific elements of information about archival materials are recorded at every phase of their management (e.g., creation, appraisal, accessioning, conservation, arrangement) if the material is to be on the one hand securely preserved and controlled, and on the other hand made accessible at the proper time to all who have a right to consult it.

Archival description in the widest sense of the term covers every element of information no matter at what stage of management it is identified or established. At every stage the information about the material remains dynamic and may be subject to amendment in the light of further knowledge of its content or the context of its creation. Computerized information systems in particular may serve to integrate or select elements of information as required, and to update or amend them.³²

For example, specific ISAD(G) sections cover information about appraisal, arrangement, destruction or conditions of access within specific areas of description. Standards, such as ISAD(G), DACS and RAD, can also be used to create accession records, database records, catalogue records, inventories and guides.

²⁹ The template has been developed by everis based on several sources of description of standards, such as the one provided by the Library of Congress (<https://www.loc.gov/preservation/digital/formats/fdd/fdd000020.shtml>).

³⁰ http://www.cdncouncilarchives.ca/RAD/RAD_Frontmatter_July2008.pdf

³¹ Millar, Laura A. Archives. Principles and practices in records management and archives. 2017. 2nd Ed. Facet publishing <http://www.archives.org.uk/about/sections-interest-groups/archives-a-technology/news-and-events.html>

³² https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf

ELEMENTS OF DESCRIPTION	
3.1 IDENTITY STATEMENT AREA	
3.1.1 Reference code(s)	
3.1.2 Title	
3.1.3 Date(s)	
3.1.4 Level of description	
3.1.5 Extent and medium of the unit of description (quantity, bulk, or size)	
3.2 CONTEXT AREA	
3.2.1 Name of creator(s)	
3.2.2 Administrative / Biographical history	
3.2.3 Archival history	
3.2.4 Immediate source of acquisition or transfer	
3.3 CONTENT AND STRUCTURE AREA	
3.3.1 Scope and content	
3.3.2 Appraisal, destruction and scheduling information	
3.3.3 Accruals	
3.3.4 System of arrangement	
3.4 CONDITIONS OF ACCESS AND USE AREA	
3.4.1 Conditions governing access	
3.4.2 Conditions governing reproduction	
3.4.3 Language/scripts of material	
3.4.4 Physical characteristics and technical requirements	
3.4.5 Finding aids	
3.5 ALLIED MATERIALS AREA	
3.5.1 Existence and location of originals	
3.5.2 Existence and location of copies	
3.5.3 Related units of description	

Figure 56: Sections of ISAD(G)

Family fonds (family papers): description of the fonds level and one of its series and files.
Language of description: English (U.S.A.)

Fonds level

3.1.1 Reference code(s)	US DNA NW-XP
3.1.2 Title	The Robert E. Peary Family Collection
3.1.3 Date(s)	1798-1976 (dates of accumulation)
3.1.4 Level of description	Fonds
3.1.5 Extent and medium of the unit of description (quantity, bulk, or size)	312 linear feet of photographs, maps and charts, and textual records
3.2.1 Name of creator(s)	Robert E. Peary Family
3.4.3 Language / scripts of material	English

Series level

3.1.1 Reference code(s)	US DNA NWDNS-XPE
3.1.2 Title	Peary Family Photographs
3.1.3 Date(s)	1890-1916 (dates of creation) ca. 1960 (dates of accumulation)
3.1.4 Level of description	Series
3.1.5 Extent and medium of the unit of description (quantity, bulk, or size)	260 photographic prints, copy negatives, and transparencies
3.2.4 Immediate source of acquisition or transfer	The records described in this series were among the initial Instrument of Gift by Marie Peary Kuhne and Robert E. Peary, Jr., signed by Wayne Grover, the Archivist, on May 20, 1964 (accession III-NCA-438).

Figure 57: Example of transfer record included in ISAD(G)

11.2.1 Data content standards

Data content standards are a set of formal rules that specify the content, order, and syntax of information to promote consistency. A content standard goes beyond identifying the general type of information and

indicates how to select between different equivalent representations of the information as well as the manner in which the information will be structured. They cover, but are not limited to, such issues as punctuation and capitalisation, formats for expressing dates and quantities, and required vs. optional inclusion of specific items of information.

For example, NODAC is the Catalan translation of the General International Standard Archival Description ISAD(G) and its adaptation to the Catalonia scope.

NODAC defines the following levels of hierarchy in a classification system:

- Level 1: fonds
- Level 2: sub-fonds
- Level 3: series group
- Level 4: series
- Level 5: file level description
- Level 6: item level description

And, exceptionally, the level unit of installation.

Levels 1 and 2 (fonds and sub-fonds) correspond to the producers, levels 3 and 4 (series group and series) correspond to levels of classification in accordance with functions, competences and activities, and levels 5 and 6 (file and item level descriptions) correspond to individual documents which can be made up of more than one or one document. The unit of installation level corresponds to the physical installation of one or more documents.

NODAC is articulated into seven areas of descriptive information that assembles the twenty-six elements identified and defined by ISAD(G), and can be combined in the way thought most convenient to prepare the description of a specific archival unit.

Of the twenty-six elements, only a few are considered indispensable in an archival description or essential for the exchange of information. The incorporation of more elements than are considered essential or obligatory in a specific archival description depends on the criteria used by the archivist, the nature of the unit of description, the available human resources, the available information, etc. A table is included below which reflects the elements that are considered obligatory, recommended or optional for the distinct levels of description. Those that are obligatory must always be included; the recommended are always used when the circumstances or the information at their disposal permits; and the optional, are used at the discretion of the archivist.

TABLE OF OBLIGATORY, RECOMMENDED AND OPTIONAL ELEMENTS

		OB	Obligatory	RE	Recommended	op	Optional		
1		fonds	sub-fonds	series group	series	file-level description	item-level description	Unit of install	
1.1	Reference code	OB	OB	OB	OB	OB	OB	OB	
1.2	Level of description	OB	OB	OB	OB	OB	OB	OB	
1.3	Title	OB	OB	OB	OB	OB	OB	OB	
1.4	Date(s)	OB	OB	OB	OB	OB	OB	OB	
1.5	Volume and medium	OB	OB	OB	OB	OB	OB	OB	
2		CONTEXT AREA							
2.1	Name (s) of the producer(s)	OB	OB	op	op	op	op	op	
2.2	History of the producer(s)	OB	OB	op	op	op	op	op	
2.3	Archival history	OB	OB	op	op	op	op	op	
2.4	Admission information	OB	OB	op	op	op	op	op	
3.		CONTENT AND STRUCTURE AREA							
3.1	Scope and content	OB	RE	op	RE	op	op	op	
3.2	System of organisation	RE	RE	op	op	op	op	op	
3.3	Information on evaluation, selection and deletion	op	op	op	RE	op	op	op	
3.4	Additions	op	op	op	op	op	op	op	

Standards for data contents and data construction provide guidelines for the content of the data structures. Many data content standards are guidelines, namely 'rules for activities that should be applied as consistently as possible but which, by their nature, will not necessarily produce identical results even when followed'.³³

The standards list provided covers the main archival description standards developed by international and national associations (from European countries and also USA and Canada

General International Standard Archival Description (ISAD(G))

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	The General International Standard Archival Description - ISAD(G)
	URL	https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf
	Basic description	The General International Standard Archival Description - ISAD(G), an internationally agreed metadata standard, has been developed by the International Council on Archives (ICA) committee on descriptive standards. The main purpose of ISAD(G) is to ensure consistency across archival finding aids to enhance resource discovery. This is with the purpose to a) identify and explain context and content of archival materials, b)

³³ <https://www2.archivists.org/glossary/terms/c/content-standard>

Category	Evaluation criteria	Evaluation criteria values
		improve accessibility through search facilitation, and c) enable file sharing, both within and across organisations.
	Owner	International Council on Archives (ICA)
	Group of standards	Data content standard
Use	Business area	Arrangement & Description
	Features	<p>ISAD(G) is designed for hierarchical description and defines 26 elements, which together constitute the information required to adequately control an archival fonds. When all elements are used for a 'collection level description' they serve to a) identify, contextualise, describe content and structure, b) define usage rights and restrictions c) identify related material, and d) provide administrative information.</p> <p>Only six of the elements are compulsory at every level: reference code; title; creator; date(s); extent of the unit of description; and level of description. Careful specification of the ISAD(G) elements used, at each level of archival description, ensures the best possible descriptions. Extensions to the available elements may be required for greater functionality.</p>
	Area of use	International
Evolution	Standard maturity	<p>The ISAD (G) 2nd edition is the latest version of the standard, it replaced its previous version from 1994. ISAD(G) has gained wide international acceptance and is used as the structure standard in many countries.</p> <p>After initial activities since 1988 supported by UNESCO, a subgroup of the ad-hoc Commission on Descriptive Standards discussed the first draft from 1990. The first version was released in 1993/94. In 2000 the ICA published a revised version, the second edition, sometimes abbreviated as ISAD(G)2 which remains the current standard.</p> <p>ISAD(G) has been adopted as a standard by various members. In the United States, for example, the local implementation of ISAD(G) is Describing Archives: A Content Standard (2006) Adopted by several countries and translated to other languages³⁴.</p>

Table 40: ISAD(G)

³⁴ For instance: <https://www.mecd.gob.es/cultura-mecd/areas-cultura/archivos/recursos-profesionales/normas-archivisticas.html>

International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR(CPF))

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	International Standard Archival Authority Record for Corporate Bodies, Persons and Families - ISAAR(CPF)
	URL	https://www.ica.org/sites/default/files/CBPS_Guidelines_ISAAR_Second-edition_EN.pdf
	Basic description	<p>The International Standard Archival Authority Record for Corporate Bodies, Persons and Families - ISAAR(CPF) has been developed by ICA committee on descriptive standards. The latest version of the standard is the 2nd edition of 2004. The ISAAR(CPF), as internationally agreed metadata structure standard, provides guidance for the preparation of archival authority records that describe the corporate bodies, persons, and families associated with the creation and maintenance of archives. It is a set of general rules for the separate capture of information regarding records creators.</p> <p>This standard is intended to be used in conjunction with ISAD(G) - General International Standard Archival Description, 2nd edition. When these standards are used together within the context of an archival descriptive system or network, authority records will be linked to descriptions of archives, and vice versa. Descriptions of archives and records can be linked to archival authority records in the Name of creator(s) element (3.2.1) and the Administrative/Biographical history element (3.2.2) of an ISAD(G) compliant description.</p> <p>Furthermore, this standard is intended to be used in conjunction with national standards and conventions. For example, archivists may be guided by national standards when deciding which elements may or may not be repeatable. In many countries, archival descriptive systems require a single authorised form of name for a given entity, while in other countries it is permitted to create more than one authorised form of name.</p>
	Owner	International Council on Archives (ICA)
Group of standards	Data content standard	
Use	Business area	Arrangement & Description
	Features	The ISAAR(CPF) is a standard published by the International Council on Archives to establish controls for the creation and

Category	Evaluation criteria	Evaluation criteria values
		<p>use of access points in archival descriptions and to identify the kinds of information that should be used to describe a corporate body, person, or family.</p> <p>The standard comprises a set of elements, which together serve to uniquely identify the corporate body, person or family to establish standardised access points; to describe the history, roles, context and activities of the corporate body, person or family; and to establish relationships with other corporate bodies, persons and families.</p> <p>Four elements are mandatory: type of entity, authorised form of name, dates of existence, and authority record identifier.</p>
	Area of use	International
Evolution	Standard maturity	ISAAR(CPF) 2 nd edition is the latest version of the standard; it replaced its previous version from 1996 2004 2 nd edition

Table 41: ISAAR(CPF)

International Standard for Describing Institutions with Archival Holdings (ISDIAH)

The main purpose of the standard is to facilitate the description of institutions whose primary function is to keep archives and to make them available to the general public. But other entities such as cultural institutions (libraries, museums), businesses, families or individuals may hold archives. This standard, or an appropriate subset of its elements, can be applied to all entities which provide access to the records they hold.

In addition, this standard makes provision for linking information about institutions to descriptions of the records they hold and their creators. These descriptions should comply with ISAD(G) and ISAAR(CPF).

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	International Standard for Describing Institutions with Archival Holdings - ISDIAH
	URL	https://www.ica.org/sites/default/files/CBPS_2008_Guidelines_ISDIAH_First-edition_EN.pdf
	Basic description	International Standard for Describing Institutions with Archival Holdings (ISDIAH) first edition (2008) has been developed by ICA committee on best practices and standards. It is mainly used to provide general rules for the standardisation of descriptions of institutions with archival holdings.
	Owner	International Council on Archives (ICA)
	Group of standards	Data content standard

Category	Evaluation criteria	Evaluation criteria values
Use	Business area	Arrangement & Description
	Features	A standard published by the International Council on Archives to establish controls for the creation and use of access points in archival descriptions and to identify the kinds of information that should be used to describe a corporate body, person, or family. The standard can be used to develop access points for institutions with archival holdings, as well as to describe institutions as units within an archival network and establish relationships between institutions and the archives they hold.
	Area of use	Europe, USA, Canada, Australia
Evolution	Standard maturity	First published 2008. Last edition September 2011. Translation to other languages

Table 42: ISDIAH

International Standard for Describing Functions (ISDF)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	International Standard for Describing Functions - ISDF
	URL	https://www.ica.org/sites/default/files/CBPS_2007_Guidelines_I_SDF_First-edition_EN.pdf
	Basic description	The International Standard for Describing Functions (ISDF) has been developed by ICA committee on descriptive data. The standard first edition is an internationally agreed metadata structure standard which provides guidance for preparing descriptions of the functions of corporate bodies associated with the creation and maintenance of archives.
	Owner	International Council on Archives (ICA)
	Group of standards	Archival description standard - Data content standard
Use	Business area	Arrangement & Description
	Features	The standard provides guidance on how to record such information in discrete descriptions. The standard comprises a set of descriptive elements organised into four information areas. Only three elements are mandatory: type, authorised form(s) of name and function description identifier.
	Area of use	Europe, USA, Canada, Australia

Category	Evaluation criteria	Evaluation criteria values
Evolution	Standard maturity	First published 2007. Closed activities

Table 43: ISDF

Describing Archives: A Content Standard (DACS)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Describing Archives: A Content Standard - DACS
	URL	https://www2.archivists.org/groups/technical-subcommittee-on-describing-archives-a-content-standard-dacs/dacs https://www2.archivists.org/standards/DACS/appendices/appendix_c_crosswalks
	Basic description	<p>The Describing Archives: A Content Standard (DACS) has been developed by The Society of American Archivists. The latest version of the standard is the third edition of 2007. The main purpose of DACS, as US content standard, is the description of archival collections and their creators.</p> <p>DACS is related to other standards. Descriptions created according to DACS are shared electronically using encoding standards, such as Machine-Readable Cataloguing (MARC 21), Encoded Archival Description (EAD), and Encoded Archival Context (EAC). There are also close connections with Resource Description and Access (RDA) and with standards promulgated by the International Council on Archives (ICA), including International Standard Archival Description—General (ISAD[G]), the International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR[CPF]), and the International Standard for Describing Functions (ISDF).</p> <p>In particular, DACS largely conforms to the standards created by the ICA: ISAD(G) and ISAAR(CPF). All of the data elements of ISAD(G) and ISAAR(CPF) are incorporated into DACS—in some cases, virtually word for word. The exception is the exclusion of the Level of Description element from ISAD(G). It is hoped that these close ties will allow U.S. archivists to readily share information about their collections around the world. This revision continues to rely heavily on the ICA standards while</p>

Category	Evaluation criteria	Evaluation criteria values
		recognising that there is a growing convergence between museum, library, and archival practice ³⁵ .
	Owner	The Society of American Archivists
	Group of standards	Data content standard
Use	Business area	Arrangement & Description
	Features	<p>The standard is based on generally accepted archival principles, listed here but with extensive commentary in the standard.</p> <ul style="list-style-type: none"> - Records in archives possess unique characteristics. - Respect des fonds is the basis of archival arrangement and description. - Arrangement involves the identification of groupings within the material. - Description reflects arrangement. - Description applies to all archival materials regardless of form or medium. - The principles of archival description apply equally to records created by corporate bodies and by individuals or families. - Archival descriptions may be presented in a variety of outputs and with varying levels of detail. - The creators of archival materials, as well as the materials themselves, must be described. <p>The usage of DACS can be divided into three parts, a) archival description, b) description of the context of creation, and c) set out rules of the formation of authorised versions of personal, family and corporate names as well as geographic names, which may also be the name of corporate bodies.</p> <p>DACS facilitates the consistent, appropriate, and self-explanatory description of archival materials and creators of archival materials.</p>
	Area of use	<p>USA, Canada</p> <p>DACS is the United States implementation of rules proposed by the Canadian-United States Task Force on Archival Description.</p>
Evolution	Standard maturity	<p>DACS 3rd edition is the latest version of the standard, it replaced its previous version from 1989 and it has APPM as ancestor.</p> <p>The Society of American Archivists adopted Describing Archives: A Content Standard (DACS) as the official content standard of the U.S. archival community in 2005. DACS was</p>

³⁵ <https://www2.archivists.org/book/export/html/17474>

Category	Evaluation criteria	Evaluation criteria values
		designed to be used to create a variety of archival descriptions, including finding aids and catalogue records. Last edition 2013, revised on March 2015

Table 44: DACS

Rules for Archival Description (RAD)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Rules for Archival Description (RAD)
	URL	http://www.cdncouncilarchives.ca/archdesrules.html
	Basic description	Rules for Archival Description (RAD) is the Canadian content standard for archival description. Its rules are based on archival principles such as respect des fonds and description reflecting arrangement. RAD contains chapters devoted to the description of several different types of resources, including moving images, sound recordings, and objects. Its structure is similar to that of AACR2.
	Owner	Canadian Council of Archives
	Group of standards	Data content standard
Use	Business area	Arrangement & Description
	Features	RAD rules aim to provide a consistent and common foundation for the description of archival material based on traditional archival principles. The rules can be applied to the description of archival fonds, series, collections, and discrete items. The application of the rules will result in descriptions for archival material at various levels, e.g., fonds, series, file, and item levels, and will aid in the construction or compilation of finding aids of all kinds. The rules cover the description of, and the provision of access points for, all forms of material, e.g., text, graphic material, moving images, commonly found in Canadian archives at the present time. The integrated structure of the text makes the general rules usable as a basis for the description of uncommon material and material yet unknown. RAD is divided into two parts: Part I deals with the provision of information describing the material itself and Part II deals with the determination and establishment of headings (access

Category	Evaluation criteria	Evaluation criteria values
		<p>points) under which the descriptive information is to be presented and with the making of references to those headings. In both parts, the rules proceed from general to specific. In Part I, the specificity relates to the broad class of material being described (e.g., text, graphic material, sound recordings, moving images), to the level of detail required for each element of the description, and to the analysis of the parts of the unit being described (e.g., series, file, item).</p> <p>In Part II, the rules are based on the proposition that one or more provenance, author and/or non-subject access point are chosen to make accessible the units of material at various levels of description.</p>
	Area of use	Canada
Evolution	Standard maturity	1990 initial development. Revised version July 2008.

Table 45: RAD

RiC-CM, Records in Contexts, Conceptual Model

RiC-CM intends to incorporate the four existing ICA description standards in a single standard.

- From Unit of Description to Record and Record Set
- From Multilevel Description to Multidimensional Description:

ISAD(G) is based on a “multilevel” model. The standard assumes (though does not prescribe) that the scope of a single description is a fonds, and that the description begins with a description of the fonds, as a single and complete entity, and may then proceed to describe parts of the whole, and parts of the parts, all linked together to form a single, self-contained hierarchy. RiC-CM models what may be described as “multidimensional description.” Rather than a hierarchy, the description may take the form of a graph or network. The multidimensional model thus enables the description of the fonds, but also sees the fonds existing in a broader context, in relation to other fonds. In a multidimensional approach to description, the Records and Sets of Records, their interrelations with one another, their interrelations with Agents, Functions, Activities, Mandates, etc., and each of these with one another, are represented as a network within which individual fonds are situated³⁶.

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Records in Contexts (RiC) - Conceptual model (CM) for archival description
	URL	https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf

³⁶ <https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>

Category	Evaluation criteria	Evaluation criteria values
		https://www.ica.org/en/egad-ric-conceptual-model
	Basic description	RiC is the newest ICA standard for the description of records based on archival principles.
	Owner	The International Council on Archives (ICA)
	Group of standards	Data content standard
Use	Business area	Arrangement & Description
	Features	<p>The main purpose of the standard is to reconcile, integrate, and build on the four existing standards - from unit of description to record and record set and from multilevel description to multidimensional description: General International Standard Archival Description (ISAD(G)); International Standard Archival Authority Records—Corporate Bodies, Persons, and Families (ISAAR(CPF)); International Standard Description of Functions (ISDF); and International Standard Description of Institutions with Archival Holdings (ISDIAH).</p> <p>The standard has two parts: a conceptual model (RiC-CM) and an ontology (RiC-O).</p> <p>RiC-CM will resemble the existing four standards, defining the major archival descriptive entities and their properties, and the interrelations among them.</p> <p>RiC-O will be expressed as a W3C OWL (Ontology Web Language), and will have as its primary focus enabling archival descriptions to be expressed in semantic technologies.</p>
	Area of use	USA, Canada
Evolution	Standard maturity	<p>The standard in the initial stage of development by EGAD is named Records in Contexts (RiC). EGAD has begun its work by developing RiC-CM, identifying and defining the primary descriptive entities and the interrelations that constitute archival description.</p> <p>The development of the RiC-O is in the preliminary stages. Full development is pending a stable draft of RiC-CM, as the ontology must be aligned with the conceptual model. RiC-O will translate RiC-CM using the W3C standard OWL. RiC-O will provide the archival community with the ability to make archival description available using the techniques of Linked Open Data (LOD) and will employ a conceptual vocabulary and structure that is specific to archival description.</p> <p>Draft of RiC-O on 2016.</p>

Table 46: RiC-CM

11.2.2 Data structure standards

Data structure standards are formal guidelines specifying the elements into which information is to be organised. By establishing a set of elements to be included, a data structure standard also excludes other types of information.

Data structure standards define the elements of information contained in the components of an information system, including input formats (e.g., accession sheets, deeds of gift); output formats (e.g., registers, catalogues, inventories, shelf lists); and record types (e.g., holdings, donors, creators). Uniform data structure standards adopted across repositories must recognise legitimate needs of distinctive organisations for different methods and mechanisms of control. Data structure standards were developed with the goal of establishing a codified standard for description instruments electronic devices without a registered trademark, which would provide a stable and long-term deposit mechanism, and improve navigation and search on the World Wide Web³⁷.

Functional requirements that standards are intended to accomplish include³⁸:

- Present extensive and interrelated information found in the archival description instruments;
- Preserve the existing hierarchical relationship between description levels;
- Represent descriptive information that has been inherited by one hierarchical level from another;
- Navigate within a hierarchical information structure; and
- Provide support for element-specific indexing and recovery.

The result of this was the creation of standards, for example EAD which stands for Encoded Archival Description, a non-proprietary de facto standard for the encoding of finding aids³⁹ for use in a networked (online) environment. EAD allows the standardisation of collection information within and across repositories.

Furthermore, established relationships exist, implemented as crosswalks between main data content standards and data structure standards to facilitate the mapping between fields and the encoding of finding aids⁴⁰. For example, descriptions created according to DACS are shared electronically using

³⁷ <https://www.loc.gov/ead/eaddev.html> ; <http://www.archivists.org/catalog/stds99/intro.html>

³⁸ <http://www.arxiv.org/index.php/documents/publicacions/revista-lligall-1/lligall-17-1/404-04-descripcion-archivistica-codificada-ead-desarrollo-y-potencial-internacional-1/file>

³⁹ Finding aids are inventories, indexes, or guides that are created by archival and manuscript repositories to provide information about specific collections. While the finding aids may vary somewhat in style, their common purpose is to provide detailed description of the content and intellectual organisation of collections of archival materials. <https://www.loc.gov/ead/eadabout.html>

Finding aid is a single document that places the materials in context by consolidating information about the collection, such as acquisition and processing; provenance, including administrative history or biographical note; scope of the collection, including size, subjects, media; organisation and arrangement; and an inventory of the series and the folders: <https://www2.archivists.org/glossary/terms/f/finding-aid>

⁴⁰ DACS is related to other standards. Descriptions created according to DACS are shared electronically using encoding standards, such as Machine-Readable Cataloguing (MARC 21), Encoded Archival Description (EAD), and Encoded Archival Context (EAC). There are also close connections with Resource Description and Access (RDA) and with standards promulgated by the International Council on Archives (ICA), including International Standard Archival Description—General (ISAD[G]), the International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR[CPF]), and the International Standard for Describing Functions (ISDF).

encoding standards, such as Machine-Readable Cataloguing (MARC 21), Encoded Archival Description (EAD), and Encoded Archival Context (EAC).

EAD Family

EAD

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Encoded Archival Description - EAD
	URL	https://www.loc.gov/ead/ http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_for_archives_ead.pdf
	Basic description	Encoded Archival Description (EAD) is the standard for the electronic capture of hierarchical archival finding aids as well as standardised collection information in finding aids within and across repositories. EAD is a non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment.
	Owner	Non-proprietary de facto standard
	Group of standards	Data structure standard
Use	Business area	Access, Publication
	Features	EAD provides a set of elements (or tags) with which a finding aid can be marked up (or encoded) to produce an XML or SGML text file. As such EAD files can be created using any text or word processing software. There are two parts to the EAD document. The EAD Header provides control information about the finding aid itself: its identity, ownership, creation and revision. The content of the finding aid is encoded within the Archival Description section.
	Area of use	International
Evolution	Standard maturity	EAD started life as a Standard Generalised Markup Language (SGML) Document Type Definition (DTD) first published in 1998. This was revised taking into account international usage, and to bring it online with the revised 2 nd edition of the General International Standard Archival Description (ISAD(G)) as EAD 2002. And, EAD3 is the current version (https://www.loc.gov/ead/eaddev.html)

In particular, DACS largely conforms to the standards created by the ICA: ISAD(G) and ISAAR(CPF). All of the data elements of ISAD(G) and ISAAR(CPF) are incorporated into DACS—in some cases, virtually word for word. The exception is the exclusion of the Level of Description element from ISAD(G). It is hoped that these close ties will allow U.S. archivists to readily share information about their collections around the world. This revision continues to rely heavily on the ICA standards while recognising that there is a growing convergence between museum, library, and archival practice. <https://www2.archivists.org/book/export/html/17474>

Category	Evaluation criteria	Evaluation criteria values
		EAD3 is the current version

Table 47: EAD

EAD 2002

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Encoded Archival Description - EAD 2002
	URL	https://www.loc.gov/ead/eadschema.html
	Basic description	The EAD 2002 version incorporates a small number of newly-defined elements, deprecates eight previously used elements, and modifies the structure (content model) for a few elements to allow the inclusion of other valid EAD elements at different levels within a finding aid. The standard was designed to be somewhat prescriptive, providing and in some cases requiring that finding aids be structured in a particular way. This will occasionally result in differences between the structure of an encoded archival description and existing printed finding aids when they are used as the source for content.
	Owner	Non-proprietary de facto standard
	Group of standards	Data structure standard
Use	Business area	Access, Publication
	Features	Version 2002 of the EAD DTD is designed to function as both an SGML and XML DTD. To be used as an XML DTD, "switches" have been included in the DTD for turning off features used only in SGML applications, and turning on features used in XML applications.
	Area of use	International
Evolution	Standard maturity	EAD3 has replaced EAD 2002 as the official version of EAD. EAD 2002 is currently maintained

Table 48: EAD 2000

EAD 3

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Encoded Archival Description - EAD 3
	URL	https://www.loc.gov/ead/index.html

Category	Evaluation criteria	Evaluation criteria values
		https://www2.archivists.org/groups/encoded-archival-standards-section/frequently-asked-questions-about-ead-and-ead3
	Basic description	<p>EAD3 has replaced EAD 2002 as the official version of EAD. EAD3 seeks to simplify EAD and to update EAD to connect more easily with other standards like EAC-CPF.</p> <p>The goal of EAD3 was to move away from the presentation angle toward full data encoding. A number of presentation-only elements have been deprecated and other data-centric elements added. However, EAD3 remains a continuation of the 2002 schema and, in order to encourage migration, the transition from mixed presentation and data to full data is not complete.</p>
	Owner	Non-proprietary de facto standard
	Group of Standards	Data structure standard
Use	Business area	Access, Publication
	Features	<p>EAD is expressed in XML, which is a structural and preservation format. XML facilitates repurposing of data. A finding aid in XML can be converted into a variety of different formats for display and access.</p> <p>Technology, as we all know, has been moving fast. From collection management systems like Archivists' Toolkit, Archon, AtoM, and the new ArchivesSpace to the emerging possibilities of Linked Data and the release of Encoded Archival Context - Corporate bodies, Persons, and Families (EAC-CPF), the technology available to complete archival description has rapidly evolved since the release of EAD 2002. In addition, years of working with EAD had given archivists more experience with how they felt EAD would work better. In 2010, the SAA Standards Committee charged a new Technical Subcommittee for Encoded Archival Description (TS-EAD) to complete a revision of EAD within 5 years.</p>
	Area of use	International
Evolution	Standard maturity	EAD3 has replaced EAD 2002 as the official version of EAD. EAD3 seeks to simplify EAD and to update EAD to connect more easily with other standards like EAC-CPF.

Table 49: EAD 3

apeEAD

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	apeEAD Initiatives: APEnet, APEx
	URL	http://www.apenet.eu/ http://apex-project.eu/index.php/en/outcomes/standards/apeead
	Basic description	apeEAD is based on EAD and within the Archives Portal Europe EAD is used with a concept of three interrelated and interconnected layers of description. Each of them consists of individual documents structured internally with the levels of EAD.
	Owner	Non-proprietary de facto standard
	Group of Standards	Archival description standard - Data structure standard
Use	Business area	Access, Publication
	Features	The complete apeEAD schema is defined as a subset of the EAD 2002 schema fulfilling the requirements for the joint presentation in the union finding aid of the Archives Portal Europe. It is optimised for the use inside the union finding aid and contains only those parts of EAD needed for common display, for linking and identification purposes. It was drafted on the basis of a comparison of EAD profiles and practices of the National Archives participating in the project. In cases of different possibilities it names those elements and attributes commonly used – and describes how they are used.
	Area of use	EU
Evolution	Standard maturity	APEnet (Archives Portal Europe network) was a Best Practice Network project supported by the European Commission in the eContentplus programme. It ran from the 15th of January 2009 until the 15th of January 2012. Its objective was to build an Internet Gateway for Documents and Archives in Europe where seventeen European National Archives in close cooperation with the Europeana initiative were to create a common access point to European archival descriptions and digital collections. APEx – the Archives Portal Europe network of excellence was the framework for European archives to collaborate for wider and enhanced accessibility of their content on the web. It continued the work of the APEnet project (2009–2012) in which 19 European national archives and Europeana established and released the Archives Portal Europe. With the continued expansion and enhancement of the Archives Portal Europe – in

Category	Evaluation criteria	Evaluation criteria values
		<p>terms of participating countries and institutions, plus available content and services – APEX wanted to contribute to the development of a coherent digital infrastructure which opens up the portal to our common cultural heritage of archives, equally accessible for citizens, researchers, businesses and governments alike. Funded and supported by the European Commission within the framework of the Information and Technology Policy Support Programme (ICT-PSP),</p> <p>APEX will cooperate actively with Europeana on the interoperability of metadata formats and rights management of archival material.</p> <p>Finally, APEX's ultimate goal was to provide easy access via the Archives Portal Europe to as much archival content of as many European institutions as possible and equally, to channel all digitised and digital archival material to Europeana.</p> <p>The project finished at 2015</p>

Table 50: apeEAD

EAC-CPF Encoded Archival Context – Corporate Bodies, Persons, and Families

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	EAC-CPF Encoded Archival Context – Corporate Bodies, Persons, and Families
	URL	https://www.loc.gov/ead/index.html http://eac.staatsbibliothek-berlin.de/index.php?id=61 http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_articleEAC-CPF.pdf
	Basic description	EAC-CPF Encoded Archival Context – Corporate Bodies, Persons, and Families (latest version is the 1.0 from 2010) is a communication standard for encoding in Extensible Markup Language (XML) information about the corporate bodies, persons and families associated with the creation and maintenance of archives. Its primary purpose is to standardise the encoding of descriptions about agents to enable the sharing, discovery and display of this information in an electronic environment.
	Owner	Non-proprietary de facto standard
	Group of standards	Data structure standard

Category	Evaluation criteria	Evaluation criteria values
Use	Business area	Access, Publication
	Features	EAC-CPF provides for the independent description of individuals, families and corporate bodies that create, preserve, use and are responsible for and/or are associated with records. The standard supports the linking of information about one agent to other agents to show the relationships amongst record-creating entities, and the linking to descriptions of records and other contextual entities.
	Area of use	International
Evolution	Standard maturity	<p>The 2nd edition at 2003.</p> <p>In 2011, EAC became an adopted standard of the Society of American Archivists (SAA). A Technical Subcommittee (TS-EAC-CPF) was established under the SAA's Standards Committee. In 2015, the Technical Subcommittees on EAD and EAC-CPF were merged to form the Technical Subcommittee on Encoded Archival Standards (TS-EAS), responsible for the ongoing maintenance of EAD and EAC-CPF.</p> <p>Since August 2016, a subgroup within TS-EAS has been discussing questions and issues regarding EAC-CPF, which came up in different projects implementing the standard and during the revision process of EAD 2002 to EAD3. Feedback came from partners and institutions all over the world.</p> <p>In summer 2017, the TS-EAS agreed to undertake a revision of the standard EAC-CPF. To ensure the greatest possible input from users of EAC-CPF and other relevant standards, the deadline for change proposals is 11 December 2017.</p>

Table 51: EAC-CPF

Encoded Archival Guide (EAG)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	The Encoded Archival Guide (EAG) DTD and the "Censo-Guia de los Archivos de España e Iberoamerica" Project: An Electronic Guide to Spanish and Iberian American Archives
	URL	<p>https://eric.ed.gov/?id=EJ885621</p> <p>https://www.bundesarchiv.de/archivgut_online/standards_werkzeuge/02255/index.html</p> <p>http://apex-project.eu/images/docs/EAG_2012_guide_2013-05-27.pdf</p> <p>http://www.archivschule.de/uploads/Forschung/Retrokonversion/Vortraege_Kolloquium/13_Desantes.pdf</p>

Category	Evaluation criteria	Evaluation criteria values
	Basic description	<p>Encoded Archival Guide (EAG) is intended to provide a standard for the description of archives. The standard is a digital reference instrument that contains the elements and attributes necessary to provide information about the archive, its history and formation, timetables and user services, amongst others.</p> <p>There is a need for specific standards for the description of archive repositories. With this objective in mind a DTD has been created, called Encoded Archival Guide (EAG) that provides an electronic format for the storage, publication and exchange of information related to archive repositories. The different elements and attributes that make up the EAG DTD are described, as well as their application to the Electronic Guide to Spanish and Iberian American Archives, which includes information about more than 42,000 institutions.</p>
	Owner	Non-proprietary de facto standard
	Group of standards	Archival description standard - Data structure standard
Use	Business area	Access, Publication
	Features	<p>The main purpose of EAG is to facilitate an associated XML structure for encoding information about archives.</p> <p>EAG is a XML DTD, which regulates the market from general information on Archive Institutions, providing an electronic format for the storage, publication, and exchange of Archive Institutions representation.</p>
	Area of use	International
Evolution	Standard maturity	<p>The standard has been developed within the APEX project in the context of expanding and enhancing the Archives Portal Europe. Its 2012 (latest) version is the result of reviewing the existing version EAG 0.2, created initially in the context of the <i>Censo-Guía de los Archivos de España e Iberoamérica</i>, and relating this to the ISDIAH.</p>

Table 52: EAG

ISO 15836 Dublin Core - DC ()

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO 15836:2009, Information and documentation – The Dublin Core metadata element set
	URL	http://dublincore.org/

Category	Evaluation criteria	Evaluation criteria values
		http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/web_standards_for_archives_dublinc.pdf
	Basic description	<p>A standard (ISO 15836, ANSI Z39.85) that defines metadata elements used to describe and provide access to online materials.</p> <p>The Dublin Core Metadata Element Set is focused on information and documentation.</p> <p>It is a simple metadata standard that can be used across different domains. It defines fifteen elements for item description and discovery that are generally regarded as the minimum information required to adequately identify, describe and administer items in any format.</p> <p>A crosswalk is the relationships between the elements of two or more data structures. The relationship between standards as DC and ISAD(G), EAD, RAD, MARC21, MODS is defined</p> <p>Crosswalks are intended to enable comparison of EAD elements with the data elements defined in three related metadata standards or frameworks: ISAD(G), Dublin Core, and MARC. Use of these crosswalks may facilitate mapping of data between and among these metadata tools, such as for exporting data from EAD-encoded finding aids to create MARC records</p> <p>DC record is being created for an archival collection itself,</p>
	Owner	ISO Standard
	Group of standards	Data structure standard
Use	Business area	Access, Publication
	Features	<p>Dublin Core is often used to provide a common bridge between different metadata schemes, although some projects use Dublin Core as their native descriptive metadata. Originally intended to describe online resources, DC is often used to describe offline, non-digitised materials or digital surrogates of those materials.</p> <p>The data elements of unqualified Dublin Core include title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, and rights. In qualified Dublin Core, elements may be refined for greater specificity.</p> <p>Simple Dublin Core consists of 15 elements. The use of each of these is optional, and elements can be repeated.</p>
	Area of use	International
Evolution	Standard maturity	DC is maintained by the Dublin Core Metadata Initiative. ISO

Category	Evaluation criteria	Evaluation criteria values
		15836 (2009)

Table 53: ISO 15836 Dublin Core DC

MARC Family⁴¹

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	MARC Family
	URL	https://www.loc.gov/marc/marc.html
	Basic description	The Machine-Readable Cataloguing (MARC) formats are standards for the representation and communication of bibliographic and related information in machine-readable form. The MARC family is widely used for representation and exchange of authority, bibliographic, classification, community information, and holdings data in machine-readable form. They consist of a family of five coordinated formats: 1) MARC 21 Format for Authority Data; 2) MARC 21 Format for Bibliographic Data; 3) MARC 21 Format for Classification Data; 4) MARC 21 Format for Community Information; and 5) MARC 21 Format for Holdings Data.
	Owner	Library of Congress
	Group of standards	Data structure standard
Use	Business area	Access, Publication
	Features	MARC standard prevents duplication of work and allows libraries to better share bibliographic resources and make use of commercially available library automation systems to manage library operations. Also, the standard allows libraries to replace one system with another with the assurance that their data will still be compatible. It provides the mechanism by which computers exchange, use, and interpret bibliographic information, and its data elements make up the foundation of most library catalogues used today.
	Area of use	International

⁴¹ The introduction of Encoded Archival Description (EAD) as an electronic standard for displaying finding aids on the World Wide Web has not diminished the importance of bibliographic cataloguing, especially in institutions with collections including archives, published material, and other formats. <https://www2.archivists.org/glossary/terms/m/marc-format-for-archival-and-manuscripts-control>

Category	Evaluation criteria	Evaluation criteria values
Evolution	Standard maturity	The LC MARC (original) format evolved into MARC 21 and has become the standard used by most library computer programs. The MARC 21 bibliographic format, as well as all official MARC 21 documentation, is maintained by the Library of Congress.

Table 54: MARC Family

11.2.3 Data value standards

Data value standards are established lists of normalised terms used as data elements to ensure consistency. They provide lists or tables of terms, names, alphanumeric codes, or other specific entities that are acceptable for entry in a particular data element. They are used to “populate” or fill metadata elements, such as controlled vocabularies or authorities that assist with documentation and cataloguing. These metadata are used as research tools with vocabularies containing rich information and contextual knowledge, or as search assistants in database retrieval systems as well as with online collections. They include thesauri, controlled vocabularies, and authority files. Only a representative list is included in this section. Their benefit is to increase data integrity and shareability.

Controlled vocabulary	<p><i>A limited set of terms and phrases used as headings in indexes and as access points in catalogues.</i></p> <p><i>Terms in a controlled vocabulary are selected so that only one term represents a concept, allowing all material relating to that concept to be retrieved using that term even if the term does not appear in the text. The term used for indexing purposes is the preferred term. The vocabulary may also include cross-references from non-preferred terms to preferred terms.</i></p>
Classification schemes	<p><i>A diagram or chart that describes standard categories used to organise materials with similar characteristics.</i></p> <p><i>Classification schemes are often hierarchical in nature and frequently associating codes with each class. Typically used in an office of origin to file active records or in archives as a finding aid. Libraries commonly use either the Library of Congress Classification System or the Dewey Decimal Classification to organise their books. These bibliographic standards have only limited use in archives, which maintain the records in their original order.</i></p>
Authority file	<p><i>A compilation of records that describe the preferred form of headings for use in a catalogue, along with cross-references for other forms of headings.</i></p> <p><i>Authority files may be lists, card catalogues, databases, or printed publications.</i></p>
Thesaurus	<p><i>1. A list of words with related meanings, including synonyms and antonyms. - 2. A specialised vocabulary of words and phrases, commonly used for indexing, that indicates a preferred term among synonyms and shows relationships between terms.</i></p> <p><i>A thesaurus is distinguished from dictionaries and glossaries, the latter not establishing a preferred term among synonymous terms. A thesaurus often contains nothing more than headings and their relationships, where dictionaries and glossaries usually contain definitions and annotations on word use.</i></p>

Taxonomy	<i>Structure used for classifying materials into a hierarchy of categories and subcategories. In theory, the development of a good taxonomy takes into account the importance of separating elements of a group (taxon) into subgroups (taxa) that are mutually exclusive, unambiguous, and taken together, include all possibilities. In practice, a good taxonomy should be simple, easy to remember, and easy to use (TechTarget Network, 2017).</i>
Ontology	<i>The study of concepts and their relationships in an information system. As used in [computer science] circles, an ontology is a family tree of taxonomies. Where a taxonomy orders members of a single taxon, an ontology specifies relationships among taxa whose respective members, minus the ontology, might not be seen as having any relationship. (Thibodeau, 2003)</i>

Table 55: Conceptual definitions of different lists of terms (Society of American Archivists, 2017)

EuroVoc

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	EuroVoc - Multilingual Thesaurus of the European Union Interface Language:
	URL	https://data.europa.eu/euodp/en/data/group http://eurovoc.europa.eu/drupal/
	Basic description	EuroVoc is a multilingual, multidisciplinary thesaurus managed by the Publications Office. It helps to make information accessible and searchable in multiple languages using a common vocabulary. It is currently available in all 24 official EU languages as well as in some non-EU languages. EuroVoc is a common language for all documentary systems that deal with the EU's activities. It is used not only in the EU Institutions but also in national and regional institutions and in the private sector. The thesaurus is constantly evolving to meet the needs of users and indexers and to reflect the EU's activities.
	Owner	EuroVoc users include the EU Institutions, the EU Publications Office, national and regional parliaments in Europe, plus national governments and private users around the world.
	Group of standards	Archival description standard - Data value standard
Use	Business area	Arrangement & Description
	Features	EuroVoc is a multilingual thesaurus originally built up specifically for processing the documentary information of the EU Institutions. It is a multi-disciplinary thesaurus covering fields which are

Category	Evaluation criteria	Evaluation criteria values
		<p>sufficiently wide-ranging to encompass both community and national points of view, with a certain emphasis on parliamentary activities. EuroVoc is a controlled set of vocabulary which can be used outside the EU Institutions, particularly by parliaments. EuroVoc grants terminological standardisation of indexing vocabularies, allowing more accurate documentary searches as well as multilingualism, establishing language equivalences between identical concepts expressed in different languages allows documents to be indexed in the language of the archivist and searches to be made in the user's language.</p> <p>However, EuroVoc has been designed to meet the needs of systems of general documentation on the activities of the European Union. It is not suitable for indexing and searching for specialised documents. Also, it cannot claim to cover the various national situations at a sufficiently detailed level.</p>
	Area of use	Europe
Evolution	Standard maturity	<p>EuroVoc is managed by the Publications Office, which moved forward to ontology-based thesaurus management and semantic web technologies conformant to W3C recommendations as well as latest trends in thesaurus standards.</p> <p>Linked Open Data initiative launched in December 2012 in beta mode, the portal was formally established by Commission Decision of 12 December 2011 (2011/833/EU) on the reuse of Commission documents to promote accessibility and reuse.</p> <p>On 30 June 2017, a new release of EuroVoc (4.6) was made available.</p>

Table 56: EuroVoc

Library of Congress Subject Headings (LCSH)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Library of Congress Subject Headings (LCSH)
	URL	http://id.loc.gov/authorities/subjects.html https://www2.archivists.org/groups/standards-committee/library-of-congress-subject-headings-lcsh
	Basic description	Library of Congress Subject Headings (LCSH) is a controlled vocabulary maintained by the Library of Congress, covering topical subjects, genres, and geographic places.
	Owner	Library of Congress

Category	Evaluation criteria	Evaluation criteria values
	Group of standards	Data value standard
Use	Business area	Arrangement & Description
	Features	LCSH is used internationally, often in translation. Its services include all Library of Congress Subject Headings, free-floating subdivisions (topical and form), Genre/Form headings, Children's (AC) headings, and validation strings* for which authority records have been created. All of LCSH are cross-linked with RAMEAU (Répertoire d'autorité matière encyclopédique et alphabétique unifié), an authority file from the Bibliothèque nationale de France.
	Area of use	International
Evolution	Standard maturity	Work on creating LCSH list was started in 1898, with the first edition being published in 1909. The first offering of the LC Linked Data Service was the Library of Congress Subject Headings (LCSH) dataset, which was released in April 2009. The 39 th edition of LCSH (current version) contains headings established by the Library through April 2017

Table 57: LCSH

Library of Congress Name Authority File (NAF)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Library of Congress Name Authority File (NAF)
	URL	http://id.loc.gov/authorities/names
	Basic description	The Library of Congress Name Authority File (NAF) file provides authoritative data for names of persons, organisations, events, places, and titles. Its purpose is the identification of these entities and, through the use of such controlled vocabulary, to provide uniform access to bibliographic resources.
	Owner	Library of Congress
	Group of standards	Data value standard
Use	Business area	Arrangement & Description
	Features	Name descriptions also provide access to a controlled form of name through references from unused forms (e.g. a search under: Snodgrass, Quintus Curtius, 1835-1910 will lead users to

Category	Evaluation criteria	Evaluation criteria values
		the authoritative name for Mark Twain, which is, "Twain, Mark, 1835-1910."). Names may also be used as subjects in bibliographic descriptions, so they may be combined with controlled values from subject heading schemes, such as LCSH. Library of Congress Names includes over eight million descriptions created over many decades and according to different cataloguing policies. LC Names is officially called the NACO Authority File and is a cooperative effort in which participants follow a common set of standards and guidelines.
	Area of use	International
Evolution	Standard maturity	The LC Linked Data Service is an initiative of the Library of Congress that publishes authority data as linked data. It is commonly referred to by its URI: id.loc.gov. The first offering of the LC Linked Data Service was the Library of Congress Subject Headings (LCSH) dataset, which was released in April 2009. LC/NACO Authority File program allows participants to contribute authority records for personal, corporate, and jurisdictional names, uniform titles, and series.

Table 58: NAF

Library of Congress Classification (LCC)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Library of Congress Classification (LCC)
	URL	https://www.loc.gov/catdir/cpsolcc.html https://www2.archivists.org/groups/standards-committee/library-of-congress-classification-lcc
	Basic description	The Library of Congress Classification (LCC) is a system of library classification developed by the Library of Congress. It is used by most research and academic libraries in the U.S. and several other countries.
	Owner	Library of Congress
	Group of standards	Data value standard
Use	Business area	Arrangement & Description
	Features	Library of Congress Classification (LCC) is divided into 21 basic classes, each of which start with one or more uppercase letters. Full class numbers use a mixture of letters and numbers, with

Category	Evaluation criteria	Evaluation criteria values
		subtopics offset by a period. The LCC main purpose is to bring related items together in a helpful sequence from the general to the specific
	Area of use	International
Evolution	Standard maturity	The LCC is a classification system that was first developed in the late 19 th and early 20 th centuries to organise and arrange the book collections of the Library of Congress. Over the course of the 20 th century, the system was adopted for use by other libraries as well, especially large academic libraries in the United States. Updated and maintained every year.

Table 59: LCC

Virtual International Authority File (VIAF)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Virtual International Authority File (VIAF)
	URL	https://viaf.org http://www.oclc.org/en/viaf.html
	Basic description	Virtual International Authority File (VIAF) is an international authority file. It is a joint project of several national libraries and operated by the Online Computer Library Centre (OCLC). The goal of VIAF is to lower the cost and increase the utility of library authority files by matching and linking widely-used authority files and making that information available online.
	Owner	OCLC Online Computer Library Centre
	Group of standards	Data value standard
Use	Business area	Arrangement & Description
	Features	The Virtual International Authority File (VIAF) helps to make library authority files less expensive to maintain and more generally useful to the library domain and beyond. VIAF: a) Links national and regional level authority records, creating a cluster record for each unique name; b) Expands the concept of universal bibliographic control by (1) allowing national and regional variations in authorised form to coexist; and (2) supporting needs for variations in preferred language, script and spelling; and c) Plays a role in the emerging Semantic Web. In addition to providing web-accessible identification of entities of

Category	Evaluation criteria	Evaluation criteria values
		interest to libraries, VIAF builds a foundation to enable localisation of bibliographic data by making local versions of names available for searching and display.
	Area of use	International
Evolution	Standard maturity	<p>VIAF transitioned from a joint experimental activity of the United States Library of Congress (LC), the German National Library (Deutsche Nationalbibliothek, or DNB) National Library of France (Bibliothèque nationale de France, or BnF) and OCLC to become an OCLC service in early 2012.</p> <p>Since VIAF's transition to become an OCLC service in early 2012, the number of agencies participating as VIAF Contributors has grown from 19 agencies in 22 countries to 34 agencies in 29 countries (as of July 2014). 24 of the VIAF contributors are national libraries. An additional 11 national libraries provide data to VIAF through federal library agencies, consortia or other arrangements, bringing the total of national libraries represented in VIAF to 35 national libraries from 30 countries.</p>

Table 60: VIAF

UKAT (UK Archival Thesaurus)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	UKAT (UK Archival Thesaurus)
	URL	http://www.ukat.org.uk/
	Basic description	The UK Archival Thesaurus (UKAT) is a subject thesaurus which has been created for the archive sector. It is a controlled vocabulary that archiving entities can use when indexing their collections and catalogues.
	Owner	UK archives
	Group of standards	Data value standard
	Typology	Data value standard
Use	Business area	Arrangement & Description
	Features	The main uses for UKAT are: a) to improve access to archives by subject; b) to ensure that users of archives can carry out effective subject searches of the national archives network; and c) to promote the involvement in archives.

Category	Evaluation criteria	Evaluation criteria values
	Area of use	United Kingdom
Evolution	Standard maturity	The backbone of UKAT is the UNESCO Thesaurus (UNESCO), a high-level thesaurus with terminology covering education, science, culture, the social and human sciences, information and communication, politics, law and economics.

Table 61: UKAT

UNESCO Thesaurus

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	UNESCO thesaurus
	URL	http://vocabularies.unesco.org/thesaurus http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/unesco.pdf
	Basic description	The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication and information. Continuously enriched and updated, its multidisciplinary terminology reflects the evolution of UNESCO's programmes and activities. It enables subjects to be expressed consistently across different datasets and services that use it.
	Owner	UNESCO
	Group of Standards	Data value standard
Use	Business area	Arrangement & Description
	Features	The thesaurus is used for the resources available in the UNESCO Documentation Network (UNESDOC), the database of the organisation's documents and publications. The thesaurus enables users to cross-search more effectively by subject. It provides links to related subjects through a browse function, and is more innovative with linking to other kinds of data through subjects. UNESCO thesaurus links terms in a number of ways: a) It provides broader and narrower terms; b) It provides associative relationships, showing similar or related terms; and c) It steers from 'non-preferred terms' that are not part of the thesaurus, to 'preferred terms' that are.

Category	Evaluation criteria	Evaluation criteria values
		Concepts are grouped into seven broad subject areas, which are broken down into micro thesauri. The UNESCO Thesaurus is compliant with the ISO 25964 standard.
	Area of use	International
Evolution	Standard maturity	<p>The first edition of the Thesaurus was released in English in 1977, with French and Spanish translations in 1983 and 1984. The second revised and restructured version was released in 1995. Today the Thesaurus is available in English, French, Russian and Spanish.</p> <p>The latest version of the Thesaurus is available for download and view in various formats. Furthermore, it is accessible through a SPARQL endpoint</p>

Table 62: UNESCO Thesaurus

Getty Thesaurus of Geographic Names (TGN)

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Getty Thesaurus of Geographic Names (TGN)
	URL	http://www.getty.edu/research/tools/vocabularies/tgn/index.html
	Basic description	The Getty Thesaurus of Geographic Names ® (TGN), The Art & Architecture Thesaurus ® (AAT), the Union List of Artist Names ® (ULAN), the Cultural Objects Name Authority ® (CONA), and the Iconography Authority (IA) are structured resources that can be used to improve access to information about art, architecture, and material culture. Through rich metadata and links, it is hoped that the Getty vocabularies will provide a powerful conduit for research and discovery for digital art history and related disciplines.
	Owner	Getty
	Group of standards	Data value standard
Use	Business area	Arrangement & Description
	Features	The TGN is an evolving vocabulary, growing and changing thanks to contributions from Getty projects and other institutions. It is a structured resource that can be used to improve access to information about art, architecture, and material culture. Through rich metadata and links, it is hoped that Getty TGN provides a powerful conduit for research and discovery for digital art history and related disciplines.

Category	Evaluation criteria	Evaluation criteria values
		<p>TGN is intended to aid cataloguing, research, and discovery of art historical, archaeological, and other scholarly information. However, its unique thesauri structure and emphasis on historical places make it useful for other disciplines in the broader Linked Open Data cloud.</p> <p>The data for the TGN is compiled and edited in an editorial system that was custom-built by Getty technical staff to meet the unique requirements of compiling data from many contributors, building complex and changing poly-hierarchies, merging, moving, and publishing in various formats. Final editorial control of the TGN is maintained by the Getty Vocabulary Programme, using well-established editorial rules.</p>
	Area of use	International
Evolution	Standard maturity	<p>The initial core of the TGN was compiled from thousands of geographic names in use by various Getty cataloguing and indexing projects, enlarged by information from US government databases, and further enhanced by the manual entry of information from published hard-copy sources.</p> <p>TGN was first published in 1997 in machine-readable files.</p> <p>The Getty vocabularies are constructed to allow their use in linked data. Releasing the Getty vocabularies as Linked Open Data is part of the Getty's ongoing effort to make our knowledge resources freely available to all. The AAT, TGN, and ULAN are now available as LOD. LOD and all of the other data releases for the Getty vocabularies are published under the Open Data Commons Attribution License (ODC-By) 1.0.</p> <p>The TGN grows and changes via contributions from the user community and editorial work of the Getty Vocabulary Programme.</p>

Table 63: TGN

GeoNames

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	GeoNames
	URL	<p>http://www.geonames.org/</p> <p>The Ontology for GeoNames is available in OWL : http://www.geonames.org/ontology/ontology_v3.1.rdf, mappings</p>
	Basic description	GeoNames is a geographical database available and accessible through various web services, under a Creative Commons attribution license.

Category	Evaluation criteria	Evaluation criteria values
	Owner	Getty
	Group of standards	Data value standard
Use	Business area	Arrangement & Description
	Features	<p>The GeoNames geographical database contains over 10 million geographical names and consists of over 9 million unique features whereof 2.8 million populated places and 5.5 million alternate names. All features are categorised into one out of nine feature classes and further subcategorised into one out of 645 feature codes.</p> <p>The data is accessible free of charge through a number of web services and a daily database export.</p> <p>GeoNames is integrating geographical data such as names of places in various languages, elevation, population and others from various sources. All lat/long coordinates are in WGS84 (World Geodetic System 1984).</p> <p>Semantic Web integration:</p> <ul style="list-style-type: none"> • Each GeoNames feature is represented as a web resource identified by a stable URI. • The GeoNames Ontology makes it possible to add geospatial semantic information to the World Wide Web. All over 11 million GeoNames place names now have a unique URL with a corresponding RDF web service. Other services describe the relation between place names.
	Area of use	International
Evolution	Standard maturity	<p>The core of GeoNames database is provided by official public sources, the quality of which may vary.</p> <p>Through a wiki interface, users can manually edit and improve the database by adding or correcting names, etc.</p> <p>GeoNames Ontology was created October 2006. Last edition November 2012.</p>

Table 64: GeoNames

11.3 ARCHIVES INTEROPERABILITY

11.3.1 Semantic Web Standards

Semantic Web technologies are based on shared and common data models called Ontologies (formalised in RDF, OWL, SKOS, etc.). This kind of technologies permits to deploy data stores (triple store or semantic repository) over the Web. Then, the Web becomes the giant knowledge base that can be queried with Federated SPARQL queries.

Linked Data	<p><i>In computing, linked data (often capitalised as Linked Data) is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried.</i></p> <p><i>Also, various technologies allow you to embed data in documents (RDFa, GRDDL) or to store the RDF data in semantic repositories that can be shared through a SPARQL endpoint.</i></p>
Ontologies	<p><i>Ontology is the hierarchical structuring of knowledge about things by subcategorising them according to their essential (or at least relevant and/or cognitive) qualities.</i></p> <p><i>This is an explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them. Using OWL (to build shared vocabularies, or “ontologies”) and SKOS (for designing knowledge organisation systems such as taxonomies or thesaurus) it is possible to enrich data with additional meaning, which allows more people (and more machines) to do more with the data.</i></p>
Query	<p><i>Query languages go hand-in-hand with databases. If the Semantic Web is viewed as a global database, then it is easy to understand why one would need a query language for that data. SPARQL is the query language for the Semantic Web.</i></p>
Inference (or reasoning techniques)	<p><i>A semantic reasoner, reasoning engine, rules engine, or simply a reasoner, is a piece of software able to infer logical consequences from a set of asserted facts or axioms. The notion of a semantic reasoner generalises that of an inference engine, by providing a richer set of mechanisms to work with. The inference rules are commonly specified by means of an ontology language, and often a description logic language.</i></p> <p><i>Rule Interchange Format (RIF) is W3C recommendation that aims to formalise the triple patterns used to generate the new facts in the knowledge base.</i></p>

Table 65: Semantic Web Standards (W3C, 2017)

RDF

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	RDF
	Initiative name	Resource Description Framework (RDF) – W3C specification
	URL	https://www.w3.org/TR/2004/REC-rdf-primer-20040210/ https://www.w3.org/2003/01/21-RDF-RDB-access/#Introduction
	Basic description	The Resource Description Framework (RDF) is a framework for representing information on the Web. It has come to be used as a general method for conceptual description or modelling of information that is implemented in web resources, using a variety of syntax notations and data serialisation formats.

Category	Evaluation criteria	Evaluation criteria values
		<p>RDF is intended for situations in which this information needs to be processed by applications, rather just than being displayed to people.</p> <p>RDF is a language for expressing Directed Labelled Graphs (DLGs) using URIs (a superset of URLs) as node and arc identifiers. It can be expressed in XML or in non-XML syntaxes like n3 or n-triples</p>
	Owner	W3C (World Wide Web Consortium)
	Group of standards	Linked Data
Use	Business area	Access, Publication
	Features	<p>The RDF is particularly intended for representing metadata about web resources, such as the title, author, and modification date of a web page; copyright and licensing information about a web document; or the availability schedule for some shared resource. However, by generalising the concept of a “web resource”, RDF can also be used to represent information about things that can be identified on the web, even when they cannot be directly retrieved there.</p> <p>Since it is a common framework, application designers can leverage the availability of common RDF parsers and processing tools. The ability to exchange information between different applications means that the information may be made available to applications other than those for which it was originally created.</p>
	Area of use	International
Evolution	Standard maturity	Version 1.1
	Maintenance status	The RDF Working Group ended its activities on 1 July 2014.

Table 66: RDF

RDF Schema

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	RDF Schema
	Initiative name	Resource Description Framework Schema (RDFS) – W3C recommendation
	URL	https://www.w3.org/TR/2014/REC-rdf-schema-20140225/
	Basic description	RDF Schema provides a data-modelling vocabulary for RDF data.

Category	Evaluation criteria	Evaluation criteria values
		It is an extension of the basic RDF vocabulary that provides mechanisms for describing groups of related resources and the relationships between these resources. RDF Schema is written in RDF using the terms described in this document. These resources are used to determine characteristics of other resources, such as the domains and ranges of properties
	Owner	W3C (World Wide Web Consortium)
	Group of standards	Linked Data
Use	Business area	Access, Publication
	Features	RDF Schema provides a data-modelling vocabulary for RDF data. It is complemented by several companion documents which describe the basic concepts and abstract syntax of RDF [RDF11-CONCEPTS], the formal semantics of RDF [RDF11-MT], and various concrete syntaxes for RDF, such as Turtle [TURTLE], TriG, [TRIG], and JSON-LD [JSON-LD].
	Area of use	International
Evolution	Standard maturity	Version 1.1
	Maintenance status	The RDF Working Group ended its activities on 1 July 2014.

Table 67: RDF Schema

OWL

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	OWL
	Initiative name	Web Ontology Language - W3C recommendation
	URL	https://www.w3.org/standards/techs/owl#w3c_all https://www.w3.org/TR/2012/REC-owl2-overview-20121211/
	Basic description	The Web Ontology Language - OWL 2 (latest version) is an ontology language for the Semantic Web with formally defined meanings. The OWL languages are characterised by formal semantics. They are built upon a W3C XML standard for objects called the Resource Description Framework (RDF).
	Owner	W3C (World Wide Web Consortium)
	Group of standards	Linked Open Data

Category	Evaluation criteria	Evaluation criteria values
Use	Business area	Access, Publication
	Features	OWL 2 ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents. OWL 2 ontologies can be used along with information written in RDF. OWL 2 ontologies themselves are primarily exchanged as RDF documents.
	Area of use	International
Evolution	Standard maturity	Version 2. Closed activities.

Table 68: OWL

SKOS

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	SKOS
	Initiative name	Simple Knowledge Organisation System (SKOS) - W3C recommendation
	URL	https://www.w3.org/2004/02/skos/
	Basic description	SKOS - Simple Knowledge Organisation System is an area of work developing specifications and standards to support the use of knowledge organisation systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.
	Owner	W3C (World Wide Web Consortium)
	Group of standards	Linked Open Data
Use	Business area	Access , Publication
	Features	SKOS provides a standard way to represent KOS using the Resource Description Framework (RDF). Encoding this information in RDF allows it to be passed between computer applications in an interoperable way. Using RDF also allows KOS to be used in distributed, decentralised metadata applications.
	Area of use	International
Evolution	Standard maturity	Version 1.0 (August 18, 2009) Closed activities.

Table 69: SKOS

SPARQL

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	SPARQL
	Initiative name	SPARQL Query Language for RDF – W3C recommendation
	URL	https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/
	Basic description	SPARQL 1.1 (latest version) is a set of specifications that provide languages and protocols to query and manipulate RDF graph content on the web or in an RDF store.
	Owner	W3C (World Wide Web Consortium)
	Group of standards	Linked Open Data
Use	Business area	Access , Publication
	Features	SPARQL allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns. SPARQL provides a full set of analytic query operations such as JOIN, SORT, and AGGREGATE for data whose schema is intrinsically part of the data rather than requiring a separate schema definition. The SPARQL query language for RDF is designed to meet the use cases and requirements identified by the RDF Data Access Working Group in RDF Data Access Use Cases and Requirements.
	Area of use	International
Evolution	Standard maturity	Version 1.1 (26 March 2013) Closed activities.

Table 70: SPARQL

11.3.2 Standards for Repository Interoperability

Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) is a protocol designed to collect metadata from various repositories.

There are two main actors in this approach. On one side, there are the Data Provider(s), who expose and make their metadata available via OAI-PMH. On the other side is the Service Provider, who makes OAI-PMH service requests to harvest that metadata.

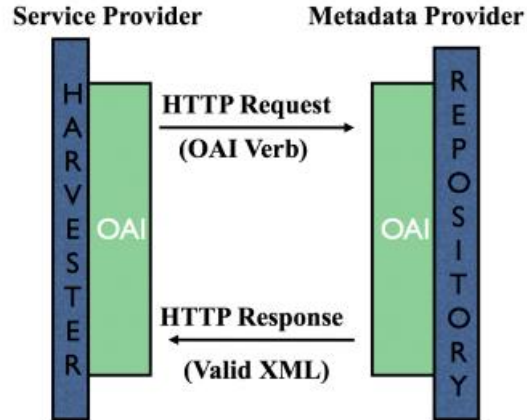


Figure 58: OAI-PMH Interchange

OAI-PMH

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	OAI-PMH
	Initiative name	Open Archives Initiative
	URL	https://www.openarchives.org/pmh/
	Basic description	The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability. Data Providers are repositories that expose structured metadata via OAI-PMH. Service Providers then make OAI-PMH service requests to harvest that metadata. OAI-PMH is a set of six verbs or services that are invoked within HTTP.
	Owner	OAI Open Archives Initiative
	Group of standards	Technical standards
Use	Business area	Access and publishing
	Features	The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH in the remainder of this document) provides an application-independent interoperability framework based on metadata harvesting. There are two classes of participants in the OAI-PMH framework: <ul style="list-style-type: none"> - Data Providers administer systems that support the OAI-PMH as a means of exposing metadata;

Category	Evaluation criteria	Evaluation criteria values
		<ul style="list-style-type: none"> - Service Providers use metadata harvested via the OAI-PMH as a basis for building value-added services. <p>A repository is managed by a data provider to expose metadata to harvesters. To allow various repository configurations, the OAI-PMH distinguishes between three distinct entities related to the metadata made accessible by the OAI-PMH.</p> <p>A resource is the object or "stuff" that metadata is "about". The nature of a resource, whether it is physical or digital, or whether it is stored in the repository or is a constituent of another database, is outside the scope of the OAI-PMH.</p> <p>An item is a constituent of a repository from which metadata about a resource can be disseminated. That metadata may be disseminated on-the-fly from the associated resource, cross-walked from some canonical form, actually stored in the repository, etc.</p> <p>A record is metadata in a specific metadata format. A record is returned as an XML-encoded byte stream in response to a protocol request to disseminate a specific metadata format from a constituent item.</p>
	Area of use	International
Evolution	Standard maturity	Version 2.0 Closed activities.

Table 71: OAI-PMH

OAI-ORE

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	OAI-ORE
	Initiative name	Open Archives Initiative
	URL	https://www.openarchives.org/ore/
	Basic description	The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of web resources.
	Owner	OAI Open Archives Initiative

Category	Evaluation criteria	Evaluation criteria values
	Group of standards	Technical Standard
Use	Business area	Access and publishing
	Features	Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of web resources. These aggregations, sometimes called compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualisation, reuse, and preservation. Although a motivating use case for the work is the changing nature of scholarship and scholarly communication, and the need for cyber infrastructure to support that scholarship, the intent of the effort is to develop standards that generalise across all web-based information including the increasing popular social networks of “web 2.0”.
	Area of use	International
Evolution	Standard maturity	Version 1.0 Closed activities.

Table 72: OAI-ORE

11.4 RECORDS MANAGEMENT

Records management refers to a set of activities required for systematically controlling the creation, distribution, use, maintenance, and disposition of recorded information maintained as evidence of business activities and transactions.

The reference model defined for records management covers the business processes such as appraisal⁴², selection, transfer or elimination, all of them high-level processes identified in the business processes inventory.

The list of standards about the processes involved in records management systems is extensive and includes standards of other areas related. The inventory has collected the most representative.

⁴² Notes: In an archival context, appraisal¹ is the process of determining whether records and other materials have permanent (archival) value. Appraisal may be done at the collection, creator, series, file, or item level. Appraisal can take place prior to donation and prior to physical transfer, at or after accessioning. The basis of appraisal decisions may include a number of factors, including the records' provenance and content, their authenticity and reliability, their order and completeness, their condition and costs to preserve them, and their intrinsic value. Appraisal often takes place within a larger institutional collecting policy and mission statement. (<https://www2.archivists.org/glossary/terms/a/appraisal>)

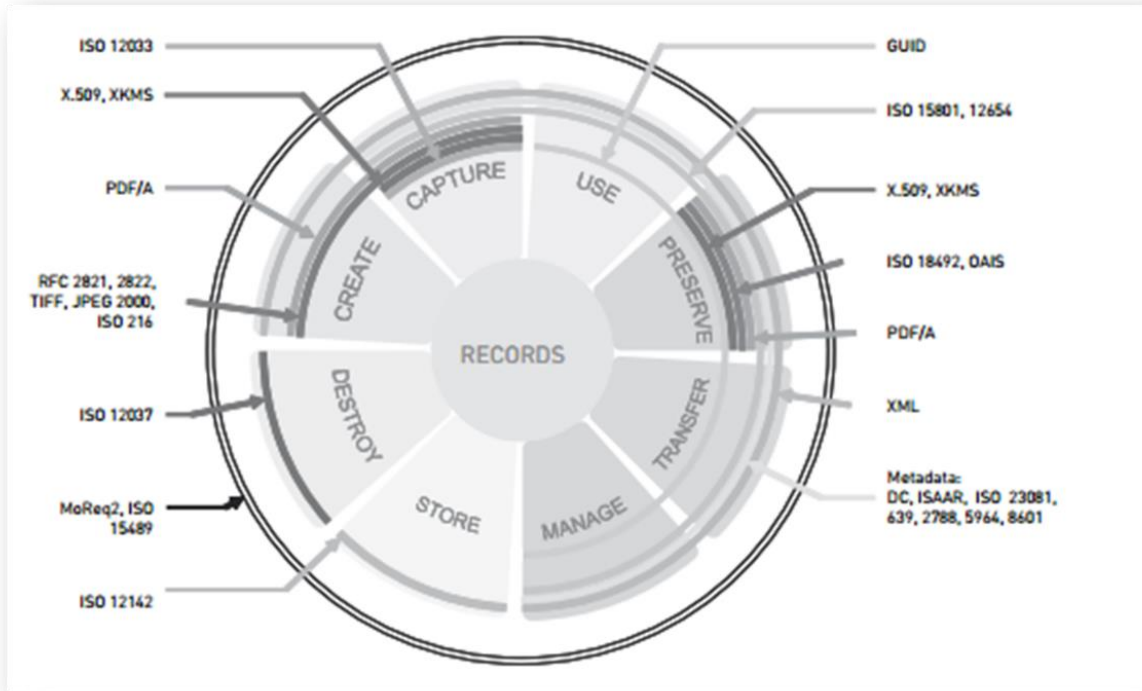

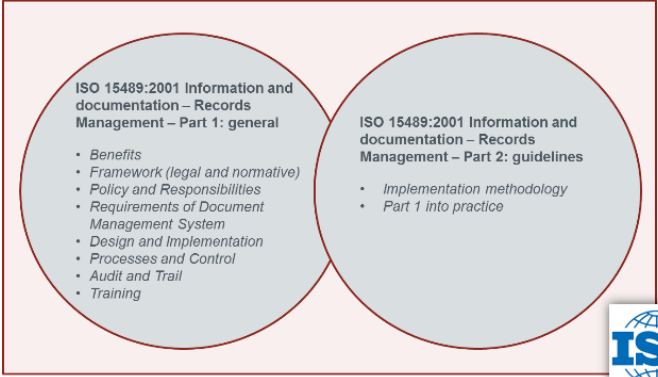


Figure 59: Standards in Records Management scope

ISO 15489:2016

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO 15489-1:2016 Information and documentation -- Records management. (Part I, II)
	URL	https://www.iso.org/standard/62542.html
	Basic description	<p>ISO 15489 Information and documentation – Records management provides high-level best practice and procedures to ensure that all records are managed appropriately and that the information they contain is retrievable.</p> <p>ISO 15489-1 Part 1: General, a best practice framework for managing records to ensure that adequate records are created and captured in to a records management system, and managed effectively. Part 2: Guidelines, aimed specifically at records management professionals within an organisation and builds on the above framework, providing more detailed recommendations for managing records.</p>

Category	Evaluation criteria	Evaluation criteria values
		<p>ISO International Standards (1) : Normative documents that provide reference material and guidelines of implementation of Records Management System</p>  <p>ISO 15489:2001 Information and Documentation – Records Management – Part 1 General</p> <p>ISO 15489:2001 Information and Documentation – Records Management – Part 2 Guidelines</p> <p>Define generic needs for recordkeeping in a business environment</p> <p>Oriented to government and non-government organisations</p>
	Owner	ISO
	Group of standards	ISO Standard
Use	Business area	Appraisal & selection
	Features	<p>ISO 15489 emphasises the importance of policies and procedures for records management and recommends having the following documentation. Classification system – to classify the records within a type framework; controlled vocabulary – for controlled subject access; dispositions – record retention schedules, to provide details of what records should be created and how long they should be retained; classification scheme – for access rights and restrictions.</p>
		<p>ISO 15489 Family</p>  <p>ISO 15489:2001 Information and documentation – Records Management – Part 1: general</p> <ul style="list-style-type: none"> • Benefits • Framework (legal and normative) • Policy and Responsibilities • Requirements of Document Management System • Design and Implementation • Processes and Control • Audit and Trail • Training <p>ISO 15489:2001 Information and documentation – Records Management – Part 2: guidelines</p> <ul style="list-style-type: none"> • Implementation methodology • Part 1 into practice
	Area of use	International

From: <http://www.iso.org/iso/30300_briefing-note.pdf>

Category	Evaluation criteria	Evaluation criteria values
Evolution	Standard maturity	2016 current edition

Table 73: ISO 15489

ISO 30300, ISO 303001, ISO 30302 Family -

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Information and documentation – management systems for records – Fundamentals and vocabulary (ISO 30300:2011) Information and documentation – management systems for records – Requirements (ISO 30301:2011) Information and documentation – management systems for records – Guidelines for implementation (ISO 30302:2015)
	URL	https://www.iso.org/standard/53732.html https://www.iso.org/obp/ui/#iso:std:iso:30300:ed-1:v1:en
	Basic description	ISO 30300:2011 defines terms and definitions applicable to the standards on management systems for records (MSR) prepared by ISO/TC 46/SC 11. It also establishes the objectives for using a MSR, provides principles for a MSR, describes a process approach and specifies roles for top management. ISO 30300 is part of a series of International Standards under the general title: Information and documentation — Management systems for records: a) ISO 30300 (Information and documentation — Management systems for records — Fundamentals and vocabulary) specifies the terminology for the Management systems for records (MSR) series of standards, and the objectives and benefits of a MSR; while b) — ISO 30301 (Information and documentation — Management systems for records — Requirements) specifies requirements for a MSR where an organisation needs to demonstrate its ability to create and control records from its business activities for as long as they are required.

ISO International Standards (2) :
A family of normative standards that provide reference and guidelines in order to define a Management System of Records that will work in any environment

ISO 30300:2011 Management system for records Fundamentals and vocabulary

ISO 30301:2011 Management system for records Requirements

ISO 30302 Management system for records Guidelines for implementation

ISO

Reference point of information management
and for the creation and control of documents

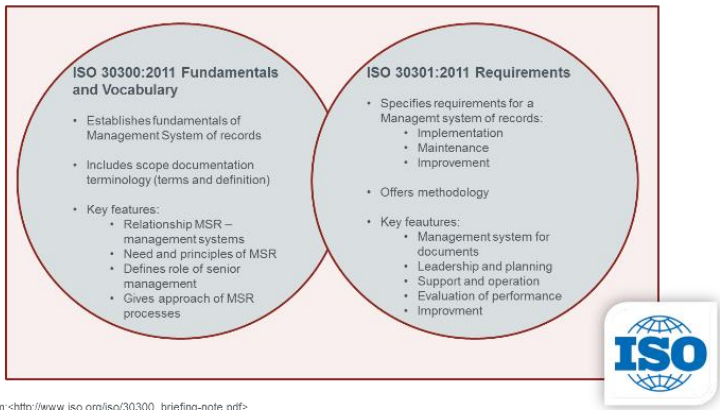
Category	Evaluation criteria	Evaluation criteria values
	Owner	ISO
	Group of standards	ISO Standard
Use	Business area	Appraisal & selection
	Features	<p>ISO 30300:2011 is applicable to any type of organisation that aims to establish, implement, maintain and improve a MSR to support its business; assure itself of conformity with its stated records policy; or demonstrate conformity with this International Standard.</p> <p style="text-align: center;">ISO 30300 Family</p> <div style="text-align: center;">  </div> <p style="text-align: center; font-size: small;">From: <http://www.iso.org/iso/30300_briefing-note.pdf></p>
	Area of use	International
Evolution	Standard maturity	ISO 30300, Fundamentals and vocabulary, and ISO 30301, Requirements widely accepted. Forthcoming ISO 30302, guidelines for implementation

Table 74: ISO 30300 Series

MoReq Family

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Model requirements for the management of electronic records (MoReq)
	URL	http://www.moreq.info/files/MoReq2010%20-%20Executive%20Summary%20EN.pdf http://moreq2.eu/attachments/article/189/MoReq2_typeset_version.pdf
	Basic description	MoReq2010® is international functional specification that defines how compliant records systems should operate and interoperate.

Category	Evaluation criteria	Evaluation criteria values
		Foundational approaches of internationally recognised standards in records management, such as ISO 15489, ISO 23081, ISO 16175, and its predecessor MoReq2, MoReq2010® goes beyond any of these in its scope, definition and ambition.
	Owner	DLM Forum
	Group of standards	Framework
Use	Business area	Appraisal & selection
	Features	MoReq focuses mainly on the functional requirements for the management of electronic records by an Electronic Records Management System (ERMS). The specification is intended to be used by (potential) ERMS users, training organisations, academic institution, ERMS suppliers and developers, record management services, and potential users of outsourced record management services.
	Area of use	Europe
Evolution	Standard maturity	The latest edition of the MoReq® specification is MoReq2010

Table 75: MoReq Family

ISO 23081

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Information & Documentation - Records Management Processes - Metadata for Records (ISO 23081)
	URL	https://www.iso.org/standard/57121.html http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_for_archives_iso_23081.pdf
	Basic description	ISO 23081 Information and documentation – Records management processes – Metadata for records. This standard comprises two parts: Part 1 Principles (2006) and Part 2 Conceptual and implementation issues (2009). The standard outlines the principles that govern records management metadata and establishes a framework in which to create, manage, and use records management metadata elements.
	Owner	ISO
	Group of standards	ISO Standard - Framework

Category	Evaluation criteria	Evaluation criteria values
Use	Business area	Appraisal & selection
	Features	ISO 23081 Principles (Part 1) covers those principles that support and govern records management metadata while (Part 2) Conceptual and implementation issues establishes a framework for defining metadata elements consistent with the principles and implementation considerations. The ISO 23081:2006 (Part 1) is linked to ISO 15489-1 by outlining the five types of metadata required to support the records management process. Metadata about a) a record itself; b) business rules or policies and mandates; c) agents; d) business activities or processes; and e) record management processes.
	Area of use	International
Evolution	Standard maturity	ISO 23081-1:2006 replaces ISO/ TS 23081:2004 and ISO 23081-2:2009 replaces ISO 23081-2:2007. Last edition 2011

Table 76: ISO 23081

11.5 PRESERVATION

Digitisation processes and long-term preservation systems are related with specific technical metadata, preservation metadata and structural metadata, and are necessary to manage, preserve and provide access to digital assets.

Within the OAIS model, three types of information packages are identified.

- Submission Information Package (SIP), which is sent from the information producer to the archive;
- Archive Information Package (AIP), which is the information package actually stored by the archive; and
- Dissemination Information Package (DIP), which is the information package transferred from the archive in response to a request by a consumer.

A consumer accesses information (packages) stored within OAIS system and executes searches provided by Finding Aid Tools or other access systems. This involves the provision of a user interface to the archive's holdings for both search and retrieval purposes.

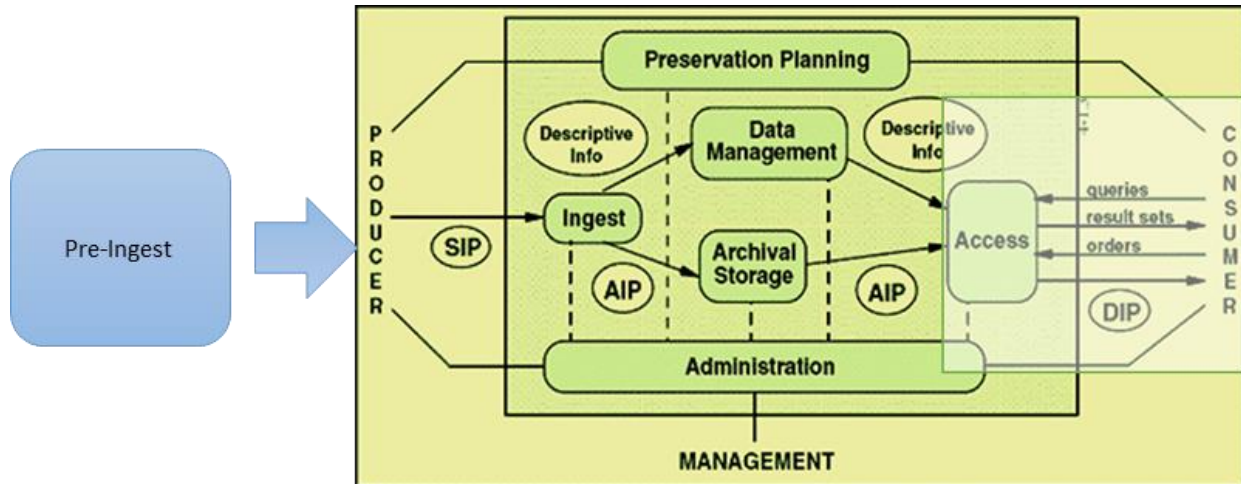


Figure 60: Interaction with OAIS

Internally, the Access Module of OAIS system helps consumers to identify and obtain descriptions of relevant information in the archive, and delivers information from the archive to consumers.

The Access Module generates a DIP in response to a user request by obtaining copies of the appropriate AIP(s) from Archival Storage; obtaining relevant descriptive information from Data Management in response to a query; and finally, delivering the DIP or query result set to consumers.

ISO 14721:2012

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO 14721:2012 (CCSDSS 650.0-P-1.1) Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model
	URL	https://www.iso.org/standard/57284.html https://public.ccsds.org/pubs/650x0m2.pdf
	Basic description	Open Archival Information System (OAIS) is known as a reference model, defining concepts and responsibilities essential for ensuring preservation of digital information. The reference model addresses a full range of archival information preservation functions including ingest, archival storage, data management, access, and dissemination. It also addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among Archives.
	Owner	ISO

Category	Evaluation criteria	Evaluation criteria values
	Group of standards	ISO Standard
Use	Business area	Pre-ingest, Ingest, Access, Publication
	Features	<p>The OAIS reference model provides a framework for the understanding and increased awareness of archival concepts needed for long-term digital information preservation and access; provides the concepts needed by non-archival organisations to be effective participants in the preservation process; provides a framework, including terminology and concepts, for describing and comparing architectures and operations of existing and future Archives; provides a framework for describing and comparing different long-term preservation strategies and techniques; provides a basis for comparing the data models of digital information preserved by Archives and for discussing how data models and the underlying information may change over time; provides a framework that may be expanded by other efforts to cover long-term preservation of information that is NOT in digital form (e.g., physical media and physical samples); expands consensus on the elements and processes for long-term digital information preservation and access and promotes a larger market which vendors can support; and guides the identification and production of OAIS-related standards.</p>
		<p>The diagram illustrates the OAIS Reference Model. It is a central box containing several components. At the top is 'Preservation Planning'. Below it are 'Data Management' and 'Access'. In the middle is 'Archival Storage'. At the bottom is 'Administration'. On the left side, 'Ingest' is shown with 'SIP' (Source Information Package) entering it. 'Descriptive Info' is shown entering 'Data Management'. 'AIP' (Archival Information Package) is shown entering 'Archival Storage'. On the right side, 'Access' is shown with 'queries' and 'result sets' going to the 'CONSUMER', and 'orders' going from the 'CONSUMER' to 'Access'. 'DIP' (Destination Information Package) is shown exiting 'Access'. The entire model is supported by 'MANAGEMENT' at the bottom.</p>
	Area of use	International
Evolution	Standard maturity	<p>OAIS has become the standard model for digital preservation systems at many institutions and organisations. Last edition 2012</p>

Table 77: OAIS (ISO 14721)

ISO 20652:2006

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO 20652:2006 (CCSDS 651.0-B-1:2004) Space data and information transfer systems -- Producer-archive interface -- Methodology abstract standard
	URL	https://www.iso.org/standard/39577.html http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_for_archives_paimas.pdf
	Basic description	PAIMAS looks in detail at the accessioning process, primarily for digital materials, and the actions which should be included in the workflow for it to be undertaken adequately and successfully.
	Owner	ISO
	Group of standards	ISO Standard
Use	Business area	Pre-ingest, Ingest, Access, Publication
	Features	PAIMAS identifies and defines specific actions for four phases in the interactions between the Producer and the Archive, which should be included when preparing and acting on the agreement. 1). Preliminary Phase (pre-ingest); 2) Formal Definition Phase; 3) Transfer Phase; and 4) Validation Phase. PAIMAS considers accessioning in terms of the first phases of the Ingest Function and the Administration Function of the OAIS Reference Model (Open Archival Information Systems Reference Model) and the transferral of material from the Producer to an OAIS compliant Archive for preparation for management and storage.
	Area of use	Preservation
Evolution	Standard maturity	PAIMAS was developed by the Consultative Committee for Space Data Systems (CCSDS), which also developed OAIS.
	Development status	2006 current version. Last reviewed and confirmed in 2014

Table 78: PAIMAS (ISO 20652)

PREMIS

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Data Dictionary for Preservation Metadata (PREMIS)
	URL	https://www.loc.gov/standards/premis/ https://www2.archivists.org/groups/standards-

Category	Evaluation criteria	Evaluation criteria values
		committee/preservation-metadata-implementation-strategies-premis
	Basic description	Preservation Metadata Implementation Strategies (PREMIS) is a data dictionary and XML Schema for the encoding of information necessary to support the digital preservation process. Its data elements are divided into five categories, reflecting information on the PREMIS container, objects, events, agenda, and rights.
	Owner	Library of Congress
	Group of standards	Data Content Standard
Use	Business area	Pre-ingest, Ingest, Access
	Features	<p>The main aims of PREMIS 3.0 are overcoming interoperability problems due to different underlying data models of digital repositories. This is with the purpose of gaining a deeper use of preservation metadata, empowering services for cross-repository search, and helping stakeholders in improving the management of preservation metadata.</p> <p>A key feature of the PREMIS model (considering immutable objects) is the definition of Objects as made up of Representations, Files, and Bit streams.</p> <p>The PREMIS Data Dictionary <i>defines semantic units</i>.</p>
	Area of use	Preservation
Evolution	Standard maturity	PREMIS latest (3.0) edition (2015) mainly introduced three main improvements: The “technical” facilitation of the semantic expression; the inclusion of external vocabularies for PREMIS semantic units using, whenever possible, external vocabularies; and guidelines and mappings from PREMIS semantic units to RDF constructs.

Table 79: PREMIS

METS

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Metadata Encoding & Transmission Standard (METS)
	URL	http://www.loc.gov/standards/mets/
	Basic description	The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the

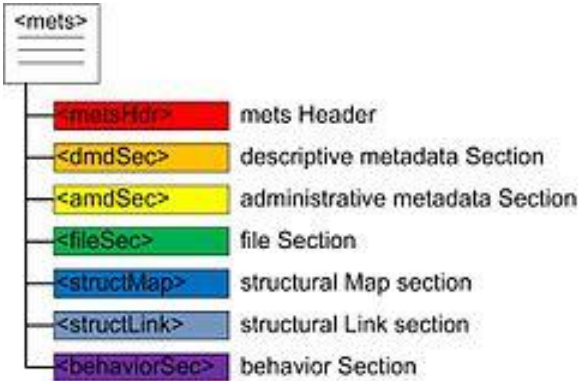
Category	Evaluation criteria	Evaluation criteria values
		World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation.
	Owner	Library of Congress
	Group of standards	Data structure standard
Use	Business area	Pre-ingest, Ingest, Access
	Features	<p>Metadata Encoding and Transmission Standard (METS) is an XML metadata standard intended to package all the information needed to represent a complex object, including both primary files and metadata that describes them. It defines its own structure for representing files and the relationships between them, and allows embedding or referencing descriptive, technical, rights, source, and digital provenance metadata defined by other schemas. METS has various levels of support in digital asset management systems.</p> 
	Area of use	International
Evolution	Standard maturity	<p>METS 1.x schema is a maturity data model and new registered METS profiles exist, the latest from 2017.</p> <p>METS data models are used by commercial vendors, and key institutional users (e.g., ccs:docWorks, Ex Libris, International Image Interoperability Framework, various academic and national libraries, open source and library-based developers).</p> <p>METS 1.x schema, 2015 last revision.</p> <p>Draft METS 2.0 data model. METS 2.0 data model is designed to be compatible with Semantic web and other technologies.</p>

Table 80: METS

MODS

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Metadata Object Description Schema (MODS)
	URL	http://www.loc.gov/standards/mods/
	Basic description	Metadata Object Description Schema (MODS) is an XML-based bibliographic description schema which was designed as a compromise between the complexity of the MARC format used by libraries and the extreme simplicity of Dublin Core metadata.
	Owner	Library of Congress
	Group of Standards	Data structure standard
Use	Business area	Pre-ingest, Ingest, Access
	Features	<p>MODS could potentially be used as follows: a) an SRU specified format; b) an extension schema to METS; c) to represent metadata for harvesting; d) for original resource description in XML syntax; e) for representing a simplified MARC record in XML; and f) for metadata in XML that may be packaged with an electronic resource.</p> <p>MODS is frequently used as a descriptive metadata structure standard inside METS metadata wrappers for storage or exchange of digital objects.</p> <p>The standard is MARC-compatible metadata format expressed in XML and using language-based element names.</p>
	Area of use	International
Evolution	Standard maturity	<p>Metadata Object Description Schema (MODS) is a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress with input from users.</p> <p>MODS 3.6 is the current version of the schema.</p> <p>MODS 3.6 Changes from version 3.5 is available.</p> <p>Recent draft MODS 3.7 schema</p>

Table 81: MODS

LoC Format Guidelines

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Digital Preservation at the Library of Congress
	URL	https://www.loc.gov/preservation/digital/ https://www.loc.gov/preservation/digital/formats/index.html https://www.loc.gov/preservation/resources/rfs/TOC.html
	Basic description	List of guidelines related to digital content packaging and ingest, monitoring and reporting of digital storage, sustainable digital file formats, metadata and more.
	Owner	Library of Congress
	Group of Standards	Technical metadata standard
Use	Business area	Pre-ingest, Ingest, Access, Publication
	Features	<p>Sustainability of Digital Formats website:</p> <ul style="list-style-type: none"> Provides in-depth descriptions of over 440 formats sorted into content categories including: still image, sound, textual, moving image, Web archive, datasets, geospatial and generic formats with more to come. Important features of the format descriptions include documenting relationships between formats and factors to consider when evaluating formats including sustainability factors and quality and functionality factors. <p>Recommended Formats Statement: Identifies hierarchies of the physical and technical characteristics of creative formats, both analogue and digital, which will best meet the needs of all concerned, maximising the chances for survival and continued accessibility of creative content well into the future.</p>
	Area of use	International
Evolution	Standard maturity	<p>Preservation is a top priority of the Library of Congress. The list of guidelines is maintained and updated.</p> <p>Reviewed continuously, the Library of Congress answers dozens of questions every week from the public and from colleagues at other institutions in the U.S. and abroad through the Ask-A-Librarian online reference service. The Preservation FAQ and the design of the Library's preservation webpages are based on these inquiries.</p>

Table 82: LoC Format Guidelines

FADGI Federal Agencies Digital Guidelines Initiative

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	FADGI Federal Agencies Digital Guidelines Initiative
	URL	https://www.loc.gov/preservation/digital/
	Basic description	Established in 2007, FADGI is a collaborative effort of 20 federal agencies to articulate common sustainable practices and guidelines for digitised and born digital historical, archival and cultural content.
	Owner	Library of Congress
	Group of Standards	Technical metadata standard
Use	Business area	Pre-ingest, Ingest, Access, Publication
	Features	<p>Recognising that the effort would require specialised expertise, two separate working groups were formed with the possibility that more tightly focused groups might be necessary as the work progressed. The Still Image Working Group concentrates its efforts on image content such as books, manuscripts, maps, and photographic prints and negatives. The Audio-Visual Working Group focuses its work on sound, video, and motion picture film.</p> <p>Still Image Working Group:</p> <ul style="list-style-type: none"> This group is involved in a cooperative effort to develop common digitisation guidelines for historical and cultural materials that can be reproduced as still images, such as textual content, maps, photographic prints and negatives. The overall goal is to enhance the exchange of research results and development, encourage collaborative digitisation practices and projects among federal agencies, and provide the public with a product of uniform quality. It also provides a common set of benchmarks for digitisation service providers and manufacturers. In addition to digital imaging and encoding, guidelines for the metadata that is embedded in digital image files have been established. <p>Audio Visual Working Group:</p> <ul style="list-style-type: none"> The goal for this working group is to identify, establish, and disseminate information about common sustainable technical guidelines, methods, and practices for digitised and born digital historical, archival and cultural content. The effort will cover sound recordings, video recordings, motion picture film, and born digital content. Topic areas include formatting, digital file formats, metadata, methodology, and, when appropriate, development of open source tools to facilitate

Category	Evaluation criteria	Evaluation criteria values
		workflows.
	Area of use	USA
Evolution	Standard maturity	<p>Created in 2007. Last entry at September, 2017.</p> <p>Long known as FADGI, the acronym's meaning was updated in 2017 from Federal Agencies Digitisation Guidelines Initiative to the Federal Agencies Digital Guidelines Initiative to reflect this growing area of work.</p> <p>The participating agencies share the belief that common guidelines will enhance the exchange of research results and developments, encourage collaborative practices and projects for digital material among federal agencies and institutions and provide the public with a product of uniform quality. They will also serve to set common benchmarks for service providers and manufacturers.</p>

Table 83: FADGI Federal Agencies Digital Guidelines Initiative

DCC Curation Reference Manual

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Curation Reference Manual
	URL	http://www.dcc.ac.uk/resources/curation-reference-manual
	Basic description	<p>Curation Reference Manual contains advice, in-depth information and criticism on current digital curation techniques and best practice.</p> <p>The manual is designed to help data custodians, producers and users better understand the challenges they face and the roles that they play in creating, managing and preserving digital information over time. For each topic covered, suggestions for best practice and real life examples are given.</p>
	Owner	DCC
	Group Standards	of Best practices
Use	Business area	Appraisal & selection, Arrangement & description, Pre-ingest, Ingest, Access, Publication
	Features	Based on DCC Curation Lifecycle Model that provides a graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt

Category	Evaluation criteria	Evaluation criteria values
		<p>through the iterative curation cycle. The model serves to plan activities within an organisation or consortium to ensure that all of the necessary steps in the curation lifecycle are covered.</p> <p>The manual offers guidelines about: Appraisal & selection; Archival metadata; Archiving web resources; Automated metadata generator; Curating emails; File Formats; Investment in intangible Asset; Metadata; Ontologies; Open Source for Digital Curation; Preservation metadata; Preservation scenarios; Preservation strategies; Microfilming, etc.</p>
	Area of use	International
Evolution	Standard maturity	<p>To ensure that the Curation Reference Manual remains as relevant as possible, the DCC will undertake periodic reviews of the individual instalments and may commission updates where necessary to reflect any major developments in digital curation. As an added quality control, a review panel, comprising a number of international experts in the field of digital curation, reviews each instalment produced.</p> <p>Created at 2004 and currently updated (2017)</p>

Table 84: DCC Curation Reference Manual

DPC Digital Preservation Handbook

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	DPC Digital Preservation Handbook
	URL	http://www.dpconline.org/handbook
	Basic description	The Handbook aims to identify good practice in creating, managing and preserving digital materials. By providing a strategic overview of the key issues, discussion and guidance on

Category	Evaluation criteria	Evaluation criteria values
		strategies and activities, and pointers to key projects and reports, the Handbook aims to provide guidance for institutions and individuals and a range of tools to help them identify and take appropriate actions.
	Owner	DPC http://www.dpconline.org/
	Group of Standards	Best practices
Use	Business area	Appraisal & selection, Pre-ingest, Ingest
	Features	The Handbook offers practical guidance about: Preservation issues, Institutional strategies and policies: audit, certification, legal compliance, risk management, standards and best practices, business cases; Organisational activities: acquisition and appraisal, retention and review, storage, preservation planning, access; Technical solution and tools; Content specific preservation.
	Area of use	International
Evolution	Standard maturity	<p>Founded in 2001, the DPC acts a consortium of those organisations interested in the preservation of digital information. Participation in the coalition is open to all sectors including commercial, cultural heritage, educational, governmental, and research bodies.</p> <p>The Handbook, is maintained and updated by the DPC. This full revision (the 2nd Edition) has expanded and updated content to cover over 30 major sections (see Contents). The 2nd edition was compiled with input from 45 practitioners and experts in digital preservation.</p>

Table 85: DPC Digital Preservation Handbook

E-ARK

The main interest of the E-ARK project for this research study resides in the essential best practices of the pre-ingest phase of data export and normalisation in source systems and metadata standards specifications for information packages related to the management of electronic archives that covers the business processes identified such as Pre-Ingest, Ingest and Publication.

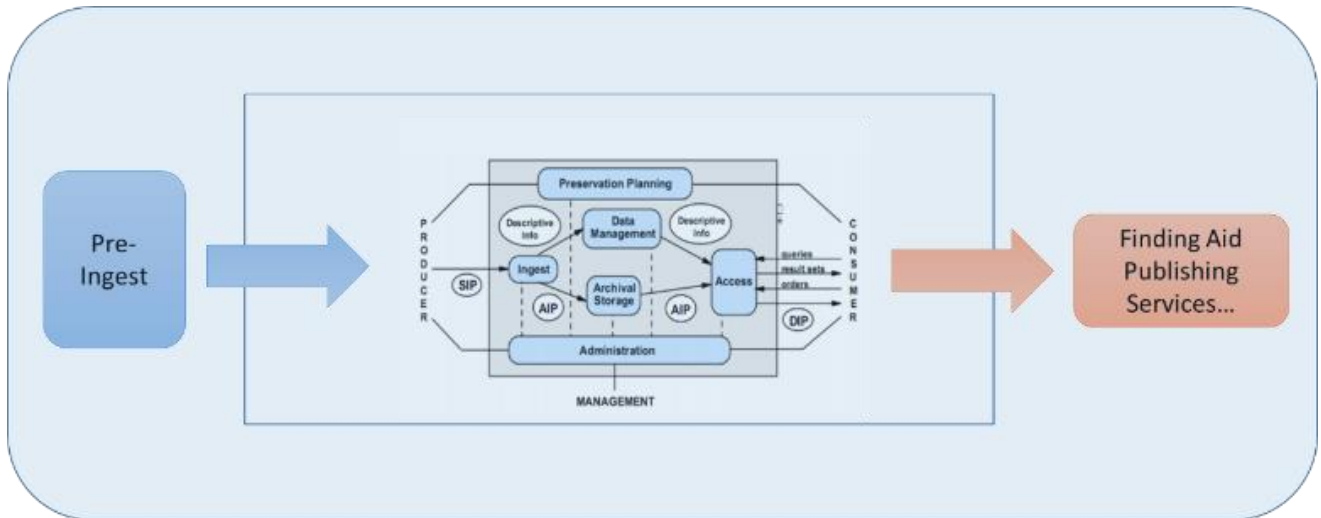


Figure 61: E-ARK project research

E-ARK Information Packages SIP, AIP, DIP

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	D3.1 E-ARK Report on Available Best Practices D3.3 E-ARK SIP Pilot Specification D4.3 E-ARK AIP Specification D5.3 Final Version of DIP Specification
	URL	http://www.eark-project.com/ http://www.eark-project.com/resources/project-deliverables
	Basic description	The set of deliverables provides specifications and examples to help archivists in the definition of Information Packages (SIP, AIP, DIP) defined in the OAIS Framework. Many different formats are used all over the world and unfortunately there is currently no central format for them. The current list of deliverables provides useful specifications.
	Owner	DLM Forum e-ARK Project
	Group of Standards	Technical standard
Use	Business area	Pre-Ingest, Ingest, Access
	Features	Best practices contains a useful appendix of guidelines in the scope of preservation of European national archives (Appendix G: Standards, guidelines and legislation used by stakeholders). E-ARK Submission Information Package (SIP) provides a


Category	Evaluation criteria	Evaluation criteria values
		<p>general structure for Submission Information Packages. SIP specification for the E-ARK project describes in detail and gives an overview of the structure and main metadata elements for E-ARK SIP and provides initial input for the technical implementations of E-ARK ingest tools. Sections of interest are:</p> <ul style="list-style-type: none"> <i>3.1 Metadata elements regarding the information package</i> <i>3.2 Metadata elements regarding submitted files</i> <i>3.3 PREMIS metadata</i> <i>3.4 EAD and EAC-CPF metadata</i> <p>E-ARK AIP pilot specification presents the E-ARK Archival Information Package AIP format specification as it will be used by the pilots (implementations in pilot organisations).</p> <p>E-ARK Dissemination Information Package (DIP) format describes the workflows and use cases of archival access services and ultimately uses these to present a set of requirements which should be followed when designing a DIP format.</p> <p>The secondary aim is to describe the access scenarios in which these DIP formats will be rendered for use.</p> <p>Annexed to this document are two papers:</p> <ul style="list-style-type: none"> • A Proposed E-ARK Standard for Vendor-Independent Archiving of Data Warehouses and • A Proposed E-ARK Standard for Relational Database Metadata Table Structure
	Area of use	UE mainly
Evolution	Standard maturity	February- March 2017 last version. First version on 2016.
	Maintenance status	DLM Forum has acquired the responsibility to help to sustain the project outputs, enhancing their longevity (DLM Forum, 2017). In addition, project partner The Digital Preservation Coalition has promoted best practices in this area (Digital Preservation Coalition, 2017).

11.6 SECURITY



The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) is a Regulation by which the European Commission wishes to strengthen and unify data protection for individuals within the European Union. Security measures around data subjects and protecting personal data are key aims of GDPR.

GENERAL DATA PROTECTION REGULATION EU GDPR

What it consists of and how it affects us


an NTT DATA Company

The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) is a Regulation by which the European Commission wishes **to strengthen and unify data protection for individuals within the European Union**

Main objectives:

- That citizens can take **control** of their own data.
- **Unify** privacy standards throughout the EU.

Scope:

- **Who does it apply to?** Any organization in any industry that stores data from **EU citizens**, within the EU and outside the EU.
- **What applies?** Applies to **personal data** stored as structured and unstructured data ("content": documents, photographs, social network content, ...)
 - Personally identifiable information (PII),
 - Payment Card Industry Security Standard (PCI)
 - Protected health information identification (PHI)

Impact:

- Organizations will have to publish their data protection policies.
- In case of **security breaches** they must inform the affected ones in a maximum of time established by the regulation (72h.)
- Failure to comply with the Regulation is penalized by **finest** of 4% of the annual overall turnover of the preceding year. *Maximum of 20 million €.*

EU GDPR DIGITAL CONTENT

Figure 62: GDPR Regulation Key Topics

Although archival institutions must comply with the GDPR by identifying whether personal data is being used and where it is located, it should be noted that Article 5 of the GDPR explicitly recognises the notion of processing of personal data for archiving purposes in the public interest. Furthermore, Article 89 establishes possible safeguards and derogations relating to the processing for archiving or statistical purposes. At this point, organisations can choose to apply protecting personal data techniques, such as anonymising⁴³ (or redacting) sensitive data properly and/or removing the data if they have no legitimate use for it.

The “ISO27k” suite comprises more than 40 standards related with security techniques, and one of them is specific for performing digital redaction on digital documents.

ISO/IEC 27038

Category	Evaluation criteria	Evaluation criteria values
----------	---------------------	----------------------------

⁴³ Anonymisation is the process of turning data into a form which does not identify individuals and where identification is not likely to take place. This technique enables wider use of personal data and allows the disclosure of data once it has been anonymised. Techniques of anonymisation are: 1-Data masking; 2-Partial data removal; 3-Data quarantining; 3- Pseudonymisation; 4-Aggregation; 5- Cell suppression; 6-Inference Control; 7- Perturbation; 8- Rounding; 9- Sampling; 10- Synthetic data; 11-Tabular reporting; 12-Derived data items and banding. <https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf>

Category	Evaluation criteria	Evaluation criteria values
Basic description	Standard name	Information technology - Security techniques - Specification for digital redaction (ISO/IEC 27038:2014)
	URL	https://www.iso.org/standard/44382.html http://www.iso27001security.com/html/27038.html
	Basic description	ISO/IEC 27038:2014 specifies characteristics of techniques for performing digital redaction on digital documents. It also specifies requirements for software redaction tools and methods of testing that digital redaction has been securely completed. ISO/IEC 27038:2014 does not include the redaction of information from databases.
	Owner	ISO
	Group of standards	ISO Standard
Use	Business area	Sensitive review
	Features	<p>Digital data sometimes has to be revealed to third parties, occasionally even published to the general public, for reasons such as disclosure of official documents under Freedom of Information laws or as evidence in commercial disputes or legal cases. However, where it is deemed inappropriate to disclose certain sensitive data within the files (such as the names or locations of people who must remain anonymous and various other personal or proprietary information that must remain strictly confidential), they must be securely removed from the files prior to their release. 'Redaction' is the conventional term for the process of denying file recipients knowledge of certain sensitive data within the original files.</p> <p>The standard covers:</p> <ul style="list-style-type: none"> - An introduction to the general principles of digital redaction and anonymisation of data; - Redaction requirements - an overview of the redaction process; - Redaction processes such as printing and physically redacting content, editing the original documents in various ways, dealing with metadata (such as document properties and change records) and, in the case of 'enhanced' redaction, considering the broader context as well as the specific content (e.g. the possibility of guessing, inferring or reconstructing redacted content from other content in redacted files, or by using other sources); - Keeping records and notes in order to be able to explain or justify redaction decisions and actions;

Category	Evaluation criteria	Evaluation criteria values
		<ul style="list-style-type: none"> - Software redaction tools - a core set of functional requirements; - Redaction testing - five simple if basic ways to check whether the redaction has been successful; and - An informative annex about redacting PDFs.
	Area of use	International
Evolution	Standard maturity	The standard was published in 2014.

Table 86: ISO/IEC 27038

11.7 STORAGE INFORMATION

BS 4971:2017

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	2017, BS 4971:2017 Conservation and care of archive and library collections
	URL	-
	Basic description	Provides recommendations for the long-term conservation of archive and library collections. Includes policy, strategy and planning, and activities and processes which protect collections, such as the building environment and preventative and remedial treatments.
	Owner	British Standards Institution
	Group of Standards	BS Standard
Use	Business area	Delivery
	Features	Not free available
	Area of use	International
Evolution	Standard maturity	Supersedes BS 4971:2002 and PD 5454:2012.: 2000, References an older version: BS 5454:2000 Recommendations for the storage and exhibition of archival documents and libraries Last version: 2017

ISO/TR 19814:2017

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO/TR 19814:2017 Information and documentation -- Collections management for archives and libraries
	URL	https://www.iso.org/standard/66263.html

Category	Evaluation criteria	Evaluation criteria values
	Basic description	<p>ISO/TR 19814:2017 provides guidance and recommendations in the planning, implementation, maintenance and improvement of the preservation of archive and library collections.</p> <p>ISO/TR 19814:2017 applies to preservation of archive and library physical collections of institutions and volumes small and large. It applies to all collections housed by an institution; their own collections and deposits or loans from other institutions. Some information on digital collections, born digital and digitised, for conservation is included for reference.</p> <p>ISO/TR 19814:2017 also applies to collections that are being managed by governmental agencies.</p>
	Owner	ISO
	Group of Standards	ISO Standard
Use	Business area	Delivery
	Features	<p>This guidance is intended for collections that are being preserved for long-term use and includes:</p> <ul style="list-style-type: none"> • recommendations and guidance for preservation planning and ongoing management of physical collections in archives and libraries; • procedures for managing collections in the stacks, research and reading rooms, conservation facilities and while on exhibit and during transportation; • guidance and recommendations for appropriate enclosures and containers for archive and library collections. <p>ISO/TR 19814:2017 covers specifically the operations required to manage the collections environment that are relevant to the preservation policy and plan of the institution. This includes the monitoring of climate stability, control of exposure to light, preventive cleanliness measures and cleaning of the collections storage areas.</p>
	Area of use	International
Evolution	Standard maturity	Last version 2017

ISO 11799:2015

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO 11799:2015 Information and documentation -- Document storage requirements for archive and library materials

Category	Evaluation criteria	Evaluation criteria values
	URL	https://www.iso.org/standard/63810.html
	Basic description	ISO 11799:2015 specifies the characteristics of repositories used for the long-term storage of archive and library materials. It covers the siting and construction and renovation of the building and the installation and equipment to be used both within and around the building.
	Owner	ISO
	Group of Standards	ISO Standard
Use	Business area	Delivery
	Features	<p>This International Standard presents some facts and general rules to be considered when a purpose-built repository is designed, when an old building originally designed for another use is converted, or when a building already in use as repository is renovated, with respect to energy efficiency and sustainable development.</p> <p>This International Standard applies to the long-term storage of archive and library materials for their lifetime. It takes into account that the materials are stored and allow current usage as well.</p>
	Area of use	International
Evolution	Standard maturity	Last version 2015

ISO 14416:2003

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO 14416:2003, Information and documentation – Requirements for binding of books, periodicals, serials and other paper documents for archive and library use – Methods and materials
	URL	https://www.iso.org/standard/20033.html
	Basic description	ISO 14416:2003 is applicable to the binding of books, periodicals and archive documents which have special requirements for durability and permanence.
	Owner	ISO
	Group of Standards	ISO Standard
Use	Business area	Delivery

Category	Evaluation criteria	Evaluation criteria values
	Features	<p>The use, as well as the wear and tear, of library and archive documents varies. The choice of binding method should therefore relate to the appropriate requirements of a specific library or archive. The quality as well as the price of the binding is dependent on this choice.</p> <p>It is applicable to the following general procedures: first-time hard-cover binding of published and unpublished materials, and any other documents requiring this type of protection; rebinding of hard-cover monographs, serials and any other documents.</p> <p>It is not intended for binding volumes identified by a customer as having high artifactual or historical value, or for any volumes that, because of their physical characteristics, cannot or should not be bound according to ISO 14416:2003. Arrangements for special treatments should be made separately.</p>
	Area of use	International
Evolution	Standard maturity	This standard was last reviewed and confirmed in 2014. Therefore this version remains current.

11.8 METRICS AND KEY PERFORMANCE INDICATORS

The purpose of this group of standards is to help archival institutions in the measurement of the quality of access and the monitoring of the services offered to users. They can also provide guidelines allowing comparison between institutions.

ACRL/ RBMS, SAA Release Standardised Statistical Measures and Metrics for Public Services in Archival Repositories and Special Collections Libraries

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	Standardised statistical measures for public services in archival repositories and special collections libraries.
	URL	https://www2.archivists.org/standards/standardized-statistical-measures-and-metrics-for-public-services-in-archival-repositories
	Basic description	The purpose of this standard is to help archival repositories and special collections libraries quantify in meaningful terms the services they provide their constituencies and evaluate the effectiveness and efficiency of the operations that support those services. Furthermore, the aim of the standards is to establish a common and precise vocabulary to facilitate conversations about service metrics and comparative studies between institutions.

Category	Evaluation criteria	Evaluation criteria values
	Owner	The Association of College and Research Libraries (ACRL) Rare Books and Manuscripts Section (RBMS) Society of American Archivists (SAA)
	Group of Standards	NA
Use	Business area	Access
	Features	This standard was developed to provide archivists and special collections librarians with a set of precisely defined, practical measures based upon commonly accepted professional practices that can be used to establish statistical data collection practices to support the assessment of public services and their operational impacts at the local institutional level. The measures were also formulated to support the aggregation of public services data from multiple institutions to provide a basis for institutional comparisons and benchmarking. Public services that can be measured based on guidelines covered by the standard are: user demographics, reference transactions, reading room visits, collection use, events, instruction, exhibitions, and online interactions.
	Area of use	International
Evolution	Standard maturity	This standard was released on October 2017

ISO 11620:2014

Category	Evaluation criteria	Evaluation criteria values
Basic description	Name	ISO 11620:2014 Information and documentation -- Library performance indicators
	URL	https://www.iso.org/standard/56755.html
	Basic description	ISO 11620:2014 offers accepted, tested, and publicly accessible (i.e. non-proprietary) methodologies and approaches to measuring a range of library service performance.
	Owner	ISO
	Group of Standards	ISO Standards
Use	Business area	Access
	Features	The standard aims to facilitate the evaluation of any type of library, through the establishment of a series of performance indicators. It includes a total of 29 indicators, grouped into 3

Category	Evaluation criteria	Evaluation criteria values
		areas: user opinion, public services and technical services. Its main purpose is to promote the knowledge and use of performance indicators in libraries.
	Area of use	International
Evolution	Standard maturity	Last review: October 2014.

11.9 PRACTICES IN OTHER AREAS

The study also examines practices of other areas around digital content and content analysis that might be of interest for the management of digital archives. A short summary of three practises is shown here under.

No standard about these techniques exists currently.

Automatic data classification

Automated data classification can be of two types:

- Text-based classification
- Semantic-based classification

Text-based classification

The techniques called **machine learning**, which fall within the artificial intelligence, are based on the creation of a decision system based on **generalization**, from a **training** set.

- The training process, which is performed through a set of previously categorized textual content, is called **supervised learning**.

D1 = "I like databases"
D2 = "I hate databases"

	I	like	hate	databases
D1	1	1	0	1
D2	1	0	1	1

Document-term matrix



The statistical information derived from the occurrence **frequency of words and their distribution** is used to calculate a relative measure of the meaning of the text. The textual content of a document can be considered as an unordered word vector (**bag of words**), or as a combination of words (**bag of n-grams**).

There are different approaches to implementation of machine learning algorithms, the most outstanding being: Bayesian Networks, Decision Tree, Support Vector Machine, Clustering

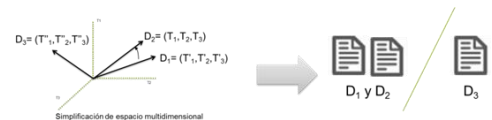
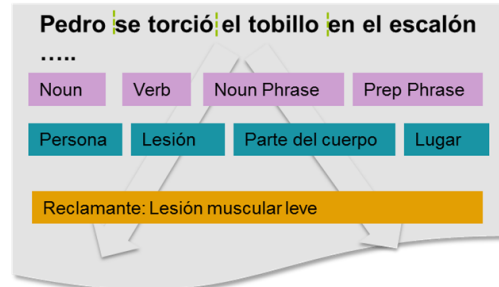


Figure 63: Text- based classification

Semantic-based classification

The techniques of **natural language processing** (NLP) allow the linguistic interpretation of textual content, to be able to process and extract information and associated concepts:

- **Part Of Speech:** Determine the linguistic category of each of the words, such as name, verb, adjective, etc
- **Annotations:** Extraction of information from linguistic rules, regular expressions and dictionary names.
- **Summarization:** Summary of the content of the text from its analysis.



There are several approaches to natural language analysis, those based on rules and those that use machine learning techniques. NLP techniques are **dependent on the language of the text**.

Figure 64: Semantic-based classification

File Analysis

Main features covered	Description
Metadata analysis	Analyse standard & customised metadata
Content awareness	Identification of sensitive data: PII, PCI and PHI data.
Tagging and classification	Ability to tag a file with metadata after metadata analysis. Classify artefacts into groups, simple categories or taxonomies.

Identify & Classify: File Analysis
What is ...?

an NTT DATA Company

Why do we need File Analysis tool? Using FA tools, is possible to automate the analysis and classification of all the data stored in the selected repositories. FA includes techniques for contextualize and identify PII,PCI,PHI best than manual searches.

Main features of FA applications:

- Analyse data in all repositories (inactive enterprise data & dark data, duplicated data, uncontrolled data,...).
- By means of graphical tools – Personal data Map gives us information of what (PII,PCI,PHI) are and where is located
- Give us enough information to take measures.
- Give us the option to take remediation actions of current situation, as migrations or secure deletion of content

Steps of FA tool:

- Identify and Index all personal data
- Organise: auto-classify (personal data map)
- Manage (apply remediation actions)

Figure 65: File Analysis

12 ANNEX C: IT TOOLS AND SOLUTIONS

12.1 AMLAD

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	AMLAD - Advance Museum Library Archive Deposit
URL	http://amlad.es/amlad/
Basic description	NTT DATA's digital archive service AMLAD™ (Advanced Museum Library Archives Deposit) is a service that enable to preserve and disseminate valuable ancient books, manuscripts, prints, visual and audio data of GLAMs (Galleries, Libraries, Archives, and Museums) into digital data, not only for storage, but also for presentation to the public as indexed data searchable from computers, smartphones, and tablets.
Owner	everis – NTT DATA
Category	Preservation system
License and delivery model	Commercial Cloud; on-Premises
Functional & Technical description	
Business processes	Acquisition, Archives Processing, Preservation, Administration, Access, Consultation, Data Exchange
Standards	AMLAD complies with several international standards in the digital archiving field including OAIS (Reference Model for an Open Archival Information System, which is a framework for archiving systems for long term preservation of digital information), Dublin-core, METS, MODS, and EAD own XML schemas. Therefore, in addition to Libraries, Museums, Archives, and other institutions can easily install and use AMLAD. Additionally, standardised linkage protocols such as OAI-PMH and SRU/SRW are available on AMLAD, thus allowing easy linking to external institutions that manage their content using different metadata formats. Authority control and controlled vocabularies have to be customised internally by the tool.
Access management	Support for management of users, groups, roles and access permissions to digital objects
Reporting and statistics	Reporting and statistics about ingestion and internal processes related to SIP, AIP and DIP management

Evaluation criteria	Description of the evaluation criteria
Functional architecture	<p>The tool covers all preservation cycle: SIP generation, AIP transformation and DIP creation.</p>
Migration	Migration support provided by owner
Integration	With Europeana. The tool provides customisable OAI-PMH access (generation and provider) With Finding AID Tools
Support	Provided by owner
Relevant aspects	
Remarks	Specifically for preservation of digital objects.
Strengths	<p>AMLAD provides:</p> <ul style="list-style-type: none"> • Support for Libraries, Museums, Archives • Cross search functions for discovering documents across various GLAM collections (• Handles documents in various formats such as Images, Videos, Audio, and Text • Display high definition images with user friendly interface • Supports various devices including PCs, Tablets, and Smartphones • Configurable XML schema for SIP, AIP, DIP and support to edit or enrich metadata stored in the AIPs
Weaknesses	No connection with archives management system

Table 87: AMLAD

12.2 ARCHIVEMATICA

Evaluation criteria	Description of the evaluation criteria
---------------------	--

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Archivemata
URL	https://www.archivemata.org/
Basic description	<p>Archivemata is a free and open-source digital preservation system (OAIS) that is designed to maintain standards-based, long-term access to collections of digital objects.</p> <p>Archivemata is packaged with the web-based content management system AtoM for access to digital objects.</p> <p>AtoM provides specific archives description support and Finding AID capabilities to the preservation system.</p>
Owner	<p>Artefactual Systems (lead development company) develops free and open software made available under the AGPLv3 open-source software license.</p> <p>Provides support through release management, public technical and user documentation, and community forum support.</p> <p>Provides other services: migration, consulting, hosting, installation, and training.</p>
Category	Preservation system
License and delivery model	Open-source Cloud; on-Premises
Functional & Technical description	
Business Process	Acquisition, Archives Processing, Preservation, Administration, Access, Consultation, Data Exchange
Standards	<p>Archivemata provides an integrated suite of free and open-source tools that allows users to process digital objects from ingest to archival storage and access in compliance with the ISO-OAIS functional model and other digital preservation standards and best practices (PREMIS, METS, DublinCore)</p> <p>Authority control and controlled vocabulary support is provided by AtoM component.</p>
Access management	Capabilities to users, groups, roles and access permissions
Reporting and statistics	Reporting and statistics available
Migration	<p>Migration support provided by owner</p> <p>Pre-ingestion tools developed for the product</p>
Integration	Integration with third-parties: AtoM, DSpace, CONTENTdm, Islandora, LOCKSS, DuraCloud, OpenStack and Archivists' Toolkit.
Support	Provided by the community of users and developers.

Evaluation criteria	Description of the evaluation criteria
	On-line support and documentation official web-sites. Specific support provided by owner.
Relevant aspects	
Remarks	Wide community of users, archivist and developers. Updated continuously.
Strengths	Support for pre-ingestion, all preservation cycle (SIP, AIP, DIP). Support for archival description and arrangement. Archival appraisal at multiple decision points:
Weaknesses	N/A

Table 88: Archivematica

12.3 PRESERVICA

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Preservica
URL	https://preservica.com/
Basic description	Preservica's standards-based (OAIS ISO 14721) active preservation software combines all the critical capabilities of successful long-term digital preservation into a single integrated platform. It keeps content safely stored, makes sure it can be found and trusted, provides secure immediate access, and automatically updates files to future-friendly formats.
Owner	Preservica. Preservica offers consulting and research services. It serves archives, libraries, government organisations, museums, and businesses in the United States and internationally. The company is based in Boston, Massachusetts. Preservica operates as a subsidiary of Tessella Ltd.
Category	Preservation system
License and delivery model	Commercial Cloud / on-Premises
Functional & Technical description	
Business Process	Acquisition, Archives Processing, Preservation, Administration, Access, Consultation, Data Exchange
Standards	METS, EAD, MODS, Dublin Core and own XML schemas. Preservica provides a service to help build other custom mappings.

Evaluation criteria	Description of the evaluation criteria
	Preservica uses its own schema, XIP
Access management	Management of users, roles and groups
Reporting and statistics	Capabilities to users, groups, roles and access permissions
Migration	Migration support provided by owner. Pre-ingestion tools developed for the product.
Integration	With ECM systems, archives management systems. Pre-ingestion out-of –the box tools provided too.
Support	Provided by owner
Relevant aspects	
Remarks	Easily manage, enrich and edit metadata over time.
Strengths	Preservica provides a rich set of out-of-the-box connectors and tools to simplify and automate the task of ingesting content from multiple sources and systems. Flexible options enable you to quickly add single records, a continuous stream of digital assets or very large volumes. Connectivity to common ECM systems such as Microsoft SharePoint, or synchronisation with popular catalogues such as ArchivesSpace, CALM and Adlib.
Weaknesses	N/A

Table 89: Preservica

12.4 RODA

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	RODA
URL	https://demo.roda-community.org/#welcome
Basic description	RODA is a complete digital repository solution that delivers functionality for all the main functional units of the OAIS reference model. RODA is capable of ingesting, managing and providing access to the various types of digital content produced by large corporations or public bodies. RODA is based on open-source technologies and is supported by existing standards such as the Open Archival Information System (OAIS), Metadata Encoding and Transmission Standard (METS), Encoded Archival Description (EAD), Dublin Core (DC) and PREMIS (Preservation Metadata).

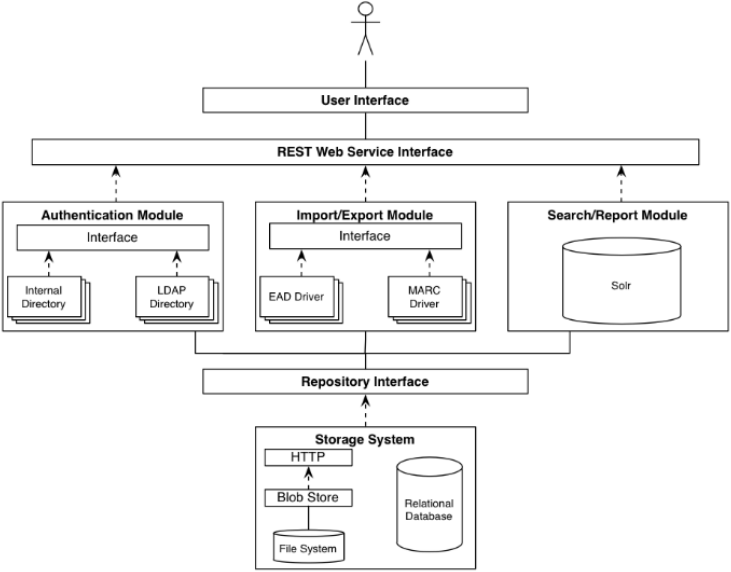
Evaluation criteria	Description of the evaluation criteria
Owner	KEEP SOLUTIONS is a European company that provides advanced services for managing and preserving digital information. The company initiated its activity in 2008 and attained the status of spin-off of the University of Minho for being an enterprise that maintains close ties to research centres and departments of this university.
Category	Preservation system
License and delivery model	Open-source On-premises, Cloud
Functional & Technical description	
Business Process	Acquisition, Archives Processing, Preservation, Administration, Access, Consultation
Standards	RODA follows open standards using EAD for description metadata, PREMIS for preservation metadata, METS for structural metadata, and several standards for technical metadata (e.g. NISO Z39.87 for digital still images).
Access management	Users must be authenticated before accessing the repository. All user actions are logged for future accountability. Permissions are granular and can be defined at repository level, all the way down to individual data objects.
Reporting and statistics	Provides: Activity log, notifications and statistics
Functional architecture and Infrastructure platform	Based on OAIS Model
Migration	This service entails the extraction, transformation and transference of data from legacy systems to the newly implemented system.
Integration	RODA exposes all its functionality via well-documented REST services API. Convenient Java libraries are available on GitHub to allow developers to interact with RODA via its Core APIs. Several tools exist to create and manipulate the SIPs and submit them to RODA's ingest workflow.
Relevant aspects	
Remarks	<ul style="list-style-type: none"> • Scalability • Open-source including libraries to manipulate SIP packages (RODA is licensed under LGPLv3 for all source-code including interoperability libraries like SIP manipulation libraries.)
Strengths	<ul style="list-style-type: none"> • Vendor independent • API for developers (using GitHub) • Advanced ingest workflow • Integration with 3rd party software • Job execution module to create, execute and control preservation

Evaluation criteria	Description of the evaluation criteria
	workflows
Weaknesses	-

Table 90: RODA

12.5 ARCHIVESPACE

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	ArchivesSpace
URL	http://archivesspace.org/
Basic description	ArchivesSpace™ is an open source, web application for managing archives information. The application is designed to support core functions in archives administration such as accessioning; description and arrangement of processed materials including analogue, hybrid, and born-digital content; management of authorities (agents and subjects) and rights; and reference service. The application supports collection management through collection management records, tracking of events, and a growing number of administrative reports. The application also functions as a metadata authoring tool, enabling the generation of EAD, MARCXML, MODS, Dublin Core, and METS formatted data.
Owner	ArchivesSpace is an open-source software. A community of archivists and developers support it. LYRASIS is the organisational home for ArchivesSpace.
Category	Archives management system
License and delivery model	Open-source On-premises
Functional & Technical description	
Business Process	Acquisition, Archives Processing, Preservation, Access, Consultation
Standards	EAD, MARCXML, MODS, Dublin Core, and METS formatted data.
Access management	The ArchivesSpace backend enforces access control, defining which users are allowed to create, read, update, suppress and delete the records in the system. The major actors in the permissions model are: <ul style="list-style-type: none"> • Repositories -- The main mechanism for partitioning the ArchivesSpace system. • Users -- An entity that uses the system--often a person, but perhaps a consumer of the ArchivesSpace API. The set of users is global to the

Evaluation criteria	Description of the evaluation criteria
	<p>system, and a single user may have access to multiple repositories.</p> <ul style="list-style-type: none"> Records -- A unit of information in the system. Some records are global (existing outside of any given repository), while some are repository-scoped (belonging to a single repository).
Reporting and statistics	Tracking of events, and a growing number of administrative reports.
Functional architecture and Infrastructure platform	<p>ArchivesSpace is divided into several components:</p> <ul style="list-style-type: none"> the backend, which exposes the major workflows and data types of the system via a REST API, a staff (internal users) interface, a public interface, and a search system, consisting of Solr and an indexer application. <p>(All these components interact by exchanging JSON data).</p> <ul style="list-style-type: none"> the product provides ArchivesSpace RESTful API for developers. <p>ArchivesSpace has been tested on Ubuntu Linux, Mac OS X, and Windows. MySQL is not required, but is strongly recommended for production use.</p>  <p>The diagram illustrates the ArchivesSpace architecture. At the top, a stick figure represents a user interacting with the 'User Interface'. Below this is the 'REST Web Service Interface'. Three modules are connected to the REST interface: the 'Authentication Module Interface' (which includes 'Internal Directory' and 'LDAP Directory'), the 'Import/Export Module Interface' (which includes 'EAD Driver' and 'MARC Driver'), and the 'Search/Report Module' (which includes 'Solr'). These modules all connect to a 'Repository Interface'. The 'Repository Interface' connects to a 'Storage System' which includes 'HTTP', 'Blob Store', and 'File System' components, as well as a 'Relational Database'.</p> <p>Figure 66: ArchiveSpace Architecture</p> <p>Supports scalability based on volume and number of users. Type supported (based on volume and typology of archive institution): very small; small; medium; large; very large.</p>
Migration	<p>The tool has specific migration tools from migration of legacy sources such as Archon and Archivists Toolkit.</p> <p>Regarding to metadata mapping, the tool provide data import and export mapping schemas.</p>
Integration	<p>Integration with LDAP authentication (security and users)</p> <p>Integration with Web content providers by OAI_PMH:</p>

Evaluation criteria	Description of the evaluation criteria
	<ul style="list-style-type: none"> The tool provides OAI-PMH interface allowing other systems to harvest records. The system provides responses to a number of standard OAI-PMH requests <p>Integration with finding AID tools:</p> <ul style="list-style-type: none"> The digital objects module can be used to describe digital objects and link to digital files stored elsewhere. The metadata created can be exported to other systems as MODS, METS, or Dublin Core or made publicly accessible through the built-in public interface. <p>Integration with Preservation systems with specific connectors:</p> <ul style="list-style-type: none"> <u>ArchivesSpace-Preservica Integration:</u> Synchronises metadata and hierarchy between ArchivesSpace and Preservica, during and at any time after ingest. <u>ArchivesSpace-Archivematica Integration:</u> Pairs Archivematica digital objects with ArchivesSpace Resources and Archival Objects and automatically generates Digital Objects in ArchivesSpace.
Relevant aspects	
Remarks	<p>Resources to the community of developers:</p> <p>https://archivesspace.atlassian.net/wiki/spaces/ADC/overview https://github.com/archivesspace/archivesspace http://archivesspace.github.io/archivesspace/ (http://mw2013.museumsandtheweb.com/paper/archivesspace-a-next-generation-archives-management-system/</p> <p>Regarding updates, 12 releases. June 2009 until September 2017 (current version)</p> <p>Regarding to hosting services, data itself can be stored locally or in the cloud.</p>
Strengths	Integration features (described above)
Weaknesses	<p>ArchivesSpace is not a digital asset or document management system and cannot manage digital files or digitisation workflow.</p> <p>The viewers in the public interface are more limited in their functionality than those of a digital asset management system or digital repository</p>

Table 91: ArchivesSpace

12.6 ATOM

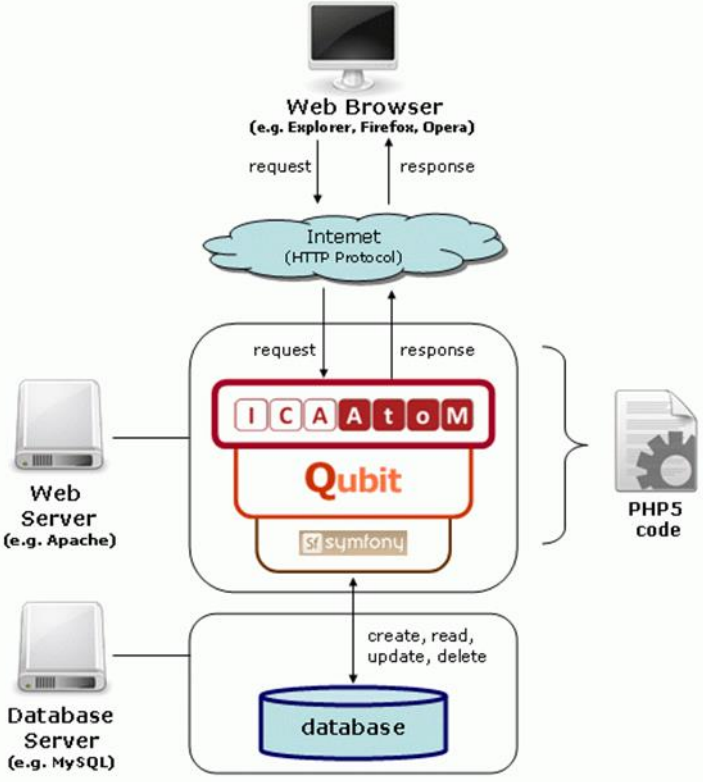
Evaluation criteria	Description of the evaluation criteria
Basic description	

Evaluation criteria	Description of the evaluation criteria
Name	AtoM
URL	https://www.accesstomemory.org
Basic description	<p>AtoM - Access to Memory is a web-based, open source application for standards-based archival description and access in a multilingual, multi-repository environment.</p> <p>It was originally commissioned by the International Council on Archives to make it easy for archival institutions worldwide to put their archival holdings online using the ICA's descriptive standards.</p>
Owner	<p>Open-source supported by Artefactual Systems (lead development company) develops free and open software made available under the AGPLv3 open-source software license. Provides support through release management, public technical and user documentation, and community forum support.</p>
Category	Archives management system
License and delivery model	<p>Open- source On premises</p>
Functional & Technical description	
Business Process	Archives Processing, Access, Consultation
Standards	<p>AtoM was originally built around International Council on Archives (ICA) descriptive standards and supports them:</p> <ul style="list-style-type: none"> • General International Standard Archival Description (ISAD) - 2nd edition, 1999 • International Standard Archival Authority Record (Corporate bodies, Persons, Families) (ISAAR) - 2nd edition, 2003 • International Standard For Describing Institutions with Archival Holdings (ISDIAH) - 1st edition, March 2008 • International Standard For Describing Functions (ISDF) - 1st edition, May 2007 <p>Other (non-ICA) descriptive standards currently supported:</p> <ul style="list-style-type: none"> • RAD is maintained by the Canadian Council of Archives⁴⁴ • DACS is maintained by the Society of American Archivists⁴⁵ • Implements unqualified Dublin Core Metadata Element Set, Version 1.1⁴⁶. <p>MODS standard, maintained by the US Library of Congress Network Development and MARC Standards Office⁴⁷. Supports the Digital Library</p>

⁴⁴ Canadian Council of Archives: <http://cdncouncilarchives.ca/>

⁴⁵ Society of American Archivists: <https://www2.archivists.org/>

⁴⁶ Dublin Core Metadata Initiative: <http://dublincore.org/documents/dces/>

Evaluation criteria	Description of the evaluation criteria
	<p>Federation second level of adoption for MODS. In the scope of Link Data, SKOS standard is supported. Provides support to authority records and terms are treated as entity types, as controlled vocabularies.</p>
Access management	Management of users/roles/ permissions.
Reporting and statistics	Tracking of events, and a growing number of administrative reports.
Infrastructure platform	 <p>AtoM supports scalability based on volume and number of users: its components and services could be deployed in a distributed manner (across multiple machines in a network) in order to accept an escalating number of users.</p>
Migration	N/A
Integration	<p>AtoM has the ability to act as a data provider to service providers interested in harvesting descriptive metadata from AtoM, by exposing simple Dublin Core XML via OAI-PMH. AtoM is integrated with several archives management system and Preservation system.</p>
Support	GitHub and User Forum.

⁴⁷ <http://www.loc.gov/marc/ndmso.html>

Evaluation criteria	Description of the evaluation criteria
	Provided on-line and specific support by Artefactual Systems
Relevant aspects	
Remarks	The tool is specific for archival description and publication. Furthermore, as a publication tool it supports import/export terms with SKOS and provides API for the access to the datasets stored.
Strengths	AtoM is designed to be flexible enough for adaptation to other descriptive standards, supporting a wide range of archival description standards and mapping resources between them (crosswalks).
Weaknesses	N/A

Table 92: AtoM

12.7 CUADRASTAR

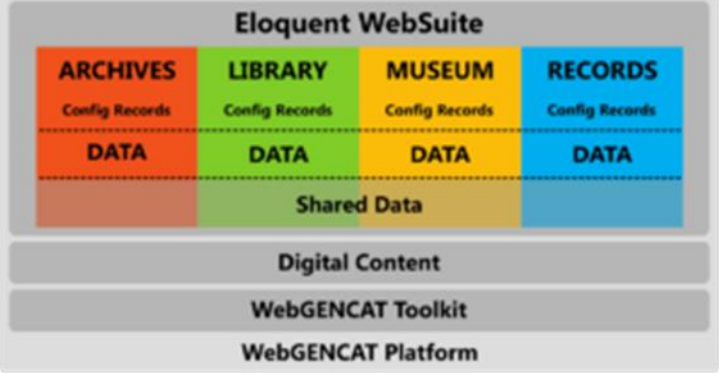
Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	CuadraSTAR
URL	https://lucidea.com/cuadrastar-skca/archives-management-software/http://amlad.es/amlad/
Basic description	<p>Composed for a set of IT tools:</p> <ul style="list-style-type: none"> • CuadraSTAR Knowledge Centre for Archives (SKCA) is a powerful solution for making archival collections visible and accessible. SKCA supports accession, catalogue, and management fonds/collections capabilities, EAD finding aids, MARC records, container lists, and management reports. • STAR Knowledge Centre for Libraries (SKCL): OIPAC and ILS system • STAR Museums: support for the full range of museum management processes, including tasks such as conservation, exhibition planning, and shipping. • STAR/RIMS: Highly customisable records management software that supports the processes that ensure records are well managed, including creation and maintenance of retention schedules, filing plans, inventory integration, request management, and circulation. • STAR Thesaurus: creation and maintenance of thesaurus.
Owner	Lucidea is in the market for over 30 years and provides applications and business process expertise that empowers information intensive organisations to easily collect, organise and leverage important knowledge assets. Its strategy has included growing both organically and via acquisitions, and has enabled Lucidea to achieve double-digit growth and

Evaluation criteria	Description of the evaluation criteria
	build a global customer base serving more than 3,000 clients in more than 50 countries.
Category	Archives management system Library and collection management system
License and delivery model	Commercial On-premises
Functional & Technical description	
Business Process	Acquisition, Archives Processing, Preservation, Access, Consultation
Standards	SKCA complies with the DACS, EAD, MARC, ISAD (G), and Dublin Core archival standards. It also provides flexible support within the full hierarchical model—collections, series, containers, and items.
Access management	Users, groups and permissions. Includes a secure public access module that makes easy for users to search within and across collections and to access both descriptions and relevant digital files.
Reporting and statistics	Produces management reports that help monitor the workload, the locations of materials, and other statistical data. Users can choose from hundreds of pre-defined reports or create their own reports easily, using own built-in capabilities.
Migration	By the owner
Integration	With the rest of IT tools of the owner and with external systems.
Support	On-line documentation and free courses. Specific support by owner.
Relevant aspects	
Remarks	Provides programming interfaces allowing the customer to integrate the tool into other environments.
Strengths	Support to traditional archival processes and a powerful search/retrieve tool. The tool is developed to serve multiple information management needs in libraries, information centres, archives, museums, records centres, and publishing organisations.
Weaknesses	No integration with external preservation systems.

Table 93: CuadraSTAR

12.8 ELOQUENT ARCHIVES

Evaluation criteria	Description of the evaluation criteria
---------------------	--

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Eloquent Archives
URL	https://www.eloquent-systems.com/products/archives/
Basic description	
Owner	<p>Eloquent Systems is a company specialised in software for Archives, Libraries, Museums and Records Management providing to customers with leading-edge software for over 40 years.</p> <p>Lucidea has recently acquired Eloquent Systems https://lucidea.com/lucidea-acquires-eloquent-systems/</p>
Category	<p>Archives management system</p> <p>Library and collection management system</p>
License and delivery model	<p>Commercial</p> <p>Cloud & On-premises</p>
Functional & Technical description	
Business Process	Acquisition, Archives Processing, Preservation, Access, Consultation
Standards	<p>Archive standards DACS, ISAD(G) and RAD</p> <p>EAD</p>
Access management	N/A
Reporting and statistics	N/A
Functional architecture	 <p>The diagram illustrates the functional architecture of Eloquent WebSuite. It is structured as follows:</p> <ul style="list-style-type: none"> Eloquent WebSuite (Overall Suite) Four main functional modules: ARCHIVES (red), LIBRARY (green), MUSEUM (yellow), and RECORDS (blue). Each module includes Config Records and DATA. Shared Data layer connecting the modules. Digital Content layer. WebGENCAT Toolkit layer. WebGENCAT Platform (Foundational layer).
Migration	Data migration and implementation services provided by the owner
Integration	<p>The tool provides several types of Finding AIDs for integration with external systems or with Google search engine:</p> <ul style="list-style-type: none"> • Module of HTML PUBLISHING - INDEX This module creates another access point for researchers. It contains a short name for each descriptive record and initially displays only the top-level records. Researchers open them to go down the hierarchy.

Evaluation criteria	Description of the evaluation criteria
	<p>Clicking on any name will invoke a detail display for the descriptive record. The published HTML index is hosted where it becomes accessible to Google and other search engines. Descriptive detail and digital content is retrieved from the dynamic Eloquent Archives database which is secure and not accessible to Google and other search engines.</p> <ul style="list-style-type: none"> • Module of HTML PUBLISHING – PAGES A static HTML finding aid can be published for posting on any server, making the data accessible to Google and other search engines. You custom tailor the publisher for a 2- or 3-level tree index on the left and descriptive data for selected items on the right. The detail on the right may include links to digital content and to other detail at lower levels in the Eloquent Archives database. • Module for EAD EXPORT This module is for exporting descriptive records in EAD/XML format for import into another system. Exported contain the hyperlink back into the Eloquent Archives database for access to the lower levels. The records are ready to load into a library system or other federated archives database. Researchers in those systems will get a seamless flow into the Eloquent Archives database for all lower levels and digital content. While in the Eloquent database they can navigate to all related material as well as conduct additional searches.
Relevant aspects	
Remarks	Mobile application oriented
Strengths	The tool is a suite of different components integrated for provide support to records, archives, library and museums.
Weaknesses	-

Table 94: Eloquent Archives

12.9 AXIELL CALM

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Axiell CALM
URL	http://alm.axiell.com/collections-management-solutions/technology/calm-archive/
Basic description	<p>Archive & Records Collections Management System provides:</p> <ul style="list-style-type: none"> • Catalogue to archival standards. • Store, link and manage data using a sophisticated hierarchy based

Evaluation criteria	Description of the evaluation criteria
	<p>data structure.</p> <ul style="list-style-type: none"> • Ensure consistency of authority terms by linking to external thesauri and people databases through Linked Open Data. • Store and link a wide variety of multimedia and text files. • Create efficient workflows to manage conservation, Reading Room requests, digitisation processes, etc. • Manage user permissions and access to records and authority files to adhere to data protection rules. • Manage exact physical locations using hierarchies of locations. <p>Manage treatments schedule and condition checks.</p>
Owner	<p>Axiell Group.</p> <p>The company comprises four business areas – Axiell Public Library, Axiell ALM (Archives, Libraries & Museums), Axiell Education and Axiell Media – which together deliver innovative solutions to the cultural sector globally.</p> <p>Leader for Archival collection management in the UK. Today are in more than 3400 institutions across 55 countries</p>
Category	<p>Archives management system</p> <p>Library and collection management system</p>
License and delivery model	<p>Commercial</p> <p>Cloud & On-premises</p>
Functional & Technical description	
Business processes	Acquisition, Archives Processing, Preservation, Access, Consultation
Standards	<p>DACS, ISAD(G), and RAD</p> <p>Integration with authority tools and standard thesaurus.</p> <p>Linked Data</p>
Access management	Easily manage user permissions and access to records and authority files to adhere to data protection rules.
Reporting and statistics	Reports of actions performed
Migration	Provided by owner
Integration	<ul style="list-style-type: none"> • Integrate with Digital Preservation Systems and DAMs (Digital Asset Management) • Open API: Allows integrate with third party solutions
Support	Offers technical support, consulting and training. Customer focused services.
Relevant aspects	
Remarks	High integration with third party solutions offered by an open API.
Strengths	Ensure consistency of authority terms by linking to external thesauri and

Evaluation criteria	Description of the evaluation criteria
	people databases through Linked Open Data
Weaknesses	-

Table 95: Axiell CALM

12.10 SCOPEARCHIV

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	scopeArchiv™ scopeOAIStm
URL	http://www.scope.ch/en/
Basic description	The suite standard software package scopeArchiv™ covers the whole work process from accession to usage.
Owner	Scope Solutions AG. The company Support for Archives, libraries and museums to respond path-breaking to the challenges in the areas of knowledge management, open data, data management and digital preservation with scopeArchiv™ and scopeOAIStm.
Category	Archives management system Preservation system
License and delivery model	Commercial Cloud & On-premises
Functional & Technical description	
Business processes	Acquisition, Archives Processing, Preservation, Access, Consultation, Administration
Standards	Supports the archival standards ISAD(G), ISAAR(CPF), EAD, METS and has been implemented with the OAIStm reference model in mind.
Access management	One access control facility for centralised administration of access rights.
Reporting and statistics	Module Reports based on Cristal reports.
Functional architecture	The platform is composed by several modules, providing specific features:

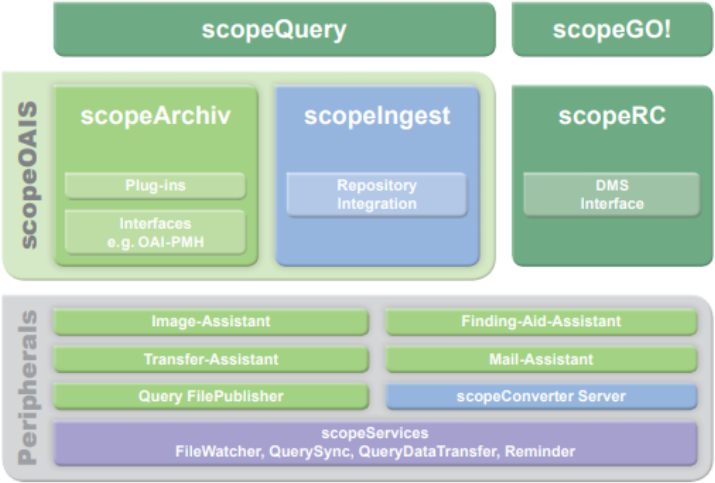
Evaluation criteria	Description of the evaluation criteria
	 <p style="text-align: center;">Figure 67: Scope Archive product range</p>
Infrastructure platform	<p>Composed by:</p> <ul style="list-style-type: none"> • Client- Server solution. • Microsoft Windows Server: Database, Web Server, ScopeArchiv Server, OAIS Server • Oracle Server Database
Migration	<p>Provided by owner. Services oriented to the user that incorporates training, installation, migration, data transfers and individual programming.</p>
Integration	<p>Integration with content management systems: The tool offers proprietary IT solutions for archive management and authentic transfer of data from document management systems.</p>
Support	<p>Services oriented support by owner.</p>
Relevant aspects	
Remarks	<p>Add-on: Inventory & Location Management Transfer of data from document management systems Covers: records management, archive management and long term preservation (OAIS)</p>
Strengths	<p>Transfer assistant module and quality control and data preparation before the data is transferred to scopeArchiv (Units of Description, Descriptors, Filing Plans)</p>
Weaknesses	<p>-</p>

Table 96: scopeArchiv

12.11 ARCHIDOC

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	ArchiDoc
URL	http://www.archidoc.pl/en/
Basic description	The tool provides operation of the warehouse and the storage and archiving of documentation in compliance with the top safety standards. Digitalisation services of original paper documents are also supported
Owner	For over 20 years the company offers business services in the field of document management and back-office support.
Category	Archives management system
License and delivery model	Commercial Cloud
Functional & Technical description	
Business processes	Acquisition, Archives Processing, Access, Consultation
Standards	ISAD(G), ISAAR(CPF), EAD
Access management	Security of users and access
Reporting and statistics	Easy access to information and reporting: <ul style="list-style-type: none"> • Quantitative reports of the documentation stored • Reports on borrowing and scanning • Reports on rejecting and destroying • Other (upon request)
Functional architecture	The services offered are: <ul style="list-style-type: none"> • Documentation storage; • Access to the originals or scanned copies of documents; • Ordering documents in A and B category; • Document rejecting and destroying;

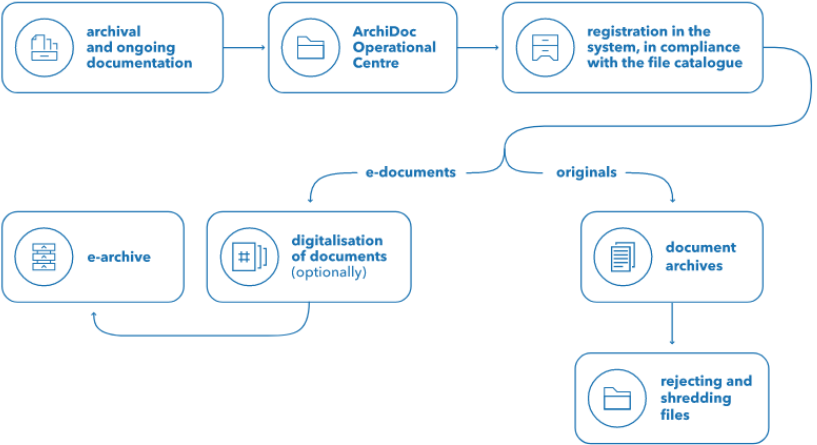
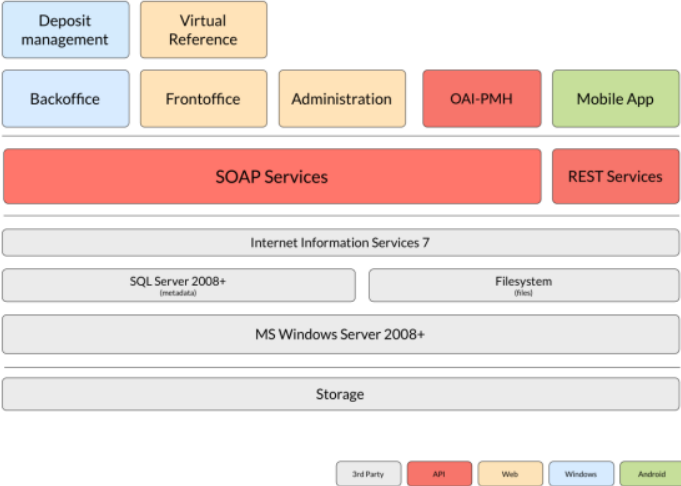
Evaluation criteria	Description of the evaluation criteria
	 <pre> graph TD A[archival and ongoing documentation] --> B[ArchiDoc Operational Centre] B --> C[registration in the system, in compliance with the file catalogue] C --> D[e-documents] C --> E[originals] D --> F[digitalisation of documents (optionally)] F --> G[e-archive] E --> H[document archives] H --> I[rejecting and shredding files] </pre>
Migration	Migration support provided by owner
Integration	Integration with digitisation module
Support	Offers technical support, consulting and training. Customer focused services.
Relevant aspects	
Remarks	Several services and products in the scope of document management instead in the scope of archives. No information available in the web site about main classical business processes of selection, appraisal, transfer, etc.
Strengths	Integration with digitisation module Configurable according to the needs of the client
Weaknesses	The product is not the typical archives management system and is more oriented to the management of the physical file and its digital copy according to the information on the web site However, the product could be adapted to implement the typical workflows of historical archive such as the management of the selection, transfers and other classic processes.

Table 97: ArchiDoc

12.12 ARCHEEVO

Evaluation criteria	Description of the evaluation criteria
Basic description	

Evaluation criteria	Description of the evaluation criteria
Name	Archeevo
URL	https://www.keep.pt/produtos/archeevo/?lang=en
Basic description	Archeevo is a management software capable of handling millions of archival records and terabytes of digital assets. This software consists of 9 functional modules that meet the needs of the most experienced archival professional, i.e. records description, management of digital assets, online publication, conservation and restoration, intermediate archive, deposit management, virtual reference room, administration and productivity management, and interoperable programmable interfaces.
Owner	KEEP SOLUTIONS is a European company that provides advanced services for managing and preserving digital information. The company initiated its activity in 2008 and attained the status of spin-off of the University of Minho for being an enterprise that maintains close ties to research centres and departments of this university.
Category	Archives management system Library and collection management system
License and delivery model	Commercial Cloud; on-Premises
Functional & Technical description	
Business processes	Acquisition, Archives Processing, Preservation, Access, Consultation
Standards	ISAD(G), ISAAR(CPF), EAD, METS, Baglt File Packaging (Library of Congress) Compatible with national authority records registry services.
Access management	Support for management of users, groups, roles and access permissions to digital objects
Reporting and statistics	Log accountability: employee performance reporting and statistics
Functional architecture	Archeevo is composed by several modules providing functional requirements:

Evaluation criteria	Description of the evaluation criteria
	 <p style="text-align: center;">Figure 68: Functional architecture of Archevo</p>
Infrastructure platform	<p>Server-side modules are:</p> <ul style="list-style-type: none"> • Database server – where the database engine will run; • Application server – where the Internet Information Services (IIS) will run to support the following modules: Core Services, Virtual Reference, Front-office, Administration, and OAI-PMH.
Migration	Migration support provided by owner
Integration	<p>Suite of tools integrated:</p> <ul style="list-style-type: none"> • LDAP authentication • Record management module integration • Preservation system (OAIS) integration <p>Integration with external OAI-PMH service providers, such as European Archives Portal, Europeana, Driver, and others</p>
Support	Offers technical support, consulting and training. Customer focused services.
Relevant aspects	
Remarks	Enhanced Digital Asset Management capabilities (DAM)
Strengths	<p>Full integration between different modules giving support to the full cycle of documents.</p> <p>Archevo enables the archival professional to do description, management of digital assets, deposit management, online publishing, productivity management, conservation and restoration, reporting, intermediate archival including transfers and records eliminations, deposit management, etc.</p>

Evaluation criteria	Description of the evaluation criteria
Weaknesses	-

Table 98: Archeevo

12.13 PRESERVICA – ARCHIVSPACE CONNECTOR

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Preservica ArchivesSpace Connector
URL	https://preservica.com/resources/knowledge-centre/archivesspace-connector-overview http://archivesspace.org/developers/integrations-with-archivesspace/
Basic description	The Preservica ArchivesSpace connector provides an easy and automated way to synchronise metadata and catalogue hierarchy between the two systems, thereby simplifying the ingest of digital content and automatically maintaining changes after ingest. This makes it easy to maintain a single ArchivesSpace catalogue view across both physical and digital artefacts, avoids the need to re-key metadata and ensures that valuable digital content is safely preserved and accessible into the future.
Owner	Preservica
Category	Connectivity Tool
License and delivery model	Commercial Cloud & on-premises
Functional & Technical description	
Business processes	N/A
Standards	XIP
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	Provides integration between AMS and Preservica
Support	By owner
Relevant aspects	
Remarks	Integration service between archives management system and preservation system at level of:

Evaluation criteria	Description of the evaluation criteria
	<ul style="list-style-type: none"> • Single Catalogue View • Automatically Synchronise after Ingest • Metadata and Hierarchy Mapping
Strengths	Integration itself
Weaknesses	-

Table 99: Preservica – ArchivesSpace Connector

12.14 PRESERVICA – AXIELL CALM CONNECTOR

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Preservica Axiell Calm Connector
URL	https://preservica.com/resources/press-releases/preservica-simplifies-digital-preservation-for-calm-catalogue-users
Basic description	This connector allows automatically synchronise Calm catalogue information with digital content and metadata held in Preservica
Owner	Preservica
Category	Connectivity Tool
License and delivery model	Commercial Cloud & on-premises
Functional & Technical description	
Business processes	N/A
Standards	XIP
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	Provides integration between AMS and Preservica
Support	By owner
Relevant aspects	
Remarks	Preservica integration provides the process of safeguarding digital content and automatically synchronising catalogue data between the two systems. Calm is used to catalogue collection of both physical and digital assets.
Strengths	Integration itself

Evaluation criteria	Description of the evaluation criteria
Weaknesses	-

Table 100: Preservica – Axiell CALM Connector

12.15 ARCHIVESSPACE – ATOM – ARCHIVEMATICA CONNECTOR

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	ArchivesSpace – AtoM – Archivemata connector
URL	https://libraries.mit.edu/digital-archives/integrating-tools/
Basic description	<p>The combination of these three products provides support to the full lifecycle and business processes:</p> <ul style="list-style-type: none"> • ArchivesSpace – archives collection management and discovery • Archivemata – digital preservation system (OAIS) • AtoM – archives and special collections discovery and delivery system and collections management system. <p><u>ArchivesSpace – AtoM scope:</u> Describe collections in ArchivesSpace + Accession, etc.</p> <ul style="list-style-type: none"> • Export that descriptive info at EAD • Import it into AtoM as a Collection Description <p><u>Archivemata – AtoM scope:</u></p> <ul style="list-style-type: none"> • Run digital collection files through Archivemata • Link the DIP files to the AtoM descriptions <p><u>Archivemata – ArchivesSpace scope:</u> Send the location and file info to ArchivesSpace</p>
Owner	<p>Artefactual Systems (lead development company) develops free and open software made available under the AGPLv3 open-source software license. Provides support through release management, public technical and user documentation, and community forum support.</p> <p>Provides other services: migration, consulting, hosting, installation, and training.</p>
Category	Connectivity Tool
License and delivery model	<p>Open-source</p> <p>Cloud & on-premises</p>
Functional & Technical description	
Business processes	N/A
Standards	The integrated suite of free and open-source tools that allows users to process digital objects from ingest to archival storage and access in

Study on Standard-Based Archival Data Management, Exchange and Publication
Final Report

Evaluation criteria	Description of the evaluation criteria
	compliance with the ISO-OAIS functional model and other digital preservation standards and best practices (PREMIS, METS, DublinCore)
Access management	N/A
Reporting and statistics	N/A
Functional architecture	<p>Proposed workflow showing human, tool, and storage interactions.</p>
Migration	N/A
Integration	Integration between archives management system and preservation system
Support	By owner
Relevant aspects	
Remarks	Full life cycle (from ingest to archival storage and access)
Strengths	Integration itself
Weaknesses	-

Table 101: ArchivesSpace – Atom – Archivemata connector

12.16 E-ARK EXTRACTION TOOLS

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	A set of tools: <ul style="list-style-type: none"> • ESSArch Tools for Producer (ETP) • RODA-in • Universal Archiving Module (UAM)

Evaluation criteria	Description of the evaluation criteria
	<ul style="list-style-type: none"> • CEF Building Block e-Archiving (CEF programme 2018)
URL	http://www.eark-project.com/resources/project-deliverables/93-d34-1
Basic description	Several tools have been created in the framework of E-ARK project that cover a full extend of parts of OAIS Standard. E-ARK Extraction Tools allow the content to be exported using the pull method and imported into one of the SIP Creation Tools.
Owner	Project E-ARK. Multinational big data research project with the objective to improve the methods and technologies of digital archiving, in order to achieve consistency on a Europe-wide scale.
Category	Connectivity Tool
License and delivery model	Open-source
Functional & Technical description	
Business processes	Pre-ingest
Standards	METS
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	Integration tool
Support	Open-source
Relevant aspects	
Remarks	Provides normalisation in pre-ingestion processes eArchiving will be a Core Service Platform, hosted and maintained by the European Commission. It will include technical specifications, software and supporting services. It is also a so-called Building Block, i.e. providing basic and re-usable digital services that can be integrated into existing infrastructures. https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/apply-funding/2018-eArchiving
Strengths	European National Archives scope
Weaknesses	-

Table 102: E-ARK extraction tools

12.17 E-ARK ACCESS TOOLS

Evaluation criteria	Description of the evaluation criteria
---------------------	--

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	A set of tools: The Access Software Platform (Search Module, Order Management Tool, and IP Viewer)
URL	http://www.eark-project.com/resources/project-deliverables/92-d54/file
Basic description	E-ARK Access tools for provide search and access capabilities using standard technologies: CMIS, OLAP and GIS. <i>CMIS: There is a set of services for adding and retrieving documents ('objects').</i> <i>OLAP is part of the broader category of business intelligence, which also encompasses relational database, report writing and data mining.</i> <i>A geographic information system (GIS) is a system designed to capture, store, manipulate, analyse, manage, and present spatial or geographic data.</i>
Owner	E-ARK Project: Multinational big data research project with the objective to improve the methods and technologies of digital archiving, in order to achieve consistency on a Europe-wide scale.
Category	Connectivity Tool
License and delivery model	Open-source
Functional & Technical description	
Business processes	Access
Standards	METS
Access management	N/A
Reporting and statistics	N/A
Functional architecture	<p>The access tools are categorised into three access modules: Search module; Order management Tool and IP viewer.</p> <p>Every module contains specific tools, covering features such as CMIS, OLAP, and GIS. Most of them contribute to add new access systems to the data stored in preservation systems.</p> <pre> graph LR ASP[Access Software Platform] SM[1. Search Module] --> OMT[2. Order Management Tool] OMT --> IPV[3. IP Viewer] OMT -- includes --> AIPDIP[AIP-DIP Conversion Tool] IPV -- has --> DPTK[Database Preservation Toolkit] IPV -- has --> DVTK[Database Visualization Toolkit] IPV -- has --> RDBMS[RDBMS] IPV -- has --> Geodata[Geodata tool] IPV -- includes --> OLAP[OLAP tool] IPV -- includes --> SMURF[SMURF (EDRMS and SFSB) Viewer] </pre>
Figure 69: Access tools E-ARK	

Evaluation criteria	Description of the evaluation criteria
Migration	N/A
Integration	Integration tool
Support	Open-source
Relevant aspects	
Remarks	Provides normalisation in access processes
Strengths	European National Archives scope
Weaknesses	-

Table 103: E-ARK access tools

12.18 ARCHIVEMATICA (DIP PROVIDER)

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Archivemata
URL	https://www.archivemata.org/en/docs/archivemata-1.6/user-manual/access/access/
Basic description	<p>During ingest, access copies of digital objects are generated and packaged into a DIP (Dissemination Information Package). The user uploads the DIP to the access system, or stores it for future use.</p> <p>Access Systems:</p> <ul style="list-style-type: none"> • Upload DIP to AtoM • Upload metadata to AtoM (AIP) • Upload DIP to ArchivesSpace • Upload DIP to Archivists' Toolkit
Owner	<p>Artefactual Systems (lead development company) develops free and open software made available under the AGPLv3 open-source software license. Provides support through release management, public technical and user documentation, and community forum support.</p> <p>Provides other services: migration, consulting, hosting, installation, and training.</p>
Category	Connectivity Tool
License and delivery model	Open-source Cloud; on-Premises
Functional & Technical description	
Business Process	Access

Evaluation criteria	Description of the evaluation criteria
Standards	OAIS METS
Access management	Capabilities to users, groups, roles and access permissions
Reporting and statistics	Reporting and statistics available
Migration	N/A
Integration	Connection tool
Support	Provided by the community of users and developers. On-line support and documentation official web-sites. Specific support provided by owner.
Relevant aspects	
Remarks	In Archivematica 1.6 and higher, you can send AIP object metadata to Atom without uploading dissemination copies of the files.
Strengths	Enhanced dissemination capabilities for DIP packages into OAIS preservation system
Weaknesses	N/A

Table 104: Archivematica (DIP provider)

12.19 PRESERVICA OAI-PMH API

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Preservica OAI-PMH API
URL	https://preservica.com/
Basic description	OAI-PMH API for metadata harvesting.
Owner	Preservica. Preservica offers consulting and research services. It serves archives, libraries, government organisations, museums, and businesses in the United States and internationally. The company is based in Boston, Massachusetts. Preservica operates as a subsidiary of Tessella Ltd.
Category	Connectivity Tool
License and delivery model	Commercial Cloud / on-Premises
Functional & Technical description	
Business processes	Access, Data Exchange

Evaluation criteria	Description of the evaluation criteria
Standards	EAD, DC, METS.
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	Integration tool
Support	By the owner
Relevant aspects	
Remarks	N/A
Strengths	N/A
Weaknesses	-

Table 105: Preservica OAI-PMH API

12.20 ATOM OAI-PMH PLUGIN

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	AtoM OAI-PMH plugin
URL	https://www.accesstomemory.org
Basic description	AtoM has the ability to act as a data provider to service providers interested in harvesting descriptive metadata from AtoM, by exposing simple Dublin Core XML via OAI-PMH, the Open Archives Initiative Protocol for Metadata Harvesting (version 2.0).
Owner	Open-source supported by Artefactual Systems (lead development company) develops free and open software made available under the AGPLv3 open-source software license. Provides support through release management, public technical and user documentation, and community forum support.
Category	Connectivity Tool
License and delivery model	Open- source On premises
Functional & Technical description	
Business processes	Access, Data Exchange
Standards	DC, EAD

Evaluation criteria	Description of the evaluation criteria
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	Connection tool
Support	By the owner
Relevant aspects	
Remarks	N/A
Strengths	N/A
Weaknesses	-

Table 106: Atom OAI-PMH plugin

12.21 CONTENTDM

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	CONTENTdm
URL	http://www.oclc.org/en/contentdm.html
Basic description	CONTENTdm is a digital collection management system for acquire, store and publish content via WorldCat, the premier database of library materials. CONTENTdm also secures and monitors master files in a cloud-based preservation archive
Owner	OCLC, is an American non-profit cooperative organisation "dedicated to the public purposes of furthering access to the world's information and reducing information costs". It was founded in 1967 as the Ohio College Library Centre. OCLC and its member libraries cooperatively produce and maintain WorldCat, the largest online public access catalogue (OPAC) in the world.
Category	Library and collection management system
License and delivery model	Commercial On premises & cloud
Functional & Technical description	
Business processes	Archives Processing, Preservation, Administration, Access, Consultation
Standards	DC, MARC21
Access management	Users, roles and permissions

Evaluation criteria	Description of the evaluation criteria
Reporting and statistics	Web Analytics and internal usage reports
Migration	Services for migration
Integration	With library catalogue WorldCat
Support	On-line manuals, training and specific support by the owner
Relevant aspects	
Remarks	CONTENTdm includes a preservation archive to save library's master files and digital originals are securely stored in a purpose-built environment.
Strengths	The product provides upload capabilities for collection metadata to WorldCat — the premier database of library materials.
Weaknesses	No integration with archives management systems.

Table 107: CONTENTdm

12.22 OMEKA

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Omeka Classic Omeka net Omeka S Omeka Curator Dashboard
URL	http://www.omeka.net/
Basic description	Omeka Classic is a web publishing platform for sharing digital collections and creating media-rich online exhibits. Cloud version is Omeka net Omeka S is a next-generation web publishing platform for institutions interested in connecting digital cultural heritage collections with other resources online. Omeka Curator Dashboard is a collection of 15 plugins designed with the aim of facilitating the import and export of data, the management of metadata and the cure of contents in Omeka
Owner	Omeka.net is a project of the Corporation for Digital Scholarship.
Category	Library and collection management system
License and delivery model	Omeka is an open-source development Cloud, on-premises (not free)
Functional & Technical description	

Evaluation criteria	Description of the evaluation criteria
Business processes	Archives Processing, Preservation, Administration, Access, Consultation
Standards	Dublin Core, OAI-PMH RDF/XML
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	Integrates digital cultural heritage collections with other resources online.
Support	On-line documentation
Relevant aspects	
Remarks	Connect to the semantic-web Publish items with Linked Open Data.
Strengths	A suite of tools with enhancements for Linked Data and Linked Open Data from ingestion to publish. Omeka S is a next-generation web publishing platform for institutions interested in connecting digital cultural heritage collections with other resources online. Capabilities to add new vocabulary files (RDF/XML with an rdf extension).
Weaknesses	No connection with archives management system

Table 108: Omeka - top-down approach

12.23 NEATLINE

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Neatline
URL	http://neatline.org/
Basic description	Neatline allows scholars, students, and curators to tell stories with maps and timelines. As a suite of add-on tools for Omeka, it opens new possibilities for hand-crafted, interactive spatial and temporal interpretation.
Owner	Neatline is a project of the Scholars' Lab at the University of Virginia Library. It has benefited from the generous support of the UVa Library, the National Endowment for the Humanities, the Institute of Museum and Library Services, and the Library of Congress.
Category	Library and collection management system

Evaluation criteria	Description of the evaluation criteria
License and delivery model	Open-source On-premises
Functional & Technical description	
Business processes	Access
Standards	DC
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	With Omeka
Support	By the owner, community of users
Relevant aspects	
Remarks	Neatline adds an interactive map-making environment that makes it possible to create Neatline exhibits, each of which is populated with its own collection of records, which can optionally be synchronised with items in the underlying Omeka collection.
Strengths	Open-source map integration system, Neatline can be used to build exhibits on top of a collection of modern-geography base layers - the OpenStreetMap tile set, the Google Maps API, for example.
Weaknesses	-

Table 109: Neatline

12.24 HPE CONTROLPOINT

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	HPE ControlPoint
URL	http://www8.hp.com/us/en/software-solutions/file-analysis-dark-data-cleanup/index.html
Basic description	File analysis software to access, understand, classify and control all enterprise information. Supports entity extraction so as to find personally identifiable information (PII), personal credit information (PCI) and personal health information (PHI) in documents and emails etc. Accelerate compliance and information governance with legacy data clean-up to reduce redundant, obsolete, trivial and dark data. ControlPoint helps to achieve

Evaluation criteria	Description of the evaluation criteria
	information compliance by making it possible to not only access information, but also understand, classify, and reduce outdated and unnecessary legacy dark data content
Owner	HPE
Category	Content analysis
License and delivery model	Commercial On-Premises & Cloud
Functional & Technical description	
Business processes	Archives Processing, Administration
Standards	N/A
Access management	N/A
Reporting and statistics	About use
Migration	N/A
Integration	With other HPE products, as content manager.
Support	By the owner
Relevant aspects	
Remarks	<ul style="list-style-type: none"> • Included in a suite of tools: from file analysis to records management. Email & User desktop applications (easy integration) • Machine learning and probability to auto classify content <ul style="list-style-type: none"> ○ Provides a system to navigate over classification and analyse the content. ○ Tag the content and update classifications • Paper format support Encryption features
Strengths	Huge dictionary and country options for discovery of personal data (PII, PCI, PHI) <ul style="list-style-type: none"> • Predefined patterns and regular expressions to identify PII, PCI, and PHI and tag them. • Customisable to configure new patterns. Included patron over different countries and jurisdictions.
Weaknesses	Unable to detect personal data in other formats such as video, sound, images.

Table 110: HPE ControlPoint

12.25 IBM WATSON

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	IBM-Watson
URL	https://www.ibm.com/support/knowledgecenter/SS8NLW_10.0.0/com.ibm.discovery.es.nav.doc/iitypofnv_prodoover_cont.htm
Basic description	<p>IBM Watson is a technology platform that uses natural language processing and machine learning to reveal insights from large amounts of data.</p> <p>IBM Watson Explorer Content Analytics collects and analyses structured and unstructured content in documents, email, databases, websites, and other enterprise repositories.</p> <p>By providing a platform for crawling and importing content, parsing and analysing content, and creating a searchable index, Watson Explorer Content Analytics helps you perform text analytics across all data in your enterprise and make that data available for analysis and search.</p>
Owner	IBM
Category	Content analysis
License and delivery model	Commercial On-Premises & Cloud
Functional & Technical description	
Business processes	Archives Processing, Administration
Standards	N/A
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	With other IBM products. Belongs to a suite
Support	By the owner
Relevant aspects	
Remarks	<p>IBM Watson combines:</p> <ul style="list-style-type: none"> • Deep cognitive computing capabilities. Watson uses natural language processing machine learning, and evidence-based hypotheses and scoring. • Enabling hardware. Watson's processing power and storage capabilities deliver cognitive functionality in real time. • Support community. IBM is building the Watson Partners Program around its Watson offering including app developers, content

Evaluation criteria	Description of the evaluation criteria
	<p>providers, and service providers.</p> <ul style="list-style-type: none"> • Watson Curator is a component of the tool. Watson Curator helps users assess and gather relevant text documents across multiple sources using a guided and intuitive review process. It helps subject matter experts create higher-quality information collections much quicker and automatically manages superseded and contested information
Strengths	Provided as an architecture of components to address new challenges as visual recognition, machine translation, relationship extraction of information, etc. and offers a platform to develop new applications based on this product.
Weaknesses	Watson requires many months of laborious training and preparation of content in order to start to use.

Table 111: IBM Watson

12.26 MORIARTY

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Moriarty
URL	http://everismoriarty.com/
Basic description	Everis Moriarty is the world's most advanced workbench for developing Big Data applications. It easily enables the integration of sophisticated analytical services (i.e. reasoning engines, machine learning, natural language processing, deep learning algorithms, semantic ontologies, etc.) to develop cognitive solutions.
Owner	everis
Category	Content analysis
License and delivery model	Commercial On-Premises & Cloud
Functional & Technical description	
Business processes	Archives Processing, Administration
Standards	N/A
Access management	N/A
Reporting and statistics	N/A

Evaluation criteria	Description of the evaluation criteria
Migration	N/A
Integration	-
Support	By the owner
Relevant aspects	
Remarks	Next release of Moriarty will offer an enterprise platform for the development and deployment of Artificial Intelligence applications.
Strengths	<ul style="list-style-type: none"> Smart Semantic Search Enrich traditional search with the semantic and syntactic understanding of documents, texts, audio and video. Document-Based Discovery Use inductive techniques to extract knowledge and entity relations from text.
Weaknesses	-

Table 112: Moriarty

12.27 SAS

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	SAS Contextual Analytics SAS Text Miner
URL	https://www.sas.com/en_us/software/visual-text-analytics.html https://www.sas.com/en_us/solutions/analytics.html#text-analytics
Basic description	SAS Visual Text Analytics uncover insights hidden within unstructured data using the combined power of natural language processing, machine learning and linguistic rules.
Owner	SAS Institute (or SAS, pronounced "sass") is an American multinational developer of analytics software based in Cary, North Carolina. SAS develops and markets a suite of analytics software (also called SAS), which helps access, manage, analyse and report on data to aid in decision-making. The company is the world's largest privately held software business and its software is used by most of the Fortune 500.
Category	Content analysis
License and delivery model	Commercial On-Premises & Cloud

Evaluation criteria	Description of the evaluation criteria
Functional & Technical description	
Business processes	Archives Processing, Administration
Standards	N/A
Access management	N/A
Reporting and statistics	N/A
Migration	N/A
Integration	-
Support	By the owner
Relevant aspects	
Remarks	SAS is a Leader in The Forrester Wave™: Predictive Analytics and Machine Learning Solutions, Q1 2017
Strengths	<p>Provides:</p> <ul style="list-style-type: none"> • Content categorisation. A linguistic-based document summary, including search and indexing, content alerts and duplication detection. • Topic discovery and modelling. Accurately capture the meaning and themes in text collections, and apply advanced analytics to text, like optimisation and forecasting. • Contextual extraction. Automatically pull structured information from text-based sources. • Sentiment analysis. Identifying the mood or subjective opinions within large amounts of text, including average sentiment and opinion mining. • Speech-to-text and text-to-speech conversion. Transforming voice commands into written text, and vice versa. • Document summarisation. Automatically generating synopses of large bodies of text. <p>Machine translation. Automatic translation of text or speech from one language to another.</p>
Weaknesses	-

Table 113: SAS

12.28 BLANCOO

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	Blancoo File Eraser – File Erasure Software

Evaluation criteria	Description of the evaluation criteria
	Blancoo Data Eraser for Government Agencies
URL	https://www.blancco.com/products/file-eraser/
Basic description	<p>File & Data eraser IT –Tool to eliminate data and content</p> <ul style="list-style-type: none"> • Blancoo Drive Eraser Securely erase sensitive data from HDDs and complex SSDs in desktop/laptop computers and servers. • Blancoo File Eraser Go beyond file shredding with secure erasure of sensitive files/folders with integration and automation of rules. • Blancoo Removable Media Eraser Permanently erase data from USB drives, SD cards, micro drives, CompactFlash cards and other flash memory storage devices. • Blancoo Management Console Centralised data erasure management reporting across every IT asset with the ability to integrate with your asset management systems. • Blancoo Enterprise Appliances Broad on-site data erasure solutions for erasing large volumes of loose drives or damaged drives
Owner	Blancoo Technology Group is a leading global provider of mobile device diagnostics and secure data erasure solutions.
Category	Secure deletion (Elimination)
License and delivery model	Commercial On-premises Cloud
Functional & Technical description	
Business processes	Archives Processing, Administration
Standards	US NIST 800-88.1 and older DoD 5220.22-M
Access management	Supports security
Reporting and statistics	Digitally signed, tamper-proof reporting Flexible and User-Friendly Interface Comprehensive Reporting
Migration	N/A
Integration	With Content Management systems
Support	By the owner
Relevant aspects	
Remarks	Blancoo Data Eraser solutions meet the data sanitization recommendations

Evaluation criteria	Description of the evaluation criteria
	set forth by US NIST 800-88.1 and older DoD 5220.22-M.
Strengths	The suite has specific solutions for government agencies
Weaknesses	-

Table 114: Blancoo

12.29 OPENTEXT REDACT-IT

Evaluation criteria	Description of the evaluation criteria
Basic description	
Name	OpenText™ Redact-It™
URL	https://www.opentext.com/what-we-do/products/enterprise-content-management/content-centric-applications/opentext-redact-it
Basic description	OpenText™ Redact-It™ effortlessly removes sensitive content and personally identifiable information from documents as a part of business processes. Redact-It can be used for desktop use, bulk redaction or integration as part of a workflow. Redact-It works with a range of document types—including PDFs, Microsoft® Office documents and scanned images. It creates a new PDF or TIFF rendition of the redacted file, leaving the source file untouched. Instead of just covering sensitive information, Redact-It removes all content and associated metadata.
Owner	OpenText
Category	Redaction tool
License and delivery model	Commercial On-premises & Cloud
Functional & Technical description	
Business processes	Archives Processing
Standards	General regulatory standards
Access management	N/A
Reporting and statistics	Reports of use
Migration	N/A
Integration	Integrated with OpenText products
Support	By the owner
Relevant aspects	

Evaluation criteria	Description of the evaluation criteria
Remarks	
Strengths	<p>Redact individual work with documents on-demand in an automated workflow process or redact entire folder hierarchies as a batch process. Can use complex expressions to automatically redact all instances of a phrase, name, social security number, phone number, account number, monetary amount and more. Or, redact predefined block out zones on common image document types. Covers:</p> <ul style="list-style-type: none"> • Automatically find and remove sensitive data in unstructured documents and forms • Replace tedious, manual redaction with highly efficient automated processes • Protect intellectual property or personal information from unintentional disclosure, theft or other misuse • Meet regulatory requirements for privacy and disclosure • Make data protection a seamless part of everyday workflows, not a separate process <p>Integrate redaction capability with content and records management</p>
Weaknesses	-

Table 115: OpenText Redact-it

12.30 LIST OF ARCHIVES MANAGEMENT SYSTEMS

Product/Tool	Information	Vendor	Brief Description
Archivist's Toolkit	http://www.archiviststoolkit.org/	Open-source applications	<p>The Archivists' Toolkit™, or the AT, is the first open source archives management system to provide broad, integrated support for the management of archives. It is intended for a wide range of archival repositories. AT is superseded by ArchivesSpace (http://www.archivesspace.org/) and AT support ended September 1, 2013. This website (www.archiviststoolkit.org) will be available for the next few months as the transition is made to ArchivesSpace</p>

Archives Space	http://archivesspace.org https://archivesspace.atlassian.net/wiki/s/ADC/overview	Open-source applications	<p>ArchivesSpace™ is an open source, web application for managing archives information. The application is designed to support core functions in archives administration such as accessioning; description and arrangement of processed materials including analogue, hybrid, and born-digital content; management of authorities (agents and subjects) and rights; and reference service. The application supports collection management through collection management records, tracking of events, and a growing number of administrative reports. The application also functions as a metadata authoring tool, enabling the generation of EAD, MARCXML, MODS, Dublin Core, and METS formatted data.</p>
Archon	http://www.archon.org	Open-source applications	<p>Archon™ is a unified platform for archival description and access. It provides both a way to record descriptive information about collections and digital objects and a means to view, search, and browse that information in a fully-functional public web site. Integrated in ArchivesSpace from 2009.</p>
AtoM	https://www.accessto-memory.org	Open-source applications	<p>AtoM™ stands for Access to Memory is a web-based, open source application for standards-based archival description and access in a multilingual, multi-repository environment that was originally commissioned by the International Council on Archives to make it easy for archival institutions worldwide to put their archival holdings online using the ICA's descriptive standards. The tool works in combination with ArchivesSpace.</p>
Cuadra Star/Archives	https://lucidea.com/cuadrastar-skca/archives-management-software/	Lucidea	<p>CuadraSTAR Knowledge Centre for Archives SKCA™ is a powerful solution for making archival collections visible and accessible. SKCA complies with the DACS, EAD, MARC, ISAD (g), and Dublin Core archival standards. It also provides flexible support within the full hierarchical model collections, series, containers, and items. SKCA supports accession, catalogue, and management fonds/collections capabilities, EAD finding aids, MARC records, container lists, and management reports.</p>

Eloquent Archives	https://www.eloquent-systems.com/products/archives/	Eloquent Systems Inc.	<p>Eloquent Archives™ is a mobile-friendly archives collections management system (ACM) for everything from Accessioning to online Public Access.</p> <p>Plus all standard ACM modules: Accession/De-accession - Space Management - Tracking & Retrieval - Research Management - Member Module - HTML Publishing - EAD Export - Google Maps Mash-up - Google Analytics.</p> <p>Lucidea acquires Eloquent Systems.</p>
CALM/Adlib	http://alm.axiell.com/collections-management-solutions/technology/calm-archive/	Axiell ALM	<p>Archive & Records Collections Management Systems provides:</p> <ul style="list-style-type: none"> -Catalogue to archival standards. -Store, link and manage data using a sophisticated hierarchy based data structure. -Ensure consistency of authority terms by linking to external thesauri and people databases through Linked Open Data. -Store and link a wide variety of multimedia and text files. -Create efficient workflows to manage conservation, Reading Room requests, digitisation processes, etc. -Manage user permissions and access to records and authority files to adhere to data protection rules. -Manage exact physical locations using hierarchies of locations. -Manage treatments schedule and condition checks.
scopeArchiv	http://www.scope.ch/en/ http://www.scope.ch/en/products/module-overview.html	Scope Solutions AG	<p>The comprehensive standard software package scopeArchiv™ supports our customers doing the daily work and covers the whole work process from accession to usage. scopeArchiv™ supports the archival standards ISAD(G), ISAAR(CPF), EAD, METS and has been implemented with the OAIS reference model in mind.</p> <p>Scope solutions aims at becoming the competent business partner for public archives in Europe and offers proprietary IT solutions for archive management and authentic transfer of data from document management systems.</p>

Archeevo	https://www.keep.pt/pr_odutos/archeevo/ https://www.keep.pt/pr_odutos/archeevo/?lang=en	KEEP SOLUTIONS	Archeevo is a state-of-the-art archives management software capable of handling millions of archival records and terabytes of digital assets. This software consists of 9 functional modules that meet the needs of the most experienced archival professional, i.e. records description, management of digital assets, online publication, conservation and restoration, intermediate archive, deposit management, virtual reference room, administration and productivity management, and interoperable programmable interfaces.
----------	--	----------------	--

13 ANNEX D: SELF-ASSESSMENT TOOL



D04

Self-Assessment Tool

14 ANNEX E: LINKED AND OPEN DATA INITIATIVES

14.1 INVENTORY OF OPEN DATA AND LINKED DATA INITIATIVES

14.1.1 Open Data Portal project (Publication Office)

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-001
Initiative name	European Union Open Data Portal
URL	http://data.europa.eu/euodp/en/home/
Basic description	This portal provides access to Open Data published by the European institutions and bodies.
Owner	EU Publications Office
Contact	http://data.europa.eu/euodp/en/contact
Type (sub-type or initiative specificity)	Open Data Portal
IPR	The hosted data is free to use for non-commercial and commercial use.
Technical description	
Technologies (language, infrastructure, etc.)	The portal is built with CKAN (Comprehensive Knowledge Archive Network) which is a management system for Open Data.
Architecture	Not available
Features	In addition to granting a standardised access to EU Open Data, this portal also provides an inventory of applications and tools (developed by both third-parties and in-house) that use these data.
Data models, formats	The description of datasets is stored in the RDF format.
Archival Data models, formats	The data model is based on the FRBR model.
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	Data is open and free to use for non-commercial and commercial use.
Open Data formats	N/A
Linked Data	Data is stored using a triple-store in the RDF format.
Linked Data formats	RDF

Evaluation criteria	Description of the evaluation criteria
Data transformation methods and tools	N/A
Data access / consumption	<p>Data can be accessed using two methods:</p> <ul style="list-style-type: none"> • A SPARQL endpoint allows users to query the knowledge base using the SPARQL language • A REST API is also available which can be consumed using any HTTP client (e.g. curl) or a specialised tool dedicated to making requests on the Comprehensive Knowledge Archive Network (CKAN). <p>These tools are available in multiple languages, including Python, Java and PHP.</p>
Data reusability	A list of applications that reuse these data is provided by the portal.

Table 116: Open Data Portal Project (Publication Office)

14.1.2 Pan-European Open Data Portal (DG CONNECT)

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-002
Initiative name	European Data Portal
URL	https://www.europeandataportal.eu/
Basic description	This portal collects metadata of Public Sector information available on portals across European countries.
Owner	DG CONNECT
Contact	https://www.europeandataportal.eu/
Type (sub-type or initiative specificity)	Open Data Portal
IPR	<p>Specific licensing exists for different datasets depending on the source entity/country. A licensing assistant is provided in the portal: https://www.europeandataportal.eu/en/content/filter-licenses</p> <p>Some dataset licenses prohibit commercial use.</p>
Technical description	
Technologies (language, infrastructure, etc.)	Drupal is used as a GUI Framework as well as CKAN for the management of datasets. Virtuoso is used as an RDF Database.

Evaluation criteria	Description of the evaluation criteria
Architecture	
Features	The portal offers eLearning and resources that introduces Open Data. In particular, the portal highlights processes that allow to prepare, publish and access Open Data.
Data models, formats	The common data formats available on the European Data Portal are CSV, JSON, GeoJSON, KML, XML and RDF Turtle.
Archival Data models, formats	N/A
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	The metadata is open and free to use. The portal doesn't host the actual datasets, links to resources are provided.
Open Data formats	N/A
Linked Data	Data is stored using a triple-store in the RDF format.
Linked Data formats	RDF
Data transformation methods and tools	The portal hosts specification about the transformation and cleaning processes to be done by the dataset host.
Data access / consumption	A SPARQL endpoint is available to consume the Open Data. A user- friendly search engine (powered by Drupal and CKAN) is available on the portal's website.
Data reusability	Through SPARQL endpoint, the data or datasets can be uploaded and reused in specific context.

Table 117: Pan-European Open Data Portal (DG CONNECT)

14.1.3 LOGD (Linked Open Government Data)

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-003
Initiative name	Linked Open Government Data (LOGD)
URL	https://joinup.ec.europa.eu/page/logd https://joinup.ec.europa.eu/document/study-business-models-linked-open-government-data-bm4logd
Basic description	Linked Open Government Data (LOGD) is a set of principles for publishing, linking and accessing open government data as a service on the Web.
Owner	SEMIC – ISA Programme of the European Commission Semantic Interoperability Community (SEMIC) is a European Commission initiative funded by ISA ² Programme
Contact	SEMIC: contact@semic.eu
Type (sub-type or initiative specificity)	Interoperability Framework (Initiative of the European Commission, Action of the ISA ² programme ⁴⁸)
IPR	Open License
Technical description	
Technologies (language, infrastructure, etc.)	Linked Data enables the provision of “data as a service” and conceives the Web as an open ecosystem where data owners, data publishers, and data consumers can interconnect and integrate disparate datasets. SEMIC explores the potential of Linked Open Government Data (LOGD), from a business and a technical point of view, as an enabler to the flexible integration of data coming from different e-Government systems.
Architecture	No architecture
Features	<p>The Study on Business Models for Linked Open Government Data - BM4LOGD identifies the number of enablers to the provision and reuse LOGD for public administrations:</p> <ul style="list-style-type: none"> • The main driver for the use of LOGD is that it allows for flexible data integration; this helps to increase data quality by allowing cross-references to authoritative data to be included and may drive future development of new services. • The use of LOGD increases the efficiency of the internal operation of the data provider and allows them to fulfil their public task more effectively and efficiently. • LOGD is applied most successfully in reference data, such as in the

⁴⁸ ISA² programme: https://ec.europa.eu/isa2/home_en

Evaluation criteria	Description of the evaluation criteria
	<p>case of the Named Authority Lists of the Publications Office, EuroVoc thesaurus and ESCO.</p> <ul style="list-style-type: none"> • LOGD makes future upgrades of data models much easier, for example to include new data or connect data from different sources together. • URIs allow a ‘follow-your-nose’ navigation structure that provides better navigation through complex data. • LOGD is mostly provided free of charge and under open licences which enables further use and reuse of data. • Availability of guidelines and dissemination of best practices create common approaches and reduce risk in implementation by enabling organisations to learn from each other.
Data models, formats	<p>Core Vocabularies:</p> <ul style="list-style-type: none"> • The Core Business Vocabulary is a simplified, reusable and extensible data model that captures the fundamental characteristics of a legal entity, e.g. the legal name, the activity, address, legal identifier, company type, and its activities. On 8 January 2013, the RDF syntax of the Core Business Vocabulary has been formally published on the W3C standards track as a Public Working Draft. The RDF syntax binding of the Core Business Vocabulary has been named as Registered Organisation Vocabulary by W3C. • The Core Location Vocabulary is a simplified, reusable and extensible data model that captures the fundamental characteristics of a location, represented as an address, a geographic name, or geometry. The Location Vocabulary is aligned with the INSPIRE data specifications. • The Core Person Vocabulary is a simplified, reusable and extensible data model that captures the fundamental characteristics of a person, e.g. the name, the gender, the date of birth, etc. • The Core Public Service Vocabulary is a simplified, reusable and extensible data model that captures the fundamental characteristics of a service offered by public administration. Such characteristics include the title, description, inputs, outputs, providers, locations, etc. of the public service. An application profile of the Core Public Service Vocabulary (CPSV-AP) has been developed for describing public services and grouping them in business events. • The Core Evidence and Criterion Vocabulary is a simplified, reusable and extensible data model for describing the principles and the means that a private entity must fulfil in order to become eligible or to be qualified to perform public services or participate in public procurement. Public entities often need to define the criteria. A Criterion is a rule or principle that is used to judge, evaluate or test something. An Evidence is the means by which a Criterion may be proven. • The Core Public Organisation Vocabulary is a simplified, reusable and extensible data model for describing public organisations in the

Evaluation criteria	Description of the evaluation criteria
	European Union.
Archival Data models, formats	There are no specific models for digital archiving but the Core Vocabularies can be used as data model to describe some metadata about the context or directly the content (people, organisations, locations, etc.).
Other relevant technical aspects	<ul style="list-style-type: none"> N/A
Relevant aspects for Data Archiving Management	
Open Data	N/A
Open Data formats	N/A
Linked Data	A state aid vocabulary for DG Competition to publish transparency information as open, machine-readable data (December 2015) ⁴⁹ : this report provides an overview of an RDF dissemination vocabulary for State Aid Transparency data.
Linked Data formats	N/A
Data transformation methods and tools	Cookbook for translating Data Models to RDF Schema (March 2013) ⁵⁰ : this cookbook provides guidance for the person who has the task of translating the Domain Model into an RDF schema.
Data access / consumption	Study on Business Models for Linked Open Government Data - BM4LOGD (October 2013) ⁵¹ : this report presents the outcome of a study on Linked Open Government Data (LOGD), a set of principles for publishing, linking and accessing open government data as a service on the Web.
Data reusability	Report on high-value datasets from EU Institutions (May 2014) ⁵² : this report elaborates on a working definition of what "high-value dataset means". This can be used as a guideline for prioritising the publication of government datasets. The definition approaches the value of datasets both from the publisher's point of view, and the re-users' point of view.

Table 118: LOGD (Linked Open Government Data)

14.1.4 Estonian – Open Data portal

Evaluation criteria	Description of the evaluation criteria
---------------------	--

⁴⁹ A state aid vocabulary for DG Competition to publish transparency information as open, machine-readable data (December 2015):

<https://joinup.ec.europa.eu/news/new-vocabulary-publishing>

⁵⁰ Cookbook for translating Data Models to RDF Schemas (SEMIC): <https://joinup.ec.europa.eu/document/cookbook-translating-data-models-rdf-schemas>

⁵¹ Study on Business Models for Linked Open Government Data - BM4LOGD: <https://joinup.ec.europa.eu/document/study-business-models-linked-open-government-data-bm4logd>

⁵² Report on high-value datasets from EU Institutions (May 2014): <https://joinup.ec.europa.eu/document/report-high-value-datasets-eu-institutions>

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-004
Initiative name	Estonian - National Archives opening data
URL	http://opendata.ra.ee/
Basic description	<p>This page describes the archives of the National Archive as open records. Archival description is information about the contents of records, regardless of their physical medium (paper, electronic media, film tape, etc.). Archival descriptions are based on all archival search tools. All archive descriptions are presented hierarchically, from the most general description unit like archive or collections - to the most detailed level description unit - archived entity. According to the principles of archival descriptions, the archival material is described as highly as possible at the level of the description hierarchy, in order to avoid repeating the same information at lower levels.</p> <p>The process of providing the National Archives content as Open Data has been completed with the support of the European Union. The creation of the apeEAD data was financed from the project "Open Archives" carried out by the National Archives under the measure for the promotion of the information society of the Operational Program for the Development of the Estonian Economic Environment through APEX, RDF.</p>
Owner	National Archives of Estonia
Contact	Kuldar Aas (National Archives of Estonia) Kuldar Aas <Kuldar.Aas@ra.ee>
Type (sub-type or initiative specificity)	Open Data Portal
IPR	Metadata published as open source may be used in accordance with the CC0 license, digital images of the material referred to in the files are subject to CC-BY license.
Technical description	
Technologies (language, infrastructure, etc.)	N/A
Architecture	N/A
Features	N/A
Data models, formats	<p>The Open Data is presented in two XML formats, apeEAD and RDF:</p> <ul style="list-style-type: none"> The apeEAD format is created by the archivesportaleurope.net development team of the European Archives Portal and is used for machine-readable archival descriptions in both the archives portal and elsewhere. The description of the data structure is presented in

Evaluation criteria	Description of the evaluation criteria
	<p>the apeEAD and XLink XML schemas. Each file in the .xml extension is aggregated from the data for each description unit for one archive. In addition, the same XML files are also packed in five large zip files.</p> <ul style="list-style-type: none"> • RDF is a universal format for providing data and interconnections. Within the Open Source RDF files of the National Archives, various specific standards are used, Dublin Core, FOAF, DC Period, OWL, RDFS, SKOS, LOCAH, VCARD, ARCH, BIBO, MODS. The ontologies of the National Archives are also separate ministries, the descriptions of which are given in the section "Ontologies." There is a separate .rdf file extension for each description unit. They are aggregated into zip files at the highest level of the description units, i.e. for each archive / collection there is one zip file for all this unit and its subordinate data. In the case of digitised records, references to digital images are also included within RDFs.
Archival Data models, formats	N/A
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	N/A
Open Data formats	<p>Concatenated files (apeEAD formatted XML files into zipped files):</p> <ul style="list-style-type: none"> • Historical archive • State Archives • Movie archive • Land archives • Tallinn City Archives
Linked Data functionalities	N/A
Linked Data formats	<p>Data structure - (RDF data file structure description as commented sample files):</p> <ul style="list-style-type: none"> • Description Units • Periods • Individual bodies • Fields • Chinese keywords • Place names • Digital Copies <p>And the National Archives Ontologies (RDF files of individual ontological elements aggregated into ZIP files) to be used in descriptive unit openings:</p> <ul style="list-style-type: none"> • Periods • Individual bodies • Fields

Evaluation criteria	Description of the evaluation criteria
	<ul style="list-style-type: none"> • Chinese keywords • Place names
Data transformation methods and tools	N/A
Data access / consumption	There is a web page to download the archives files of the National Archives: http://opendata.ra.ee/
Data reusability	RDF files can be loaded in a semantic repository and be shared through a SPARQL endpoint.

Table 119: Estonian Open Data Portal

14.1.5 Europeana

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-005
Initiative name	Europeana
URL	https://www.europeana.eu/portal/en
Basic description	Europeana.eu is the EU digital platform for cultural heritage. More than 3,000 institutions across Europe have contributed to Europeana. These range from major international names like the Rijksmuseum in Amsterdam, the British Library and the Louvre to regional archives and local museums from every member of the European Union. Together, their assembled collections let users explore Europe's cultural and scientific heritage from prehistory to the modern day.
Owner	European Commission
Contact	Web: https://pro.europeana.eu/contact-us Mail: info@europeana.eu
Type (sub-type or initiative specificity)	Open Data Portal
IPR	The Europeana Data Exchange Agreement (DEA) is the central element of the Europeana Licensing Framework. It established that Europeana publishes metadata it receives from its data providers under the terms of the Creative Commons Zero Universal Public Domain Dedication (CC0).
Technical description	
Technologies (language,	N/A

Evaluation criteria	Description of the evaluation criteria
infrastructure, etc.)	
Architecture	N/A
Features	N/A
Data models, formats	XML, JSON, RDF
Archival Data models, formats	<p>EDMObjectTemplatesProviders</p> <ul style="list-style-type: none"> • edm:ProvidedCHO • ore:Aggregation • edm:Webresource • edm:Agent • edm:Place • edm:Timespan • skos:Concept • cc:License • svcs:Service • skos:ConceptScheme • edm:Event • edm:PhysicalThing <p>EDMObjectTemplatesEuropeana</p> <ul style="list-style-type: none"> • edm:ProvidedCHO • ore:Proxy • ore:Aggregation • edm:Webresource • edm:Agent • foaf:Organisation • edm:Place • edm:Timespan • skos:Concept • cc:License • dcat:Dataset • edm:FullTextResource • svcs:Service • skos:ConceptScheme • edm:Event • edm:Physicalthing

Evaluation criteria	Description of the evaluation criteria
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	<ul style="list-style-type: none"> • Dataset search and exploration • List of tools to browse, visualise or consume data • List of applications (e.g. showcases) that reuse the data • List of curated datasets (https://pro.europeana.eu/data/charters-from-the-biblioteca-de-catalunya for example)
Open Data formats	The XML Schema for the Europeana Data Model allows for automatic validation of EDM metadata. The EDM mapping rules are defined against the official release of the EDM XML Schema and the required Schematron rules (embedded in the schema). The EDM Validation document explains how to make use of the validation rules with the Oxygen XML editor.
Linked Data functionalities	Semantic search through the SPARQL endpoint: http://sparql.europeana.eu/sparql
Linked Data formats	The data can be downloaded in JSON or RDF: https://old.datahub.io/dataset/europeana-lod
Data transformation methods and tools	Europeana provides a standardisation methodology on how to process and present data relating to cultural heritage collections, making it easier for audiences to find it, share it and use it. Using the standardisation tools like the Europeana Data Model and international interoperable rights statements help to make it easy and rewarding for cultural heritage institutions to share high-quality collections with a global audience.
Data access / consumption	APIs: <ul style="list-style-type: none"> • REST API: The Europeana REST API allows you to build applications that use the wealth of cultural heritage objects stored in the Europeana repository. The API uses the standard web technology of REST calls over HTTP. Responses are returned in the popular JSON format.

Evaluation criteria	Description of the evaluation criteria
	<ul style="list-style-type: none"> • <u>Provider and Datasets API</u>: API calls permit to retrieve information about Europeana data providers and datasets that comprise the repository. • <u>The Europeana OAI-PMH Service</u> allows you to harvest metadata using Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH v2.0). The OAI-PMH Service allows you to harvest the entirety, or a selection, of all Europeana metadata, so that it can be integrated into other services or applications. • <u>Linked Open Data</u> is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources. All Europeana datasets can be explored and queried through a SPARQL endpoint.
Data reusability	Through SPARQL endpoint, the data or datasets can be uploaded and reused in specific context.

Table 120: Europeana

14.1.6 Archives Portal Europe Foundation

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-006
Initiative name	Archives Portal Europe Foundation
URL	http://www.archivesportaleuropefoundation.eu/
Basic description	<p>Funded by the European Commission between 2009 and 2015, the European archives community has realised a tremendous achievement: the creation of the Archives Portal Europe. To guarantee the sustainability of this unique aggregation and publication platform and to ensure that contributing content to the Archives Portal Europe will always be free of charge, the national archives of the participating countries – the driving forces behind the former projects – have established the Archives Portal Europe Foundation. The Foundation is a legal entity under Dutch law and has been allocated a budget and responsibilities to further develop the portal from 1 October 2015 onwards.</p> <p>In addition, the portal acts as a data aggregator for Europeana which displays digital heritage objects held in a wide range of cultural and scientific institutions around Europe, thereby helping to preserve knowledge about European cultural heritage for future generations.</p>

Evaluation criteria	Description of the evaluation criteria
	The Archives Portal Europe is the result of the APEnet project ⁵³ (2009-2012) and the APEx project ⁵⁴ (2012-2015), both funded by the European Commission.
Owner	Archives Portal Europe Foundation
Contact	Address: Prins Willem Alexanderhof 20, The Hague, 2595 BE, The Netherlands Mail: archivesportaleurope@gmail.com
Type (sub-type or initiative specificity)	Archives Portal
IPR	The code of our tools is available on GitHub under the European Union Public Licence.
Technical description	
Technologies (language, infrastructure, etc.)	<p>The Archives Portal Europe provides tools to support the participating institutions with data management and preparation of their content for presentation in the Archives Portal Europe. They can decide themselves if they want to use the local Data Preparation Tool or to work centrally in the portal's back-end, the dashboard.</p> <p>The realisation of all the tools is based on existing data provided by the partners which is already structured in "local" versions of archival exchange standards such as EAD (Encoded Archival Description), EAC-CPF (Encoded Archival Context - Corporate Bodies, Persons and Families) and METS (Metadata Encoding and Transmission Standard) or drawn from databases compliant with ISAD(G) (General International Standard Archival Description) and ISAAR-CPF (International Standard Archival Authority Record for Corporate Bodies, Persons and Families) to describe the record creators and ISDIAH (International Standard for Describing Institutions with Archival Holdings) to describe the archival institutions themselves.</p>
Architecture	N/A
Features	N/A
Data models, formats	<p>The Archives Portal Europe Wiki is a work in progress. In due time it will contain all information and documentation provided by the APEx project, which will end per the 30th of September 2015.</p> <p>For the Archives Portal Europe the APEx project and its predecessor the</p>

⁵³ APEnet project: <http://www.apenet.eu/>

⁵⁴ APEx project: <http://www.apex-project.eu/index.php/en/>

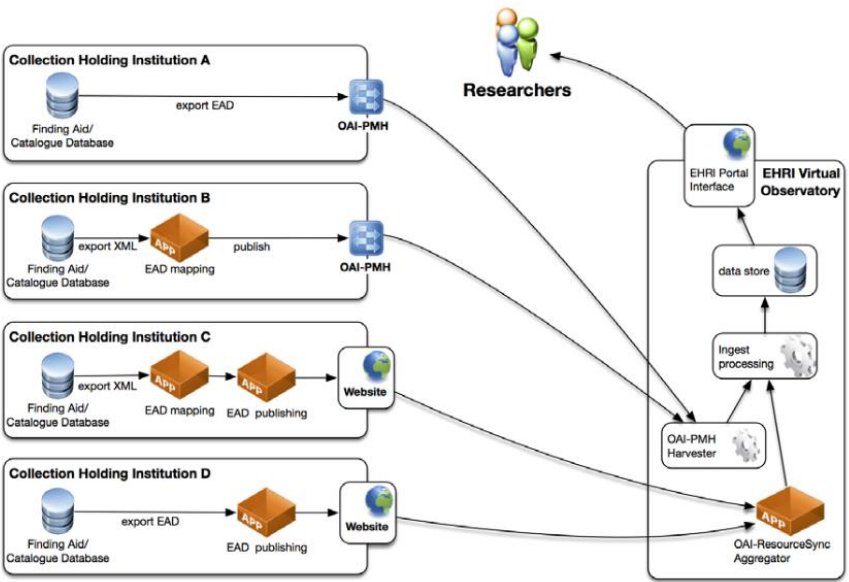
Evaluation criteria	Description of the evaluation criteria
	<p>APENet project (2009-2012) implemented a common framework across Europe that uses existing global archival standards targeted for a European context. By this, the acceptance of trans-national standards is supported and promoted as well as their use is encouraged in order to secure interoperability with the various national and institutional archives portals.</p> <p>Participation in Archives Portal Europe is open to all European archival repositories that can contribute descriptions of their holdings that are structured in accordance with the appropriate international archival standards (either in EAD, EAC-CPF, EAG and METS format or in a format that can be converted into EAD, EAC-CPF, EAG and METS preferably by themselves with the tools delivered by the APEx project).</p> <p>Read about the usage of international archival standards in the Archives Portal Europe:</p> <ul style="list-style-type: none"> • apeEAD • apeEAC-CPF • apeMETS • EAD3 • EAG2012
Archival Data models, formats	<p>The main goal of the Archives Portal Europe is to meet user expectations by providing seamless access to various archival resources held throughout Europe. To achieve this, we have developed common European profiles of the international XML schemas EAD (Encoded Archival Description), EAC-CPF (Encoded Archival Context - Corporate Bodies, Persons and Families), EAG (Encoded Archival Guide) and METS (Metadata Encoding and Transmission Standard) that act as pivotal formats enabling the Archives Portal Europe to deliver standardised data to Europeana.</p>
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	N/A
Open Data formats	N/A
Linked Data	N/A
Linked Data formats	N/A
Data transformation methods and tools	N/A
Data access / consumption	N/A

Evaluation criteria	Description of the evaluation criteria
Data reusability	N/A

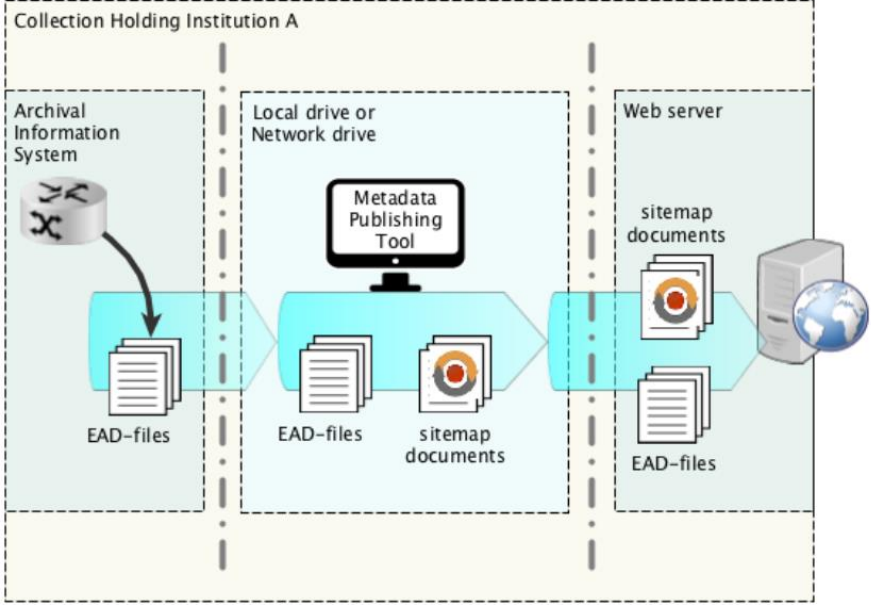
Table 121: Archives Portal Europe Foundation

14.1.7 H2020 Holocaust archiving

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-007
Initiative name	European Holocaust Research Infrastructure, EHRI H2020 project
URL	https://www.ehri-project.eu/
Basic description	EHRI is a European research project (H2020) that provides online access to information about dispersed sources relating to the Holocaust through its Online Portal, and tools and methods that enable researchers and archivists to collaboratively work with such sources.
Owner	European Commission
Contact	Address: NIOD Institute for War, Holocaust and Genocide Studies (coordinator EHRI) Herengracht 380, 1016 CJ Amsterdam, Netherlands Web: https://ehri-project.eu/content/contact-us Mail: info@ehri-project.eu
Type (sub-type or initiative specificity)	H2020 Research project
IPR	Open source
Technical description	
Technologies (language, infrastructure, etc.)	<p>The EHRI project has developed software services to assist the data integration process. To what extent the service is usable for a Collection Holding Institutions (CHI) depends on the way the local data infrastructure is organised. E.g. whether metadata on archival holdings are available in a digital form, its format and how the information infrastructure is able to communicate with the outside world.</p> <p>The software services are represented by the orange boxes the figure below:</p>

Evaluation criteria	Description of the evaluation criteria
Architecture	 <p>The diagram illustrates the EHRI distributed architecture. On the left, four 'Collection Holding Institution' boxes (A, B, C, D) show the process of data export and publication. Institution A exports EAD from its database to OAI-PMH. Institution B exports XML, uses an EAD mapping application, and publishes to OAI-PMH. Institution C exports XML, uses EAD mapping and publishing applications, and publishes to a website. Institution D exports EAD and publishes to a website. On the right, the 'EHRI Virtual Observatory' consists of an OAI-PMH Harvester that pulls data from OAI-PMH and websites, an OAI-ResourceSync Aggregator, an Ingest processing component, a data store, and an EHRI Portal Interface. Researchers are shown at the top, interacting with the OAI-PMH and the Portal Interface.</p> <p>EHRI is using a distributed architecture in which every institution can produce standardised data that can be synchronised with a central repository.</p>
Features	<ol style="list-style-type: none"> <p><u>EAD Creation Tool (ECT)</u> ftp://ftp.ontotext.com/pub/EHRI/conversion_tool/ EAD Creation Tool (ECT) is a tool that can be installed by any institutions and that permit to produce and manage metadata to archive digital content in a standardised way.</p> <ul style="list-style-type: none"> • Metadata conversion in a local format into metadata in the EAD⁵⁵ format (Encoded Archival Description) • Creation and management of the EAD metadata • Choose the mapping configuration file of your organisation or use your own; • Edit the mapping configuration to suit your needs; • Use a custom transformation type; • Convert your data to the EAD 2002 format; • Preview all validation inconsistencies; • Generate a well formed EAD 2002 data file. <p><u>Metadata Publishing Tool (MPT)</u> https://github.com/EHRI/rspub-gui/releases Metadata Publishing Tool (MPT) is a desktop application that facilitates the publishing of resources and sitemaps in conformance with the ResourceSync Framework Specification. Metadata Publishing Tool (rspub-gui and rspub-core) was developed by Data Archiving and Networked Services (DANS-KNAW) under auspices of</p>

⁵⁵ EAD format: <https://www.loc.gov/ead/>

Evaluation criteria	Description of the evaluation criteria
	<p>the European Holocaust Research Infrastructure (EHRI). MPT functionalities:</p> <ul style="list-style-type: none"> • Collect and import resources from various places within the organisation; • Select relevant resources; • Create ResourceSync sitemap metadata on relevant resources; • Export resources and sitemaps to the web server; • Verify that the exposed URL's are correct and our ResourceSync site ready to be harvested by a Destination.  <p>The diagram illustrates the workflow of the Metadata Publishing Tool (MPT) within a Collection Holding Institution A. It is divided into three main sections: <ul style="list-style-type: none"> Archival Information System: This section contains a database icon and a stack of documents labeled 'EAD-files'. An arrow points from the AIS to the Local drive. Local drive or Network drive: This central section contains a computer monitor icon labeled 'Metadata Publishing Tool'. It shows 'EAD-files' being processed into 'sitemap documents'. Web server: This section contains a server rack icon and a globe. It shows 'sitemap documents' and 'EAD-files' being exported from the local drive to the web server for public access. </p>
Data models, formats	N/A
Archival Data models, formats	<ul style="list-style-type: none"> • Input files– XML, XML EAD 1, CSV; • Output files– EAD 2002; • Mapping files– XLS, XLSX, Google Sheet. <p>The limit of such an approach is that EAD is a very permissive standard, where each institution (each archivist), and each piece of software can have their own way of creating EAD, and the same material can be described in totally different ways.</p>
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	N/A
Open Data formats	N/A
Linked Data	N/A

Evaluation criteria	Description of the evaluation criteria
Linked Data formats	N/A
Data transformation methods and tools	N/A
Data access / consumption	N/A
Data reusability	N/A

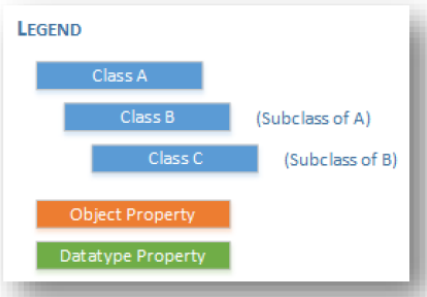
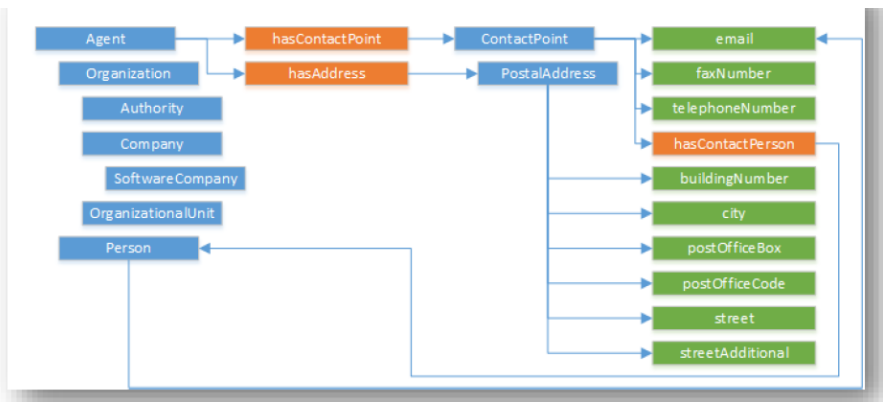
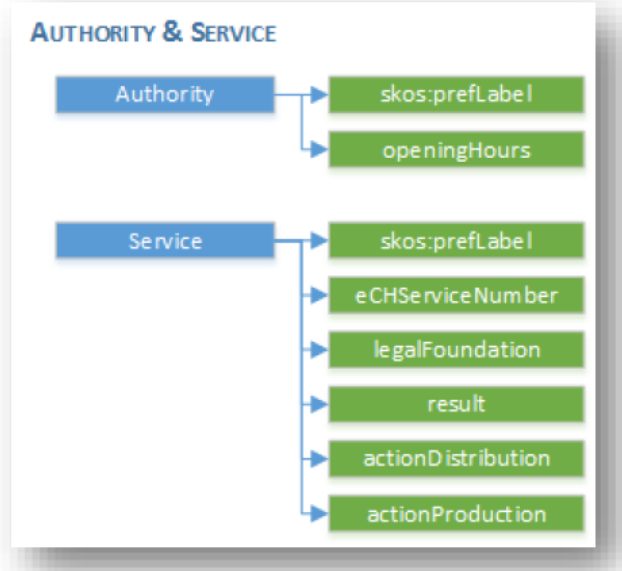
Table 122: H2020 Holocaust Archiving

14.1.8 Swiss – Federal Archives

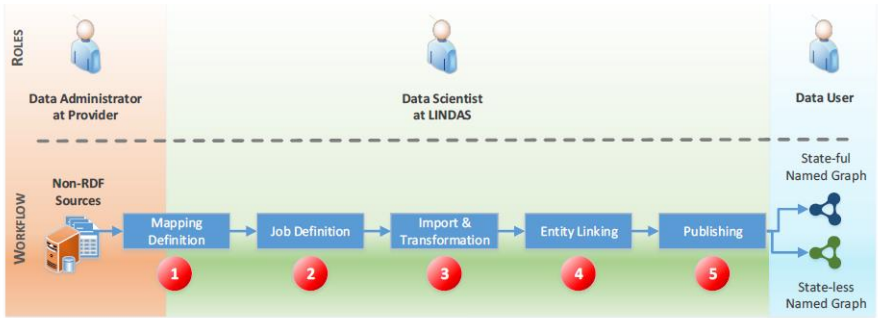
Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-008-b
Initiative name	Swiss – Federal Archives
URL	https://www.egovernment.ch/lindas/ https://www.egovernment.ch/en/umsetzung/e-government-schweiz-2008-2015/lindas/ http://www.lindas-data.ch
Basic description	<u>LINDAS Linked Data Service</u> Swiss authorities can use the LINDAS Linked Data Service at www.lindas-data.ch to make their data available as “Linked Data”. The platform also gives developers access to the data. The LINDAS service at www.lindas-data.ch is a prototype. It was developed in 2015 by the State Secretariat for Economic Affairs SECO and has been operated by the Federal Archives since the start of 2017. Further information can be found on the service’s website.
Owner	Swiss Federal Archives
Contact	Swiss Federal Archives Archivstrasse 24 3003 Bern bundesarchiv@bar.admin.ch
Type (sub-type or initiative specificity)	Linked Data Portal
IPR	According to the Open Data Swiss Free Use, the data can be freely copied, modified and distributed: https://opendata.swiss/en/terms-of-use/
Technical description	

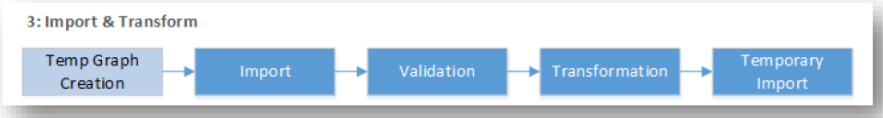
Evaluation criteria	Description of the evaluation criteria
Technologies (language, infrastructure, etc.)	<p><u>The “Linked Data Platform” project:</u></p> <p>In parallel, the Federal Archives are conducting a Linked Data project designed to develop the service further.</p> <p>Its vision is to create a service that can convert data into Linked Data simply and so make connecting them to other data much easier. The first application scenarios have already been tested and implemented using the prototype at www.lindas-data.ch.</p> <p>Here is the list of the application scenarios that have been implemented or are currently being clarified:</p> <ul style="list-style-type: none"> • aLOD Archival Linked Open Data (Federal Archives and other institutions) • Animal disease outbreak investigation (Federal Food Safety and Veterinary Office) • Excerpt from The Swiss Book (National Library) • Official data / Federal Directory and list of prices and services (Canton of Graubünden, provider of government solutions) • Energy management of rail wagons (MeteoSwiss and SBB) • Corporate data subset for a one-stop shop for companies (State Secretariat for Economic Affairs and Federal Office of Justice) • Stops and additional information (SBB) • Historical federal budgets (Federal Archives) • Historicised Official Commune Register (Federal Statistical Office) • Federal Archives • Environmental data (Federal Office for the Environment) <p>LINDAS_Technological_Concepts_v1.1.pdf⁵⁶ document contains all the technical details, especially about the architecture. The semantic repository that can be connected to the system are two different triple stores:</p> <ul style="list-style-type: none"> • Ontos OntoQUAD • or OpenLink Virtuoso
Architecture	N/A
Features	N/A
Data models, formats	N/A
Archival Data models, formats	LINDAS Top-Level Ontology, followed all the entities represented as ontological model

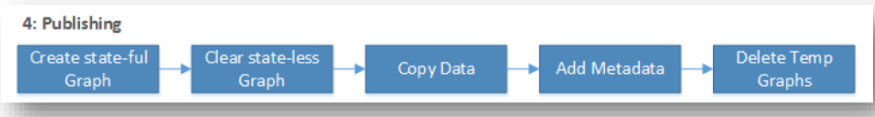
⁵⁶ LINDAS_Technological_Concepts_v1.1.pdf: https://www.egovernment.ch/index.php/download_file/force/752/3658/

Evaluation criteria	Description of the evaluation criteria
	<p>LEGEND</p>  <p>Agent and Contact information:</p>  <p>Authority and service:</p> 
Archival Data models, formats (continued)	Responsibility:

Evaluation criteria	Description of the evaluation criteria
	<p data-bbox="548 281 711 310">RESPONSIBILITY</p>  <p data-bbox="516 562 886 592">Software solution and provider:</p> <p data-bbox="535 617 808 642">SOFTWARE SOLUTION & PROVIDER</p>  <p data-bbox="516 1113 1032 1142">e-Government Service and Implementation:</p> <p data-bbox="535 1167 753 1188">EGOVSERVICE & IMPLEMENTATION</p>  <p data-bbox="516 1457 1438 1608">The current version of the LINDAS top level ontology is a first step towards a standardised vocabulary for Swiss Linked Authority Data. Especially due to new data sources or data consumer needs, new classes or properties maybe required. In order to unify the different needs, some work will be required.</p> <p data-bbox="516 1654 1107 1684">Currently, some activities are still remaining to do:</p> <ul data-bbox="565 1696 1438 1898" style="list-style-type: none"> • Classes and properties to describe contact and address information are modelled to comply with existing semantic vocabularies, primarily, schema.org as one of the most wide-spread ones. In future, it is may be required to be more conform to the eCH standards eCH-0010 and eCH-0046. However, at the current stage, they are too complex for the current needs.

Evaluation criteria	Description of the evaluation criteria
	<ul style="list-style-type: none"> • It should also be evaluated, which of the other existing eCH standards should be reused. A good foundation will be the information in eCH-0177. • It should also be investigated, which of the related vocabularies their classes and properties – could additionally be linked.
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	N/A
Open Data formats	N/A
Linked Data functionalities	<p>The main functionalities are the:</p> <ul style="list-style-type: none"> - data import, - data conversion in RDF, - and data publication. <p>The workflow is described below in the section “Data transformation methods and tools”.</p>
Linked Data formats	RDF SPARQL
Data transformation methods and tools	<p>Data import workflow:</p>  <p>The semi-automatic workflow to import, convert, and publish the data contains the following five steps:</p> <ol style="list-style-type: none"> 1. Mapping Definition: The authority information is offered in data formats, like relational data or spreadsheets. Also the data access will differ between the providers; e.g., a database connection, a REST API, or a single file upload. For that reason, in the first step we need to define how these sources map to the semantic vocabulary used in LINDAS. Together with the data provider, the data scientist selects the target class from the LINDAS vocabulary and defines the required properties, which represent a column in the CSV file. 2. Job Definition: After one or more mappings are defined, it is possible setup the import jobs. A job comprises a mapping, a target named graph with

Evaluation criteria	Description of the evaluation criteria
	<p>predefined access permissions for storing the data, and a license under which the data should be published. Furthermore, the data scientist needs to decide if the job should be executed manually or by the platform itself. In the latter case, it is possible to schedule the jobs for their iterative execution, like every day or week, what allows for keeping the data up-to date automatically.</p> <p>3. Import and Transformation: If a job is defined, the associated configuration is used to call the required services to load the data from its source, transform it to RDF, and to persist it in a temporary graph within the quad store.</p> <p>4. Entity Linking: Since the data providers have different unique IDs for maybe the same entity, we have the task to interlink them as good as possible. Thus, after the import the data scientist had to choose graphs, classes and data properties to generate owl:sameAs links. Usually, this are geo spatial and service information of authorities.</p> <p>5. Publishing: In the last step of the workflow, the data gets published. Subtasks are to combine information also from other named graphs, to add metadata (e.g. about the license or provenance information), or to move the data to the target graph, which was specified in the job configuration. It is also possible to do a revision. This includes to create a new stateful graph with the previous data of the target graph. Stateful means that the graph URI comprise the date (and time) information of its creation, i.e., <code>\url{http://lindas-data.ch/dataset/bvch/2015-01-24}</code>. The target graph with new data, which is stateless (<code>\url{http://lindas-data.ch/dataset/bvch}</code>), gets a backlink to the stateful one, thus, the data history could be traversed later on.</p> <p>(CSV) Import and transformation</p>  <pre> graph LR A[Temp Graph Creation] --> B[Import] B --> C[Validation] C --> D[Transformation] D --> E[Temporary Import] </pre> <p>If a job is started, the process above is carried out. Although, the following description focus on CSV to RDF import, it is basically the same for the D2RQ-based RDB import.</p> <ol style="list-style-type: none"> 1. For the data import, a temporary graph is required, which is used by services. It is unique by an UUID or timestamp as last part of the URL the graphs is created by the backend and it is private for the logged in user. 2. The (CSV2RDF) service is called with information of the job (temporary named graph, mapping, data source and the SPARQL endpoint URL). With this information, it queries the data. 3. After the CSV file is received, it is validated against the mapping definition. Therefore, they use the header of the CSV and the mapping file. If the structure has changed, the import is stopped. A validation is not executed for RDB import.

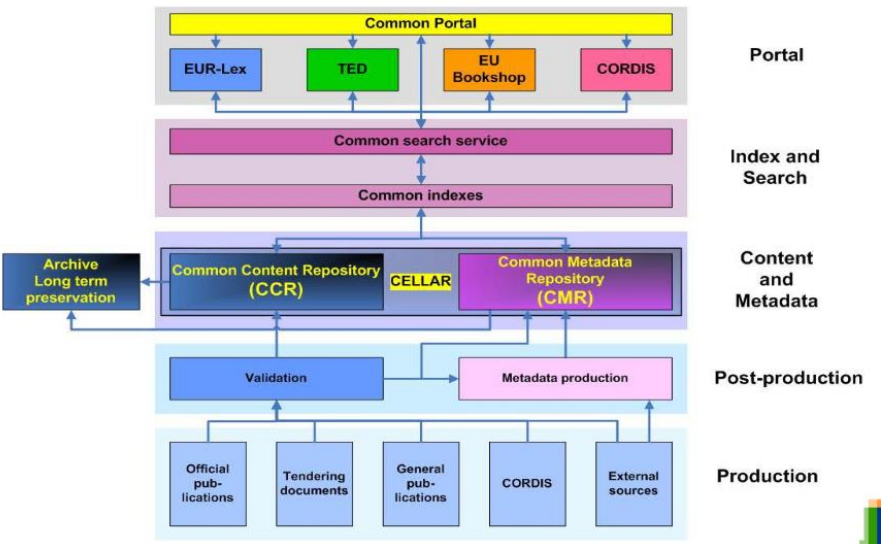
Evaluation criteria	Description of the evaluation criteria
	<p>4. Based on the mapping, the data is converted to RDF. 5. Finally, the data is uploaded to the temporary graph via the SPARQL endpoint.</p>
<p>Entity Linking Data publication</p>	<p><u>Entity Linking:</u> After importing data from various source into RDF graphs, every resource has its own ID. Unfortunately, this is also true for e.g. cantons or communes that are unique in general but could be duplicated if they are imported from different sources. A named entity service could solve this problem. Since it is not available for now, we integrated an entity linking service which allows for the semi-automatic creation of owl:sameAs links between entities. Besides the LINDAS internal linking, it is possible to link to other existing resources like DBpedia.</p> <p><u>Publishing:</u></p>  <pre> graph LR A[Create state-ful Graph] --> B[Clear state-less Graph] B --> C[Copy Data] C --> D[Add Metadata] D --> E[Delete Temp Graphs] </pre> <p>If the data transformation has finished successfully, the job scheduler initiates the publishing service. Therefore, it passes:</p> <ul style="list-style-type: none"> • the SPARQL endpoint • the list of (temporary) graphs that should be taken as data source including a flag telling if the NG should be deleted after the import • the target state-less named graph • the flag if the information should be just overwritten or also stored to a state-full graph • the metadata to add to the target graph to the publishing service. <p>The single steps of the publishing service are self-explanatory:</p> <ol style="list-style-type: none"> 1. If the existing data should be backed up in a state-full graph, it is created by adding the date (and maybe the timestamp to the URL. Afterwards, the data is read from the state-less graph and copied to the versioned one. Finally, metadata is added (e.g. the link to the stateless graph). 2. The data in the stateless graph is cleaned via SPARQL 3. The data from the import graphs is copied to the target graph. 4. The required metadata is added. Additionally, the current time stamp and, if step 1) is executed, the link to the former graph. 5. Finally, if defined, the temporary import graphs are deleted (not only cleaned)
<p>Data access / consumption</p>	<p>SPARQL endpoint:</p> <ul style="list-style-type: none"> • the data are stored in the semantic repository and shared through the SPARQL endpoint • Data can be searched with SPARQL queries
<p>Data reusability</p>	<p>SPARQL endpoint:</p>

Evaluation criteria	Description of the evaluation criteria
	<ul style="list-style-type: none"> the data are stored in the semantic repository and shared through the SPARQL endpoint Data can be consumed with SPARQL queries

Table 123: Swiss Federal Archives

14.1.9 Cellar (Publication Office)

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-009
Initiative name	CELLAR
URL	https://data.europa.eu/euodp/data/dataset/sparql-cellar-of-the-publications-office
Basic description	<p>The CELLAR is the central content and metadata repository of the Publications Office of the European Union.</p> <p>The CELLAR makes available at a single place all the metadata and digital content managed by the Publications Office in a harmonised and standardised way in order to:</p> <ul style="list-style-type: none"> to guarantee to the citizen a better access to law and publications of the European Union; to encourage and facilitate reuse of content and metadata by professionals and experts; to preserve content and metadata and access to contents and metadata over time.
Owner	Publication Office (European Commission)
Contact	Marc Wilhelm Küster marc.kuster@publications.europa.eu
Type (sub-type or initiative specificity)	Linked Open Data Portal
IPR	Open source
Technical description	
Technologies (language, infrastructure, etc.)	<p>Use of the Semantic Web technologies:</p> <ul style="list-style-type: none"> Linked Data models (RDF, OWL, SPARQL) Taxonomies represented in SKOS Semantic repository is implemented with Virtuoso solution <p>CELLAR user manual - technical documentation https://data.europa.eu/euodp/data/dataset/sparql-cellar-of-the-publications-</p>

Evaluation criteria	Description of the evaluation criteria
	office/resource/2895b9e9-4470-4eb9-a5fd-d46403216786
Architecture	 <p>The diagram illustrates a multi-layered architecture. At the top is the 'Common Portal' (yellow), which connects to four services: 'EUR-Lex' (blue), 'TED' (green), 'EU Bookshop' (orange), and 'CORDIS' (pink). Below this is the 'Index and Search' layer, containing 'Common search service' and 'Common indexes'. The 'Content and Metadata' layer includes 'Common Content Repository (CCR)' (dark blue), 'CELLAR' (yellow), and 'Common Metadata Repository (CMR)' (purple). A box for 'Archive Long term preservation' is connected to the CCR. The 'Post-production' layer consists of 'Validation' and 'Metadata production'. The 'Production' layer at the bottom includes 'Official publications', 'Tendering documents', 'General publications', 'CORDIS', and 'External sources'. Arrows indicate the flow of data and services between these components.</p>
Features	<ul style="list-style-type: none"> • Multilingual aspects (for example in EuroVoc) • Interoperability (Standards adoption) <ul style="list-style-type: none"> ○ METS (Metadata encoding and transmission standard) → ingestion protocol ○ Dublin Core → core metadata definition ○ FRBR compliant → data model/ontology ○ Linked Open Data (LOD) → access/reuse <ul style="list-style-type: none"> ▪ Web-friendly ("RESTful") Interface ▪ RDF / OWL ○ Standard Query Language (SPARQL) → access/reuse • Interoperability (Standards adoption) <ul style="list-style-type: none"> ○ Core metadata (Restricted shared set of metadata for each resource based on Dublin Core) → enable global search ○ Common authority tables → harmonise metadata ○ Exchange protocol for EU legislative procedures → ○ Interoperability ○ European Legislative Identifier (ELI) (initiative of EU member countries, supported by the Publications Office) → interoperability
Data models, formats	RDF, OWL, SKOS, SPARQL
Archival Data models, formats	<p>Common data model: the ontology is based on the FRBR model:</p> <ul style="list-style-type: none"> • Work: is a "distinct intellectual or artistic creation." • Expression: is "the specific intellectual or artistic form that a work takes each time it is 'realised.'" • Manifestation: is "the physical embodiment of an expression of a work. As an entity, manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form."

Evaluation criteria	Description of the evaluation criteria
	<pre> graph TD Endeavour -- realization --> Work Work -- realizationOf --> Endeavour Work -- embodiment --> Expression Expression -- embodimentOf --> Work Expression -- exemplar --> Manifestation Manifestation -- exemplarOf --> Expression Manifestation -- part/partOf --> Item Item -- relatedEndeavour --> Manifestation </pre>
Other relevant technical aspects	<p>In September 2016:</p> <ul style="list-style-type: none"> • 8 million requests per day served on average (peaks > 20 million) • 100'000 SPARQL queries per day • > 1 million different in > 10 million linguistic versions and > 28 million items • > 230 million of persistent identifiers • > 1500 million triple in Oracle RDF store • 4 TB Oracle DB (compressed) • Content (in Fedora repository) > 17,5 TB • 120 million files in Fedora
Relevant aspects for Data Archiving Management	
Open Data functionalities	N/A
Open Data formats	N/A
Linked Data	<ul style="list-style-type: none"> • Control Data <ul style="list-style-type: none"> ○ Ontologies / Common Data Model ○ Thesauri / authority tables (EuroVoc) • Instance Data <ul style="list-style-type: none"> ○ Instances with the classification scheme (Work > Expression > Manifestation) ○ URIs: http://publications.europa.eu/resource/{ps-i}/{obj-id}

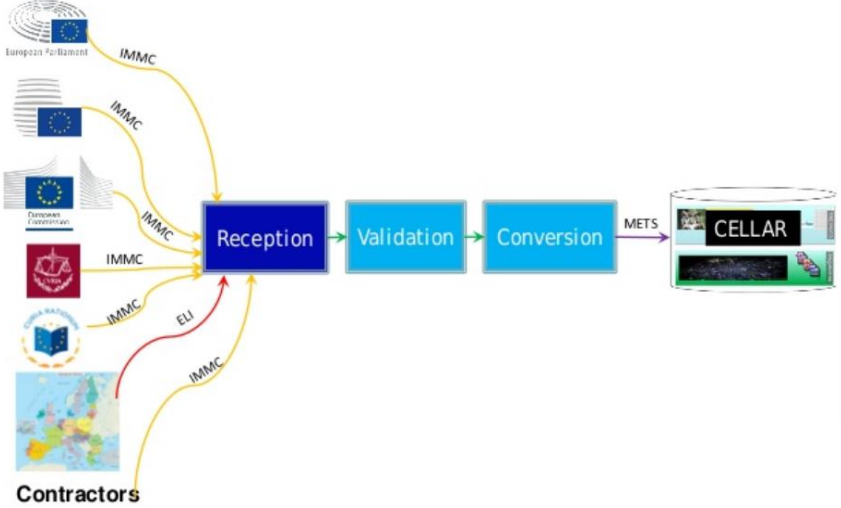
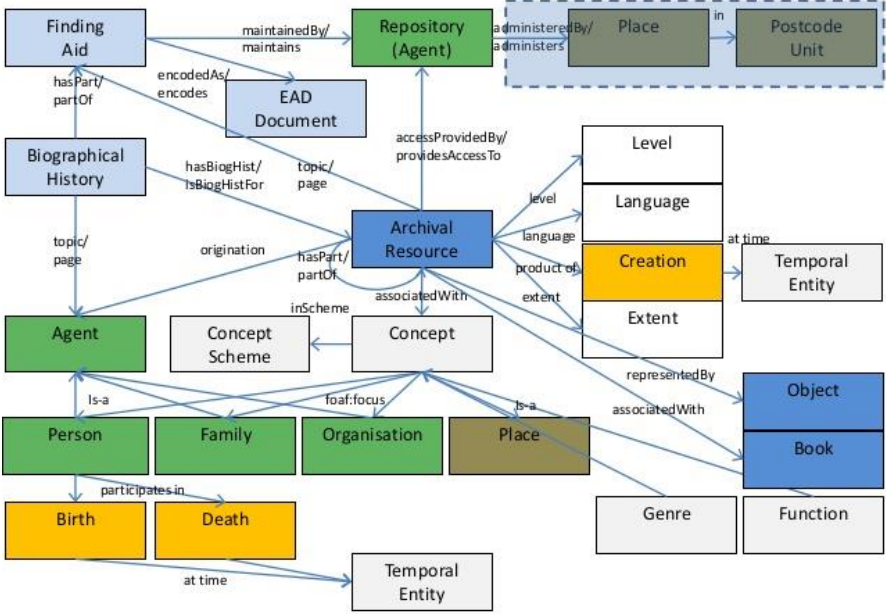
Evaluation criteria	Description of the evaluation criteria
Data transformation methods and tools	
Data access / consumption	<p>Data access is provided by:</p> <ul style="list-style-type: none"> • Notification /RSS • SPARQL endpoint • Direct access / RESTful WS • EUR-Lex OP portal
Data reusability	<p>Data reuse is provided by:</p> <ul style="list-style-type: none"> • SPARQL endpoint

Table 124: Cellar (Publication Office)

14.1.10 Archives Hub (UK)

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-010
Initiative name	Archives Hub (UK)
URL	https://archiveshub.jisc.ac.uk/ (website) http://data.archiveshub.ac.uk/ (data portal)
Basic description	<p>The Archives Hub brings together descriptions of thousands of the UK's archive collections. Representing nearly 300 institutions across the country, the Archives Hub is an effective way to discover unique and often little-known sources to support researches.</p> <ul style="list-style-type: none"> • Use the Hub to instantly scan the archival landscape and bring together diverse sources held in repositories across the country. • Historians will find the Hub an essential tool for their research. • Postgraduate students can find new sources and make new

Evaluation criteria	Description of the evaluation criteria
	<p>connections for their dissertations.</p> <ul style="list-style-type: none"> • Educators can take advantage of the Hub to introduce students to primary sources for their course work. • Archivists and librarians can use the Hub in support of their work and in support of their users. <p>Within the collections represented on the Hub there is a huge diversity of content, from the archives of industries, institutions and researchers to the letters and manuscripts of writers and poets. New descriptions of archives are being added all the time, so the Hub is always worth a visit.</p> <p>The Archives Hub team work closely with contributors to make sure that they maintain and develop the best possible service for end users.</p>
Owner	Jisc's Manchester office
Contact	Mail: archiveshub@jisc.ac.uk
Type (sub-type or initiative specificity)	Linked Open Data portal for the archives
IPR	The LOCAH Linked Archives Hub data and content is licensed under a Creative Commons CC0 1.0 licence.
Technical description	
Technologies (language, infrastructure, etc.)	<p>The Archives Hub uses the Collections Information Integration Middleware or CIIM (pronounced “sim”) provided by Knowledge Integration.</p> <p>This is a modular suite of software which sits between the archive descriptions and the web site (or other end points). It uses Elasticsearch (Elastic, 2018), a search engine based on Apache Lucene (Lucene, 2018), to represent the Hub's large volumes of complexly structured descriptions.</p>
Architecture	N/A
Features	N/A
Data models, formats	<p>The Open Archives Initiative Protocol for Metadata Harvesting is a simple protocol that allows services to expose metadata for harvesting. This means that other services can regularly gather metadata.</p> <p>OAI-PMH is a metadata harvester that locates and aggregates metadata from different datasets. A metadata harvester can be used to integrate data from a number of sources into one catalogue; users can then search the OPAC, locate the metadata within the catalogue and be linked to the content at the original repository's site. OAI allows multiple forms of metadata to be exposed but mandates DC as a minimum. Many data providers do not provide richer metadata formats. Only the metadata needs to be stored locally and not the files themselves.</p>

Evaluation criteria	Description of the evaluation criteria
	<p>OAI-PMH is based on common standards: HTTP and XML. It is intended as a low-barrier solution, but a level of manual intervention in harvesting metadata is required and incremental harvesting may be problematic. There is a high level of flexibility in the protocol and in Dublin Core, which can create obstacles for service providers.</p>
<p>Archival Data models, formats</p>	<p>The descriptions are processed in order to create a store of aggregated content that is structured and potentially re-usable. We use the International Standard Archival Description (General), or ISAD(G), but we also recognise its shortcomings for the current online world. Index terms follow recognised rules or recognised sources (e.g. NCA Rules, UKAT).</p> <p>The format used to ingest descriptions is Encoded Archival Description (EAD). The descriptions are stored in JSON. Descriptions may be at collection level or they may be multi-level, down to individual item. It is the responsibility of the Hub contributors to create and submit descriptions for inclusion on the Hub.</p> <p>Archives Hub Model</p>  <p>The diagram illustrates the Archives Hub Model, a complex network of entities and their relationships. Key entities include: <ul style="list-style-type: none"> Finding Aid (blue box) which is maintained by a Repository (Agent) (green box) and encodes an EAD Document (blue box). Repository (Agent) is administered by a Place (green box) and provides access to Archival Resources (blue box). Archival Resource is associated with Agents (green box), Concepts (grey box), and Temporal Entities (grey box). It has a Level (white box), Language (white box), Creation (yellow box), and Extent (white box). Agents include Person, Family, Organisation, and Place (all green boxes). Person and Place are associated with Birth and Death (yellow boxes). Concepts include Concept Scheme (grey box) and are associated with Agents and Archival Resources. Temporal Entities include Level, Language, Creation, Extent, Object, Book, Genre, and Function (all white boxes). Object and Book are represented by Archival Resources. Postcode Unit (grey box) is associated with Place and Repository (Agent). </p>
<p>Other relevant technical aspects</p>	<p>N/A</p>
<p>Relevant aspects for Data Archiving Management</p>	
<p>Open Data functionalities</p>	<p>N/A</p>
<p>Open Data formats</p>	<p>N/A</p>
<p>Linked Data</p>	<p>The Archives Hub Linked Data available here was produced by the Jisc</p>

Evaluation criteria	Description of the evaluation criteria
	<p>funded LOCAH project dating back to 2010. Unfortunately, the data were not updated since 2013. The data should only be used for test and example purposes as it may change substantially and cannot be relied upon. The Archives Hub is hoping to produce up to date production level data at some point in the future, but at this time it is not possible to indicate if or when this will happen.</p> <p>The SPARQL endpoint for the dataset is:</p> <ul style="list-style-type: none"> • http://data.archiveshub.ac.uk/sparql <p>A query box for trying out SPARQL queries is available at:</p> <ul style="list-style-type: none"> • http://data.archiveshub.ac.uk:8000/test/ <p>An RDF dump of the dataset is available:</p> <ul style="list-style-type: none"> • http://data.archiveshub.ac.uk/dump/ <p>The Archives Hub EAD to RDF XSLT stylesheet is available:</p> <ul style="list-style-type: none"> • http://data.archiveshub.ac.uk/ead2rdf/
Linked Data formats	<p>The data references terms from (amongst others) the following RDF vocabularies:</p> <ul style="list-style-type: none"> • http://purl.org/dc/terms/ • http://xmlns.com/foaf/0.1/ • http://www.w3.org/2004/02/skos/core# • http://www.openarchives.org/ore/terms/ • http://linkedevents.org/ontology/ http://data.archiveshub.ac.uk/def/
Data transformation methods and tools	<p><u>Archives Hub EAD to RDF XSLT Stylesheet</u></p> <p>The Archives Hub EAD to RDF XSLT stylesheet encapsulates the Linked Data model developed by the LOCAH Project and provides a simple standards-based means to transform archival descriptions to Linked Data RDF/XML. The stylesheet is simple to reuse, and can be re-purposed by anyone wishing to transform archival descriptions in EAD form to Linked Data ready RDF/XML. It is made available under the Modified BSD License.</p> <p>The style sheet is available directly from http://data.archiveshub.ac.uk/xslt/ead2rdf-nons.xml</p>
Data access / consumption	<p>Data accessibility is provided by:</p> <ul style="list-style-type: none"> • SPARQL endpoint
Data reusability	<p>Data reuse is provided by:</p> <ul style="list-style-type: none"> • SPARQL endpoint

Table 125: Archive Hub (UK)

14.1.11 Social Archive - SNAC project

Evaluation criteria	Description of the evaluation criteria
---------------------	--

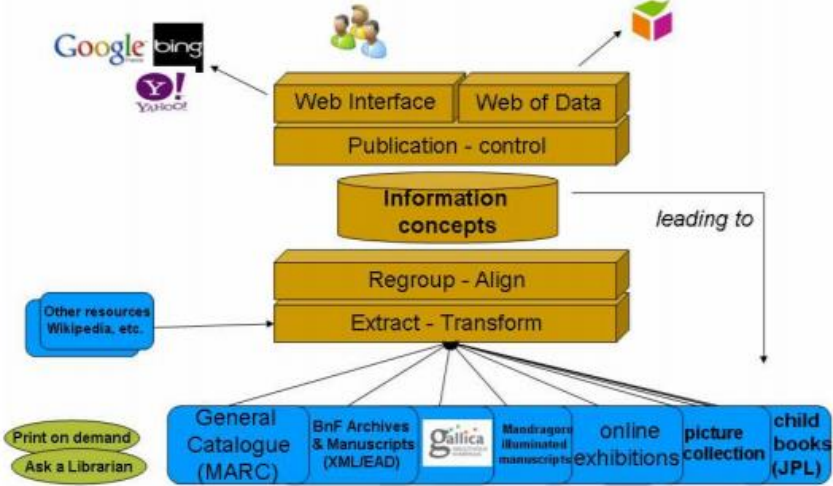
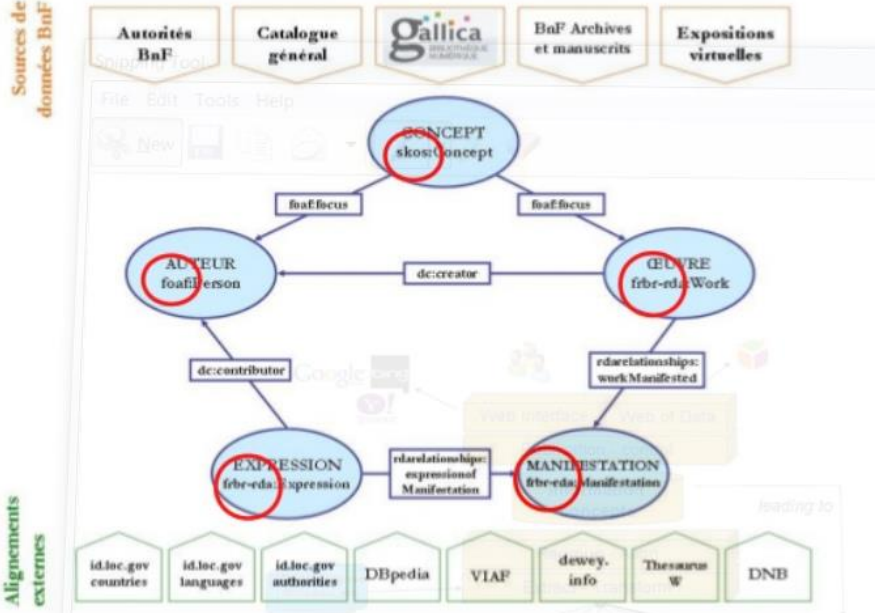
Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-011
Initiative name	Social Networks and Archival Context
URL	http://snaccooperative.org/
Basic description	The SNAC project aggregates authority data in regards to persons and organisations from all sorts of heterogeneous sources, in order to facilitate the re-use of authority data. The project involves a complex methodology for merging and matching very large volumes of semi-structured data and can be an important source of inspiration on how the EU institutions may exchange authority data.
Owner	Daniel Pitti
Contact	Mail: dpitti@Virginia.edu
Type (sub-type or initiative specificity)	Data portal
IPR	Open source software
Technical description	
Technologies (language, infrastructure, etc.)	XSLT scripts, Perl scripts, and data files to create EAC CPF (corporate bodies, persons, families) are available on https://github.com/snaccooperative/snac_eac_cpf_utils
Architecture	Collection of loose scripts to perform data transformation into a common standard
Features	Merging and matching authority records
Data models, formats	XML and RDF
Archival Data models, formats	EAC-CPF, MARC
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	Publish and share
Open Data formats	XML and RDF
Linked Data	XML and RDF
Linked Data formats	EAC-CPF

Evaluation criteria	Description of the evaluation criteria
Data transformation methods and tools	See https://github.com/snac-cooperative/snac_eac_cpf_utils
Data access / consumption	N/A
Data reusability	N/A

Table 126: Social Archive - SNAC Project

14.1.12 Bibliothèque Nationale de France (BNF)

Evaluation criteria	Description of the evaluation criteria
Basic description	
Initiative ID	LOD-INI-012
Initiative name	Bibliothèque Nationale de France (BNF)
URL	http://data.bnf.fr/
Basic description	The Bibliothèque nationale de France has designed a new project in order to make its data more useful on the Web. It involves transforming existing data, enriching and interlinking the dataset with internal and external resources, and publishing HTML pages for browsing by users and search engines. The raw data is also available in RDF following the principles of Linked Open Data architecture.
Owner	Bibliothèque Nationale de France (BNF)
Contact	Mail: romain.wenz@bnf.fr
Type (sub-type or initiative specificity)	Linked Open Data Portal
IPR	Open source
Technical description	
Technologies (language, infrastructure, etc.)	Library data can be difficult to find on the Web. At the BnF, it is of course possible to access all of the resources and services through our Library Website (www.bnf.fr). But, at present, few of them are indexed by search engines. And, even when they are, it is difficult to sort results from them. Some digital books, even when they are completely and freely available, are sometimes impossible to find if you don't already know they exist. The data.bnf.fr project can be a way to open the digital library Gallica to a wider public. Moreover, library catalogues are usually stored as relational databases: they are just no use for Web search engines. Users always

Evaluation criteria	Description of the evaluation criteria
	<p>access the BnF catalogues (mainly, the Main catalogue and the Archive and manuscript catalogue) through library portals, which they often simply don't know. As a matter of fact, users are very unlikely to find any of our resources directly from a search engine interface, unless they already know about us. Gallica platform is based on the software called CubicWeb which is a semantic web application framework licensed under the LGPL. http://data.bnf.fr/docs/databnf-presentation-en.pdf</p>
Architecture	
Features	N/A
Data models, formats	
Archival Data models, formats	<p>This way to articulate bibliographic data on the Web implies several choices. As a matter of fact, the aim of publishing HTML pages implies that BnF data</p>

Evaluation criteria	Description of the evaluation criteria
	<p>model basically enhance concepts that are relevant for creating a Web page. They chose to rely on the concepts of works, authors and subjects, which happen to be entities in the FRBR model, as they tried to make the data model compliant with the FRBR requirements. This Web interface is at the crossroads between the different resources we make available on the Web. It gathers different kinds of data at the right level: works, expressions and manifestations. For an author, users find all the links to the Web pages of the relevant works, by and about the author, in two different sections. For a work, there is a link to the author's page, but also to the different manifestations of the work (bibliographic resources, online material). In order to create these pages, they need to bring data together from different BnF datasets, which are in various formats:</p> <ul style="list-style-type: none"> • EAD5 (Encoded Archival Description) for manuscripts and archival fonds, • MARC (Intermarc) for the main catalogue, • Dublin Core6 for the digitised book from Gallica and for the virtual exhibitions. <p>Therefore, the modelling activity has a direct link with aligning and enriching the data that have to be extracted and processed.</p>
Other relevant technical aspects	N/A
Relevant aspects for Data Archiving Management	
Open Data functionalities	N/A
Open Data formats	N/A
Linked Data	N/A
Linked Data formats	<p>The subject records (RAMEAU) from the French national library. They have been converted into the RDF vocabulary SKOS (Simple knowledge organisation system), in the context of the European project TELplus (The European Library, 2007). This repository has been updated and completed with the current records from the BnF database.</p>
Data transformation methods and tools	N/A
Data access / consumption	<p>The Bibliothèque nationale de France (BnF) has opened two OAI repositories aimed at facilitating access to its collections and data.</p> <ul style="list-style-type: none"> • OAI-NUM contains the records for all documents digitised by the BnF and accessible via the Gallica digital library. However, it does not contain the records for the documents from partner libraries which are accessible via the Gallica digital library. • OAI-CAT contains all the bibliographic records of BnF catalogue general, whether documents are digitised or not.

Evaluation criteria	Description of the evaluation criteria
	<p>In accordance with the OAI-PMH protocol, the BnF uses the Simple Dublin Core for the descriptive metadata of its documents. Updated regularly, these two repositories can be freely harvested under the terms of the OAI-PMH protocol, and enable digitised documents and bibliographic data produced by BnF to be referenced in other databases.</p> <p>The BnF exposes some of its data in RDF thanks to data.bnf.fr. This site, which uses semantic web technologies, groups part of the BnF data in Authors, Works and Themes pages and links them on the web. On this model, bindings.bnf.fr offers a new dataset in RDF.</p>
Data reusability	<p>SPARQL endpoint: http://data.bnf.fr/sparql/</p> <p>Tutorial of the SPARQL endpoint: https://github.com/hackathonBnF/hackathon2016/wiki/API-Data</p>

Table 127: Bibliothèque Nationale de France (BNF)

14.2 INVENTORY OF DATA STRUCTURATION APPROACHES AND TOOLS

14.2.1 CIIM

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-001
Methodology name	Collections Information Integration Middleware (CIIM)
URL	http://www.k-int.com/products/CIIM
Basic description	An in-house modular suite of software which sits between institutional data sources (such as collections management systems, library systems, archives and DAMS) and a range of publication end points (such as the institution's online web presence or an in-gallery display).
Owner	Knowledge Integration Ltd
Contact	http://k-int.com/contact-us
Type (sub-type or methodology specificity)	AD tool
Descriptive keywords	Archiving, Linked Data, Enterprise Solution, Processing, LOD sharing
IPR	Proprietary software
Technical description (in case of tool)	
Technologies (language,	SOLR, MySQL, Apache Tomcat, Java

Evaluation criteria	Results
infrastructure, etc.)	https://jiscopenbook.wordpress.com/author/smo30/
Features	<p>Core CIIM is the central module. Its processes and data model are customised based on the needs of each customer.</p> <p>Input Modules are a set of generic modules used to extract data from collections management systems.</p> <p>Processing Modules allow an organisation to customise data to meet its specific needs. Processes are used to enrich or augment the base data to add value to it, they can be fully automated or manual.</p> <p>Sharing Modules offer options for sharing an organisation's data with third party applications.</p> <p>Presentation Modules manage the delivery of content and data to end users.</p>
Data structuration workflow	N/A
Master Data	N/A
Formats	<p>Various inputs: harvesting (OAI-PMH), SPARQL endpoint, XML, RDF and MARC formats.</p> <p>Various outputs: OAI-PMH, XML, RDF</p>
Repositories	N/A
API	APIs for integration with other systems
Exchange protocol	N/A
Competitors	N/A
Use & practices	
Community	No
Use cases	N/A
Related projects / initiatives	N/A
Dissemination channels	Only implemented by company - owner of the product
Evolution	
Methodology maturity	no information on the release
Development status	
Maintenance status	

Table 128: CIIM

14.2.2 AD tool – Preservica

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-002
Methodology name	Preservica
URL	https://preservica.com/digital-archive-software
Basic description	A single integrated application that can be quickly deployed and hosted in the cloud or run on premise. Preservica is developed, maintained and supported by digital preservation experts to ISO 9001 standards, in a secure environment that's compliant with ISO 27001:2013.
Owner	Preservica (International)
Contact	https://preservica.com/contact
Type (sub-type or methodology specificity)	AD tool
Descriptive keywords	Archiving, Preservation, Enterprise Solution, Cloud solution
IPR	Proprietary software
Technical description (in case of tool)	
Technologies (language, infrastructure, etc.)	Enterprise and Cloud editions. No details about architecture
Features	<ul style="list-style-type: none"> • Trusted standards-based (OAIS ISO 14721) repository. • Living archive that creates value for organisations through: <ul style="list-style-type: none"> ○ Multi-source content management ○ Secure sharing of content with multiple audiences: public, researchers and internal users; ○ Dynamic re-arrangement of the archives to meet changing needs; ○ Flexible metadata enrichment over time; ○ Metadata synchronisation mechanism with popular catalogue systems; ○ Automate ingestion from content management and email systems; ○ Easy-to-use for non-expert users; ○ Preservation of all types of content from digitised images, to websites, emails and video; ○ Easy upload of large gigabyte files and multi-terabyte collections; ○ Flexible, durable and cost-saving solution, combining preservation and access in one application.
Data structuration workflow	Fully customisable workflows

Evaluation criteria	Results
	Standard Preservica quality assurance and preservation steps: virus checking, fixity checking, metadata and content integrity, and characterisation to ensure the imported content is properly preserved and can easily be migrated to newer file formats overtime using Preservica Preservation Planning workflows.
Master Data	ISO 14721 Open Archival Information System (OAIS) Reference Model <ul style="list-style-type: none"> ISO 16363 Audit and Certification of a Trustworthy Digital Repository Dublin Core
Formats	Input: all possible formats
Repositories	N/A
API	APIs for integration with other systems
Exchange protocol	N/A
Competitors	http://alm.axiell.com/collections-management-solutions/archive/
Use & practices	
Community	Yes + partnership chain
Use cases	https://preservica.com/resources/case-studies/hsbc
Related projects / initiatives	N/A
Dissemination channels	Via Sales
Evolution	
Methodology maturity	N/A
Development status	Ongoing development, no information on the release
Maintenance status	Active

Table 129: AD Tool – Preservica

14.2.3 GraphDB

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-003
Methodology name	GraphDB (OntoText)
URL	https://ontotext.com/products/graphdb/
Basic description	GraphDB™ Enterprise is an enterprise level triple-store proven to scale in production environments where simultaneous loading, querying, and

Evaluation criteria	Results
	inferencing of graph data statements occur in real time. It features a new data transformation functionality that makes it easier to leverage legacy data, establish interlinked enterprise master data and ultimately build 360-degree data view.
Owner	OntoText
Contact	Europe Polygraphia Office Centre fl.4, 47A Tsarigradsko Shosse Sofia 1124, Bulgaria +359 2 974 61 60
Type (sub-type or methodology specificity)	Data Governance for LOD
Descriptive keywords	Linked Data, Semantic repository, Master Data
IPR	Proprietary software
Technical description (in case of tool)	
Technologies (language, infrastructure, etc.)	Solution implemented in Java and packaged as a Storage and Inference Layer (SAIL) for the RDF4J RDF framework.
Features	Convey the knowledge hidden in Master Data by interlinking catalogue items; clients or prospects; world's shared information of people, organisations and locations; or any other important objects of knowledge or entities in any enterprise world. Get an integrated view to all the autonomous enterprise data sources by managing the master data according to the Linked Data principles. It also takes benefit from: efficient mechanism for data integration and a many-to-one data reference; standardised paradigm for unambiguous reference to data; Continuous data integration and provenance tracking; Access to enormous world's common knowledge available as Open Data; Single entry point to all the knowledge; Instant Master Data sharing with public or private access control.
Data structuration workflow	N/A
Master Data	Used Master data ontologies can be fully custom
Formats	RDF, RDFS, OWL
Repositories	GraphDB is a semantic web repository
API	N/A
Exchange protocol	N/A
Competitors	OpenLink Virtuoso (Open Link Software, 2018), Stardog (Stardog, 2018)

Evaluation criteria	Results
Use & practices	
Community	N/A
Use cases	N/A
Related projects / initiatives	N/A
Dissemination channels	N/A
Evolution	
Methodology maturity	N/A
Development status	GraphDB 8.0
Maintenance status	Active
Open Data value chain for archival data	
Workflow	N/A
Methods	N/A
Formats	N/A

Table 130: GRAPHDB

14.2.4 Silk framework

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-004
Methodology name	Silk - The Linked Data Integration Framework
URL	http://silkframework.org/
Basic description	<p>An Open Source framework for integrating heterogeneous data sources. The primary uses cases of Silk include: The Linked Data Integration Framework is an Open Source framework for integrating heterogeneous data sources. The primary uses cases of Silk include:</p> <ul style="list-style-type: none"> • Generating links between related data items within different Linked Data sources. • Linked Data publishers can use Silk to set RDF links from their data sources to other data sources on the Web. • Applying data transformations to structured data source
Owner	Vulcan Inc. as part of its Project Halo and by the EU FP7 project LOD2 - Creating Knowledge out of Interlinked Data (Grant No. 257943).

Evaluation criteria	Results
Contact	Commercial support is provided by eccenca GmbH.
Type (sub-type or methodology specificity)	ETL for Linked Data
Descriptive keywords	ETL, Linked Data
IPR	Open source
Technical description (in case of tool)	
Technologies (language, infrastructure, etc.)	Stand-alone tool, Java EE (.war), Apache Tomcat
Features	<p>Linked data publishers can use Silk to set RDF links from their data sources to other data sources on the Web.</p> <p>Generating links between related data items within different Linked data sources.</p> <p>Applying data transformations to structured data sources</p>
Data structuration workflow	<p>Fully customisable workflows:</p> <p>Silk enables the user to create and execute lightweight transformation rules. Transformation rules may be used for:</p> <ul style="list-style-type: none"> • Data cleaning, e.g., removing unwanted values. • Mapping between different properties or adding new properties with generated values. • Converting between different data formats.
Master Data	N/A
Formats	No info
Repositories	Any SPARQL-compliant repository (the tool doesn't have own repository)
API	The Workbench provides a REST API.
Exchange protocol	N/A
Competitors	LinkedPipes
Use & practices	
Community	Developer's community (GitHub)
Use cases	https://labs.regesta.com/progettoReload/en
Related projects / initiatives	N/A
Dissemination channels	Open to download and usage
Evolution	

Evaluation criteria	Results
Methodology maturity	
Development status	Last release: Release 2.7.1 from 26 Feb 2016 Last commit: 19 Jun 2017
Maintenance status	Active

Table 131: SILK Framework

14.2.5 LinkedPipes (ETL for Linked Data)

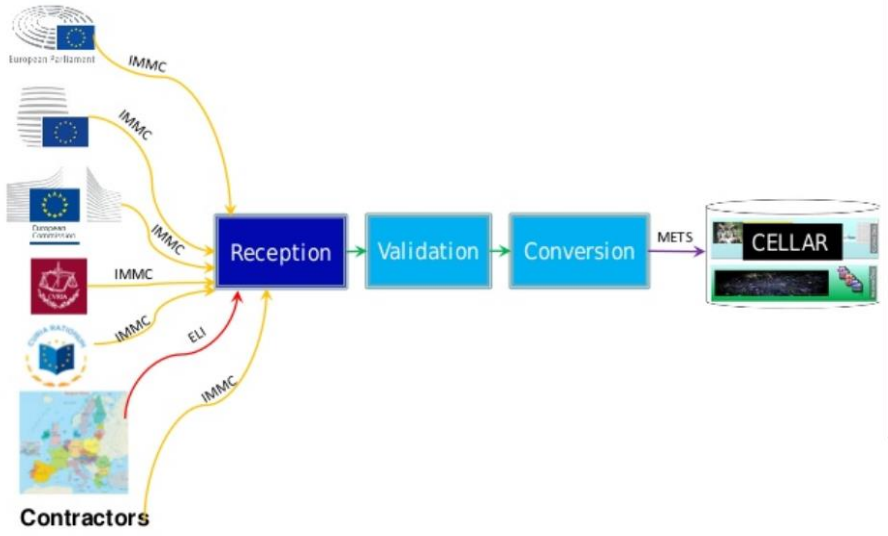
Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-005
Methodology name	LinkedPipes
URL	https://etl.linkedpipes.com/
Basic description	LinkedPipes ETL is an RDF based, lightweight ETL tool.
Owner	LinkedPipes ETL is developed partially by the OpenBudgets.eu project. OpenBudgets.eu has received funding from the European Union's H2020 EU research and innovation programme under grant agreement No 645833.
Contact	http://openbudgets.eu/contact/
Type (sub-type or methodology specificity)	ETL for Linked Data
Descriptive keywords	ETL, Linked Data
IPR	Open source
Technical description (in case of tool)	
Technologies (language, infrastructure, etc.)	Stand-alone tool, Java EE and Node.js
Features	<ul style="list-style-type: none"> • REST API based set of components for easy integration • Library of components to get you started faster • Sharing of configuration among individual pipelines using templates RDF configuration of transformation pipelines
Data structuration workflow	Fully customisable workflows. Pipelines define data transformation processes consisting of interconnected components. The structure of pipelines is completely customisable.
Master Data	N/A

Evaluation criteria	Results
Formats	<ul style="list-style-type: none"> Various input formats: tabular formats (csv, xls), sparql endpoint, archives (zip), structured (xml, json), specific (DCAT-AP) Various output formats: RDF, specific (CKAN, SCP), sparql endpoint
Repositories	Any SPARQL-compliant repository (he tool doesn't have own repository)
API	All functionality covered by REST APIs. APIs for integration with other systems
Exchange protocol	Except for the configuration file, everything is in RDF. This includes the ETL pipelines, component configurations and messages indicating the progress of the pipeline
Competitors	Silk
Use & practices	
Community	Developer's community (GitHub)
Use cases	N/A
Related projects / initiatives	N/A
Dissemination channels	Open to download and usage
Evolution	
Methodology maturity	N/A
Development status	Last commit: 29 Aug 2017 No release versioning
Maintenance status	Active

Table 132: LinkedPipes (ETL for Linked Data)

14.2.6 Publication Office Preservation (Cellar approach)

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-006
Methodology name	Public Office preservation (CELLAR approach)
URL	https://data.europa.eu/euodp/data/dataset/sparql-cellar-of-the-publications-office
Basic description	The CELLAR is the central content and metadata repository of the Publications Office of the European Union. The CELLAR makes available at a single place all the metadata and digital content managed by the Publications Office in a harmonised and standardised

Evaluation criteria	Results
	<p>way in order to:</p> <ul style="list-style-type: none"> to guarantee to the citizen a better access to law and publications of the European Union; to encourage and facilitate reuse of content and metadata by professionals and experts; to preserve content and metadata and access to contents and metadata over time. <p>To reach this goal, CELLAR receives the data in a predefined and standardised formats to integrate and store them in a semantic repository.</p>
Owner	Publication Office (European Commission)
Contact	Marc Wilhelm Küster marc.kuster@publications.europa.eu
Type (sub-type or methodology specificity)	Semantic Data Management
Descriptive keywords	N/A
IPR	Open source
Methodology description (in case of methodology)	
Workflow	 <p>The diagram illustrates the data ingestion process. On the left, several entities are shown: the European Parliament, the European Commission, and a group of Contractors. Each entity is connected to a central 'Reception' box by a yellow arrow labeled 'IMMC'. A red arrow labeled 'ELI' also points to the 'Reception' box. From 'Reception', a green arrow points to a 'Validation' box, which then points to a 'Conversion' box. Finally, a purple arrow labeled 'METS' points from 'Conversion' to a server icon labeled 'CELLAR'.</p>
Methods	<ul style="list-style-type: none"> Reception: reception of the document with the data exchange format called IMMC (Inter-institutional Metadata Maintenance Committee). The format is the minimum set of metadata elements that are to be used in the data exchange related to the legal decision making process between the institutions involved and the Publications Office. Validation: IMMC is XML-based and is validated with a set of business rules. Conversion: the data is converted in a semantic format (RDF) in order to be stored in a semantic repository.

Evaluation criteria	Results
	<ul style="list-style-type: none"> Storing and publication: after conversion, data are stored in a triple store and publish through a SPARQL endpoint.
Formats	List of formats defined by the Metadata Registry: <ul style="list-style-type: none"> IMMC Core metadata exchange protocol European Legislation Identifier (ELI) Public procurement IFC Common Vocabulary OP Core metadata element set EuroVoc thesaurus and alignments (SKOS/XML distributions) Common Data Model (CDM) - Ontology of the CELLAR (content and metadata repository) Application profiles OJEEP (Official Journal Electronic Exchange Protocol) BITS (Book Interchange Tag Suite)
Use & practices	
Community	N/A
Use cases	N/A
Related projects / initiatives	N/A
Dissemination channels	N/A
Evolution	
Methodology maturity	N/A
Development status	N/A
Maintenance status	N/A

Table 133: Publication Office Preservation (Cellar approach)

14.2.7 CORDIS – Data Structuration process

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-007
Methodology name	CORDIS Data structuration
URL	http://cordis.europa.eu/
Basic description	CORDIS is the Publication Office portal to publish information about the European research projects funded by the EC. It publishes all the different scientific documents such as project description (FP7, H2020, etc.), news, and

Evaluation criteria	Results
	scientific reports. Recently, CORDIS started a data curation process to extract the knowledge from the content, validate it and this content as Linked Open Data.
Owner	Publication Office
Contact	FERRAND Karl (OP) <Karl.FERRAND@publications.europa.eu> ; REMAOUN Baya (OP) <Baye.REMAOUN@publications.europa.eu>
Type (sub-type or methodology specificity)	Semantic Data Management
Descriptive keywords	Semantic Web, ontologies, data structuration
IPR	Open source
Methodology description (in case of methodology)	
Scope	Data curation of all the scientific results (projects, papers, news, etc.)
Principles	Following a process of data structuration, the data are cleaned, validated, enriched and linked to referenced data.
Workflow / Guidelines	<p>The diagram illustrates the data curation workflow. It starts with 'Raw data' at the bottom left. The process moves through 'Data Analysis' (Collect, Clean, Enrich) and 'Data Modelling' (Data processing: Knowledge extraction, Link, Publish). The final output is 'Linked data', which is then stored in 'CELLAR (Metadata and Content Storage)'. The 'CORDIS Portal' and 'Open Data Portal (Metadata and Content Storage)' are also shown, with 'Reuse' and 'Preserve' actions indicated.</p> <p>Design:</p> <ul style="list-style-type: none"> • Data analysis: Implies an analysis of the existing schemas of the structured documents and the possible structures to identify in the non-structured documents; • Data modelling (context, structure, content): The main entities identified and their corresponding properties. This is the basic material required to build the core ontology, extended with other domain ontologies and reference data.

Evaluation criteria	Results
	<p>Collect data: Retrieval of the relevant information fields to be enriched or curated (integrated in the validation workflow; described in the Data Management Plan):</p> <ul style="list-style-type: none"> • Data collection: All information from external and internal sources is collected and validated against the expected source format; <p>Normalise</p> <ul style="list-style-type: none"> • Data cleaning: The data is processed to strip text fields from unnecessary or meaningless information. • Data transformation: Transformation of source formats to fit in the generic structure; <p>Enrich / Complete: During this step, the data will be enriched inside the knowledge extraction pipeline through different semantic techniques:</p> <ul style="list-style-type: none"> • Text processing: Involves indexing content with NLP techniques (Tokenisation, Lemmatisation, etc.), and extracting the relevant keywords and acronyms. • Automatic classification: Consists on mapping the relevant keywords extracted from the content with the categories of any relevant taxonomy (e.g. EuroVoc,) • Knowledge extraction: Is an identification of the specific entities (e.g. acronyms, concepts or Named Entities such as Person, Organisation, Location, Event, etc.) and an automatic annotation of the content. <p>Link: This phase is based on the concept of finding connections with validated reference datasets and improve the semantic coherence of the corpus by defining with precision ambiguous elements:</p> <ul style="list-style-type: none"> • Semantic Annotation and Entity Linking steps: Discover and associate the potential reference of the entities available in the LOD repositories • Reasoning and semantic disambiguation: Replacement with the reference entity, merging of records and update missing or incomplete data. Enriched annotations with inferred data <p>Validate:</p> <ul style="list-style-type: none"> • Data validation: Application of business, semantic and technical validation rules; <p>Publish: This step will complete the Data Curation workflow with the persistence of newly generated data:</p> <ul style="list-style-type: none"> • Data publication: Involves publishing the content in an advanced data portal (e.g. as LOD through a semantic repository and exposed through a SPARQL endpoint). <p>Preservation: What and how it must be preserved, including both content and meta-data. In Linked Open Data approach, the preservation operation consists just in the</p>

Evaluation criteria	Results
	<p>extension of the conceptual model with archiving metadata and to link it to the digital objet.</p> <p>Reuse: The final step will apply known semantic techniques to promote the dissemination, exploitation and re-use of the curated data and includes following steps:</p> <ul style="list-style-type: none"> • Data search: Stands for: (i) Term-based search: users can search for content by writing the query with keywords; (ii) Semantic search: possibility to look for specific entities by selecting the type of entity and their relative keywords; (iii) Hybrid search: combines the semantic and the term-based search. • Data consumption: Is about: (i) collecting specific datasets through the SPARQL endpoint: the associated content available in LOD can be used to enrich the description; (ii) triggering federated queries: the big advantage of LOD is the full compliance between the datasets in terms of interoperability.
Expected outcomes	All the different scientific documents will be published as Linked Open Data through a SPARQL endpoint.
(Open Data) Formats	N/A
Examples	List of scientific topics (CORDIS taxonomy) List of entities (Organisation, Person, Event, Location, etc.)
Tutorial	N/A
Best practices	N/A
Use & practices	
Community	N/A
Use cases	N/A
Related projects / initiatives	N/A
Dissemination channels	N/A
Evolution	
Methodology maturity	In progress
Development status	Under integration in the legacy system
Maintenance status	In progress

Table 134: CORDIS - Data structuration process

14.2.8 Project Open Data

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-008
Methodology name	Project Open Data
URL	https://project-open-data.cio.gov/
Basic description	Project Open Data (POD) is a project initiated by the US government, however it is open to public for participation (Office of Management and Budget; Office of Science and Technology Policy, 201?). Its goal is to develop an OGD methodology which would support the public sector bodies in publishing data in line with the Open Data Policy (Executive Office of the President, 2013). This methodology provides basic definition of Open Data, recommendations for implementation of the Open Data Policy, overview of the suitable software tools and recommended readings and resources. A set of case studies is also available at project website.
Owner	Office of Management and Budget (OMB) and Office of Science and Technology Policy (OSTP)
Contact	Unknown
Type (sub-type or methodology specificity)	Semantic Data Management
Descriptive keywords	Open Data
IPR	Open source
Methodology description (in case of methodology)	
Scope	Collection of code, tools, and case studies – to help agencies adopt the Open Data Policy and unlock the potential of government data.
Principles	<p>Create and maintain an Enterprise Data Inventory (Inventory)</p> <p>Create and maintain a Public Data Listing</p> <p>Create a process to engage with customers to help facilitate and prioritise data release</p> <p>Document if data cannot be released</p> <p>Clarify roles and responsibilities for promoting efficient and effective data release</p>
Workflow / Guidelines	https://project-open-data.cio.gov/implementation-guide/
Expected outcomes	<p>Open data will be consistent with the following principles:</p> <ul style="list-style-type: none"> • Public. Consistent with OMB’s Open Government Directive, agencies must adopt a presumption in favour of openness to the extent permitted by law and subject to privacy, confidentiality, security, or

Evaluation criteria	Results
	<p>other valid restrictions.</p> <ul style="list-style-type: none"> • Accessible. Open data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable (i.e., data are reasonably structured to allow automated processing). Open data structures do not discriminate against any person or group of persons and should be made available to the widest range of users for the widest range of purposes, often by providing the data in multiple formats for consumption. To the extent permitted by law, these formats should be non-proprietary, publicly available, and no restrictions should be placed upon their use. • Described. Open data are described fully so that consumers of the data have sufficient information to understand their strengths, weaknesses, analytical limitations, security requirements, as well as how to process them. This involves the use of robust, granular metadata (i.e., fields or elements that describe data), thorough documentation of data elements, data dictionaries, and, if applicable, additional descriptions of the purpose of the collection, the population of interest, the characteristics of the sample, and the method of data collection. • Reusable. Open data are made available under an open license that places no restrictions on their use. • Complete. Open data are published in primary forms (i.e., as collected at the source), with the finest possible level of granularity that is practicable and permitted by law and other requirements. Derived or aggregate Open Data should also be published but must reference the primary data. • Timely. Open data are made available as quickly as necessary to preserve the value of the data. Frequency of release should account for key audiences and downstream needs. • Managed Post-Release. A point of contact must be designated to assist with data use and to respond to complaints about adherence to these Open Data requirements.
(Open Data) Formats	N/A
Examples	https://project-open-data.cio.gov/#6-case-studies
Tutorial	N/A
Best practices	N/A
Use & practices	
Community	N/A
Use cases	N/A
Related projects / initiatives	N/A

Evaluation criteria	Results
Dissemination channels	N/A
Evolution	
Methodology maturity	Mature
Development status	Last update: Aug 23, 2017
Maintenance status	Active

Table 135: Project Open Data

14.2.9 Guidelines on Open Government Data for Citizen Engagement

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-008
Methodology name	Guidelines on Open Government Data for Citizen Engagement
URL	https://www.topquadrant.com/products/topbraid-enterprise-data-governance/
Basic description	The Guidelines on Open Government Data for Citizen Engagement (the Guidelines on OGDCE, or simply the Guidelines) is a practical, easy-to-understand and easy-to-use set of guidelines for everyone, especially policy-makers and technologists. They show what open government data is, why it is important and how it can be of great help for citizen engagement. It will also provide detailed advice on how to assess a countries readiness and how to successfully design, implement, evaluate, and sustain an OGD initiative for citizen engagement in managing development.
Owner	United Nations, Department of Economic and Social Affairs
Contact	Division for Public Administration and Development Management (DPADM), UNDESA. dpadm@un.org 1.212.963.27.64 2 UN Plaza, New York, NY 10017
Type (sub-type or methodology specificity)	Semantic Data Management
Descriptive keywords	Open Data, Open Government
IPR	Open source
Methodology description (in case of methodology)	
Scope	Guidelines, that can be used to understand, design, implement and sustain open government data initiatives, it contains the core principles of openness,

Evaluation criteria	Results
	best practices and case studies, checklists, step-by-step guidelines and practical policy recommendations.
Principles	<p>Develop an action plan Monitor implementation plan Execution of implementation plan Feedback from stakeholders</p> <p>How to open up data:</p> <ul style="list-style-type: none"> • Choose dataset(s) <ul style="list-style-type: none"> ○ Demand driven approach ○ Supply driven approach ○ Ask the people ○ Cost vs. usage and benefit analysis ○ Apply an open license (legal openness) ○ Make data available (technical openness) ○ Online methods ○ Make data discoverable ○ Data portals and catalogues ○ Linked Open Data • Evaluate outcomes and impacts <ul style="list-style-type: none"> ○ Establishing performance indicators
Workflow / Guidelines	http://workspace.unpan.org/sites/Internet/Documents/Guidelines%20on%20OGDCE%20May17%202013.pdf
Expected outcomes	<ul style="list-style-type: none"> • Increased citizens' engagement - measures that empower citizens to engage in government affairs and participate in policy and decision-making processes and other affairs of state • Increased transparency - measures that make governments more transparent through providing unrestricted access to OGD, enabling citizens to understand how the government works, to make informed decisions • Increased accountability and public integrity - measures that make governments more accountable and address corruption and public ethics, access to information, campaign finance reform as well as media and civil society freedom • Improved government efficiency and public service delivery - measures that address the effective delivery of public services through efficient administrative and financial systems, ensuring quality, accessibility, affordability and sustainability • Increased effectiveness managing public resources - measures that address budgets, procurement, natural resources and foreign assistance and reducing transaction costs and enhancing policy coordination between the different government entities • Increased corporate accountability - measures that address corporate

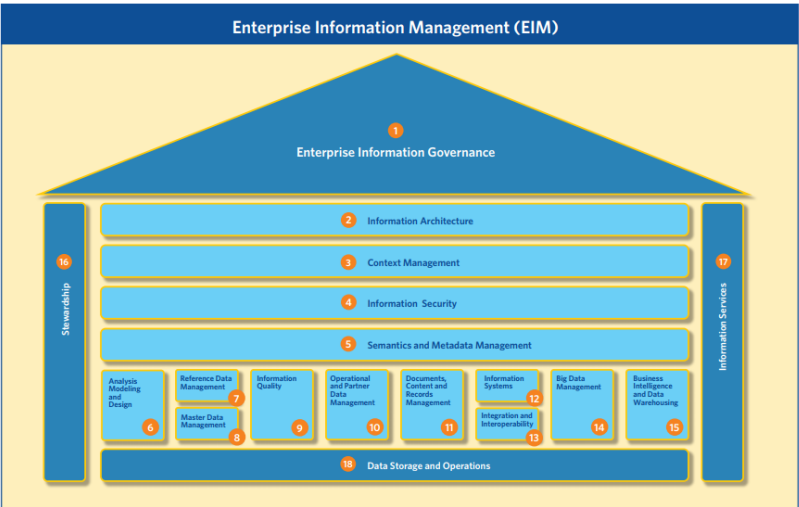
Evaluation criteria	Results
	<p>responsibility on issues such as the corporate social responsibility, anticorruption and consumer protection</p> <ul style="list-style-type: none"> Increased cooperation - measures that strengthen true multi-stakeholder dialogue and cooperation by simplifying the interaction and establish new channels of communication between all sectors of society Economic growth and job creation - measures that foster private sector innovations through unrestricted access to OGD for the creation of new services and products
(Open Data) Formats	N/A
Examples	https://publicadministration.un.org/en/Themes/ICT-for-Development/Open-Government-Data-and-Services/OGD-Projects
Tutorial	N/A
Best practices	N/A
Use & practices	
Community	N/A
Use cases	N/A
Related projects / initiatives	N/A
Dissemination channels	N/A
Evolution	
Methodology maturity	Mature
Development status	Published: 2013
Maintenance status	Active

Table 136: Guidelines on open government data for citizen engagement

14.2.10 TopBraid EDG (Enterprise Data Governance) - Data Governance for LOD

Evaluation criteria	Results
Basic description	
Methodology ID	LOD-MET-010
Methodology name	TopBraid Enterprise Data Governance
URL	https://www.topquadrant.com/products/topbraid-enterprise-data-governance/

Evaluation criteria	Results
Basic description	An agile data governance solution for today's dynamic enterprises. With EDG, you can choose one or combine two or more data governance packages to support a comprehensive but staged approach to data governance
Owner	TopQuadrant Limited
Contact	https://www.topquadrant.com/company/contact/
Type (sub-type or methodology specificity)	Data Governance for LOD
Descriptive keywords	Linked Data, Ontology Management, Taxonomy Management, Master Data
IPR	Proprietary software
Technical description (in case of tool)	
Technologies (language, infrastructure, etc.)	The EDG server uses Java servlets deployed to an Apache Tomcat web server and servlet container
Features	<ul style="list-style-type: none"> • Intuitive Graphical User Interface – with auto-completion, drag and drop, rich text editing, search and filtering, accessibility across all browsers – providing easy to use environment for the business and technical stakeholders. • Flexible data and relationship modelling – handles both complex and simple data models and their relationships across domains, allows modelling, storing and using not only codes, but all relevant associated information. • Auditability – every change is logged and time stamped, change history can be searched; usage records capture where reference data is used. • Control over versions – virtual work-in-progress copies of data assets allow parallel development of versions and enable controlled publishing, review and approval workflow. • Collaboration – enabled through access and accountability based on roles; support for task assignments, statuses and issues. • Shared semantics – provides the ability to define and share meaning of all information elements globally and in the context of specific use. • Repeatability of on-boarding – lets users capture processes and best practices for on-boarding of external reference data. Wide and diverse distribution – support for a variety of interaction patterns (e.g. batch or real-time) and integration approaches. • Data quality – offers intuitive forms for creating data validation rules. Integration – with third party systems and Linked Data clouds via Web Services interfaces and APIs. • Easy extensibility – configurable user interfaces, reports,

Evaluation criteria	Results
	<p>metamodel, import, export, web services and more including deep customisations using TopQuadrant TopBraid platform. Standards – built-in support for W3C (World Wide Web Consortium) standards for data and data models interchange on the web such as RDF and SPARQL.</p> <p>Enterprise-readiness – scalable and robust architecture with LDAP and JMS integration</p>
Data structuration workflow	N/A
Master Data	Used Master data ontologies can be fully custom
Formats	RDF
Repositories	The tool has a semantic web repository inside.
API	No
Exchange protocol	N/A
Competitors	N/A
Methodology description (in case of methodology)	
Scope	Information management and governance in the enterprise context
Principles	<p>Data governance is about creating and using policies for maximising availability, integrity, security, and usability of structured and unstructured information available to an organisation. Information governance brings into the picture the lifecycle and business context of the information. This context includes regulatory, legal, risk, environmental, and operational requirements. TopBraid EDG supports both types of governance in an integrated way — the more tactical and detail oriented data governance together with the more strategic, business policy and context oriented information governance.</p>  <p>The diagram illustrates the Enterprise Information Management (EIM) framework as a house structure. At the top is the roof, labeled 'Enterprise Information Governance' (1). Below the roof are five horizontal layers: 'Information Architecture' (2), 'Context Management' (3), 'Information Security' (4), and 'Semantics and Metadata Management' (5). The base of the house is supported by two vertical pillars: 'Stewardship' (16) on the left and 'Information Services' (17) on the right. Between these pillars are several functional blocks: 'Analysis Modeling and Design' (6), 'Reference Data Management' (7), 'Master Data Management' (8), 'Information Quality' (9), 'Operational and Partner Data Management' (10), 'Documents, Content and Records Management' (11), 'Information Systems' (12), 'Integration and Interoperability' (13), 'Big Data Management' (14), and 'Business Intelligence and Data Warehousing' (15). At the bottom of the house is the foundation, labeled 'Data Storage and Operations' (18).</p>

Evaluation criteria	Results
Workflow / Guidelines	N/A
Expected outcomes	<p>Enhanced Value and Reduced Cost</p> <ul style="list-style-type: none"> Helps organisations use data more intelligently Provides visibility into data and business operations Semantically interoperates through a standards-based data representation instead of a vendor-specific representation with proprietary data formats Enables agile responsiveness to new data requirements End User Capabilities that Preserve Meaning <p>Provides rich, standards-based information structures that enable users to capture anything they need</p> <ul style="list-style-type: none"> Provides flexible capabilities for users to access, view and link datasets Empowers users to align and enrich data assets based on their role or responsibility Readily allows extension of models of the data and metadata Increased Quality and Efficiency <p>Allows organisations to ensure that their data is consistent, reliable and usable for analysis</p> <ul style="list-style-type: none"> Enables data integration across applications for operational needs Provides sophisticated search capabilities that make it easy to create exports of tailored views of governance assets as web services Automates distribution of governance datasets and metadata <p>Provides fully transparent activity trails that enable regulatory compliance and risk mitigation</p>
(Open Data) Formats	RDF
Examples	<p>"Towards Executable Enterprise Models: Building Semantic Enterprise Architecture Solutions with TopBraid Suite"</p> <p>https://www.topquadrant.com/docs/whitepapers/WP-BuildingSemanticEASolutions-withTopBraid.pdf</p>
Tutorial	N/A
Best practices	<ul style="list-style-type: none"> • Enhanced Value and Reduced Cost <ul style="list-style-type: none"> ○ Helps organisations use data more intelligently ○ Provides visibility into data and business operations ○ Semantically interoperates through a standards-based data representation instead of a vendor-specific representation with proprietary data formats ○ Enables agile responsiveness to new data requirements

Evaluation criteria	Results
	<p>End User Capabilities that Preserve Meaning</p> <ul style="list-style-type: none"> • Provides rich, standards-based information structures that enable users to capture anything they need <ul style="list-style-type: none"> ○ Provides flexible capabilities for users to access, view and link datasets ○ Empowers users to align and enrich data assets based on their role or responsibility ○ Readily allows extension of models of the data and metadata ○ Increased Quality and Efficiency • Allows organisations to ensure that their data is consistent, reliable and usable for analysis <ul style="list-style-type: none"> ○ Enables data integration across applications for operational needs ○ Provides sophisticated search capabilities that make it easy to create exports of tailored views of governance assets as web services ○ Automates distribution of governance datasets and metadata <p>Provides fully transparent activity trails that enable regulatory compliance and risk mitigation</p>
Use & practices	
Community	N/A
Use cases	N/A
Related projects / initiatives	N/A
Dissemination channels	N/A
Evolution	
Methodology maturity	N/A
Development status	Last release: TopBraid Suite 5.3.2 Release date: no info
Maintenance status	Active

Table 137: TopBraid EDG (Enterprise Data Governance) - Data governance for LOD

TERMS AND ACRONYMS

GLOSSARY

Term	Description
Access	<p>The ability to make use of material from a fonds, usually subject to rules and conditions</p> <p>Source: ISAD</p> <p>The OAIS functional entity that contains the services and functions which make the archival information holdings and related services visible to Consumers</p> <p>Source: OAIS</p>
Administrative retention period (ARP)	<p>Period of time in which the DG/Service must preserve a file based on its administrative usefulness and any connected statutory and legal obligations. The ARP, established by the Common retention list or, where appropriate, by the Specific retention list, is calculated from the moment when the file is closed. The ARP begins to run from the date when the file is closed</p> <p>Source: IR SEC(2009)1643</p>
Appraisal	<p>Process of determining the disposal of documents based on their archival value. <u>Appraisal at the Commission is carried out by the process of the first and second review</u></p> <p>Source: IR SEC(2009)1643</p> <p>The process of determining the retention period of records</p> <p>Source: ISAD</p>
Archival Information Package (AIP)	<p>An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS</p> <p>Source: OAIS</p>
Archival Storage	<p>The OAIS functional entity that contains the services and functions used for the storage and retrieval of Archival Information Packages (AIP)</p> <p>Source: OAIS</p>
Archival description	<p>The creation of an accurate representation of a unit of description and its component parts, if any, by capturing, analysing, organising and recording information that serves to identify, manage, locate and explain archival materials and the context and records systems which produced it. This term also describes the products of the process</p> <p>Source: ISAD</p>
Archive	<p>An organisation that intends to preserve information for access and use by a Designated Community</p> <p>Source: OAIS</p>
Archives Service of the DG/Service	<p>Members of staff who, depending on the type of organisation (centralised, decentralised, mixed) in the Directorate-General/Service, are responsible for the</p>

Term	Description
	<p>preservation of the current and intermediate records of their unit, directorate or DG/service in line with a general authorisation from the appropriate tier of the administration</p> <p>Source: IR SEC(2009)1643</p>
Cataloguing	<p>The process of providing access to materials by creating formal descriptions to represent the materials and then organising those descriptions through headings that will connect user queries with relevant materials. - 2. The process of providing such access, plus additional work to prepare the materials for use, such as labelling, marking, and maintenance of authority files.</p> <p>Source: Society of American Archivist</p>
Closed file	<p>File that cannot be used for filing documents or creating new sub files. Note that a file is closed by the lead department when the case has been dealt with, i.e. when no more documents need to be added to the file</p> <p>Source: Common Commission–Level retention list for EC files– SEC(2007)970, p. 5 as amended by the 1st revision of the Common Retention List for European Commission files (SEC(2012) 713 – Ares (2012 1501883 – 17/12/2012).</p>
Consumer	<p>The role played by those persons or client systems, who interact with OAI services to find preserved information of interest and to access that information in detail. This can include other OAIss, as well as internal OAI persons or systems</p> <p>Source: OAI</p>
Current records	<p>Open files created by the departments that are regularly and frequently used in the framework of their activities for the conduct of their current business and which are generally kept nearby</p> <p>Source: IR SEC(2009)1643</p>
Date of closure	<p>Date when the status of a file is changed to “closed file”. See also “Closed file”</p> <p>Source: NOMCOM User Manual</p>
Dissemination Information Package (DIP)	<p>The Information Package, derived from one or more AIPs, and sent by Archives to the Consumer in response to a request to the OAI</p> <p>Source: OAI</p>
Elimination	<p>Regulated procedure or any other intentional action leading to the physical destruction of archives or any other operation resulting in a total or partial loss of information</p> <p>Source: IR SEC(2009)1643</p>
Historical archives	<p>All files which, in accordance with the appraisal rules in force, are transferred to the Commission’s historical archives and selected for permanent preservation (=definitive archives)</p> <p>Source: IR SEC(2009)1643</p>
Hybrid files	<p>Files containing both paper and electronic documents</p> <p>Source: IR SEC(2009)1643</p>
Information Package	<p>A logical container composed of optional Content Information and optional associated Preservation Description Information which is needed to aid in the</p>

Term	Description
	<p>preservation of the Content Information. The Information Package has associated Packaging Information used to delimit and identify the Content Information and Preservation Description Information</p> <p>Source: OAIS</p>
Ingestion	<p>The OAIS functional entity that contains the services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage and ensures that Archival Information Packages and their supporting Descriptive Information become established within the OAIS</p> <p>Source: OAIS</p>
Intermediate records	<p>Closed files which do not reach the 30 years after the most recent document in the file yet. These files may (still) be appraised or reviewed</p> <p>Source: HAS</p>
Life cycle (of a document)	<p>All the stages or periods in the life of a document from the time it is received or formally drawn up until it is transferred to the Commission's historical archives and/or opened to the public or until it is destroyed according to the rules</p> <p>Source: IR SEC(2009)1643</p>
Long Term Preservation	<p>The act of maintaining information, Independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term</p> <p>Source: OAIS</p>
Metadata	<p>Data about other data</p> <p>Source: OAIS</p> <p>Data describing the context, contents and structure of documents and their management over time</p> <p>Source: IR SEC(2009)1643</p>
Open Archival Information System (OAIS)	<p>An archive, consisting of an organisation, which may be part of a larger organisation, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS archive to be distinguished from other uses of the term 'archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums and it does not imply that access to the archive is unrestricted. It offers a reference model addressing a full range of archival information preservation functions including ingest, archival storage, data management, access, and dissemination. It also addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives. It identifies both internal and external interfaces to the archive functions, and it identifies a number of high-level services at these interfaces</p> <p>Source: OAIS</p>
Preservation	<p>Reference Information - Identifiers are stored for each object identifying it</p>

Term	Description
Description Information (PDI)	<p>globally (e.g. macrepo PID) and locally (e.g. URI).</p> <p>Provenance Information - Provenance metadata is maintained for each object that provides a history of preservation events in the object's lifetime, beginning at ingest into the digital repositories and referencing any preservation activities taken on the object (e.g., replacement due to corruption, format migration, etc.).</p> <p>Context Information - As appropriate, information on how a CDO relates to other CDOs or to other conceptual entities. Examples of these relationships can include: a newer version of an object that supersedes an older one.</p> <p>Fixity Information - Fixity information is generated at the time of ingest in order to later determine whether or not the item remains in the same state as when it was ingested. This information can be used to determine integrity of an object being copied within the system (as in the case of a change in storage location), or for periodic integrity checks.</p> <p>Source: https://digitalarchive.mcmaster.ca/node/56</p>
Permanent preservation	<p>One of two possible actions to be taken by the Historical archives service on files after their transfer to the Historical archives, the other action being second review.</p> <p>Source: IR SEC(2009)1643</p>
Preservation of a document	<p>Each Directorate-General or equivalent department shall ensure the physical protection and the short- and medium-term accessibility of the documents for which it is responsible and must be in a position to produce or reconstruct the files to which they belong</p> <p>Source: Decision 2002/47/EC, ECSC, Euratom, HAN Vision Document – Adonis D(2006)2276</p>
Producer	<p>The role played by those persons or client systems, which provide the information to be preserved. This can include other OAISS or internal OAISS persons or systems</p> <p>Source: OAISS</p>
Record	<p>Recorded information in any form or medium, created or received and maintained, by an organisation or person in the transaction of business or the conduct of affairs</p> <p>Source: ISAD(G)</p>
Sampling	<p>Method of appraisal whereby, on the basis of objective criteria, a representative portion of files is chosen for preservation from a larger body of files that will not be preserved in its entirety</p> <p>Source: IR SEC(2009)1643</p>
Second review	<p>Evaluation of certain files transferred to the Historical archives service with a view to identifying the files that have enough value to be preserved as historical archives</p> <p>Source: IR SEC(2009)1643</p>
Selection	<p>Method of appraisal whereby, on the basis of the selector's assessment, a certain number of files are chosen for preservation from a larger body of files</p>

Term	Description
	<p>that will not be preserved in its entirety</p> <p>Source: IR SEC(2009)1643</p>
Series	<p>Documents arranged in accordance with a filing system or maintained as a unit because they result from the same accumulation or filing process, or the same activity, have a particular form; or because of some other relationship arising out of their creation, receipt, or use. A series is also known as a records series</p> <p>Source: ISAD(G)</p>
Submission Information Package (SIP)	<p>An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs and/or the associated Descriptive Information.</p> <p>Source: OAIS</p>
Transfer to the historical archives	<p>Change of custody and responsibility for the Commission's files and documents from the Directorates General and equivalent departments to the Commission's Historical archives service</p> <p>Source: IR SEC(2009)1643</p>
Automatic annotation	<p>Automatic annotation refers to the process in which a set of algorithms process the digital content and is able to automatically recognise content or structural elements. The algorithms add to these elements the corresponding metadata to enrich the content and facilitate the indexing and retrieval process.</p> <p>Cf. Metadata enrichment</p>
Concept	<ol style="list-style-type: none"> 1.) a general idea or notion that corresponds to some class of entities and consists of the characteristic or essential features of the class 2.) the meaning of a predicate
Data curation	<p>Data curation is a broad term used to indicate processes and activities related to the organisation and integration of data collected from various sources, annotation of the data, and publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation. Data curation includes "all the processes needed for principled and controlled data creation, maintenance, and management, together with the capacity to add value to data" (Wikipedia).</p>
Data transformation	<p>Data transformation where the data is transformed for storing in the proper format or structure for the purposes of querying and analysis.</p> <p>In computing, data transformation is the process of converting data from one format or structure into another format or structure. It is a fundamental aspect of most data integration and data management tasks such as data wrangling, data warehousing, data integration and application integration (Wikipedia).</p>
Entity Linking	<p>In natural language processing, entity linking, named entity linking (NEL), named entity disambiguation (NED), named entity recognition and disambiguation (NERD) or named entity normalisation (NEN) is the task of determining the identity of entities mentioned in text. For example, given the sentence "Paris is</p>

the capital of France", the idea is to determine that "Paris" refers to the city of Paris and not to Paris Hilton or any other entity that could be referred as "Paris". NED is different from **Named Entity Recognition (NER)** in that NER identifies the occurrence or mention of a named entity in text but it does not identify which specific entity it is.

Extract, transform, load (ETL) In computing, extract, transform, load (ETL) refers to a process in database usage and especially in data warehousing. The ETL process became a popular concept in the 1970s. Data extraction is where data is extracted from homogeneous or heterogeneous data sources; **data transformation** where the data is transformed for storing in the proper format or structure for the purposes of querying and analysis; data loading where the data is loaded into the final target database, more specifically, an operational data store, data mart, or data warehouse.

Interoperability Interoperability is a characteristic of a product or system, whose interfaces are completely understood, to work with other products or systems, at present or future, in either implementation or access, without any restrictions. Beyond the ability of two or more computer systems to exchange information, semantic interoperability is the ability to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of both systems. To achieve semantic interoperability, both sides must refer to a common information exchange reference model, such as **Ontologies** that's; are promoted by **Semantic Web** project.

Inferred data Inferred data is information that a business has not collected either passively or actively from the user, but rather has inferred using data about a sample population.
asserted facts / axioms

Linked Data In computing, Linked Data is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. It builds upon standard Web technologies such as HTTP, **RDF** and **URIs**, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried. Tim Berners-Lee, director of the World Wide Web Consortium (W3C), coined the term in a 2006 design note about the **Semantic Web** project.

Linked Open Data (LOD) Linked Open Data is **Linked Data** that is open content. Tim Berners-Lee gives the clearest definition of Linked Open Data in differentiation with Linked Data. Linked Open Data (LOD) is Linked Data which is released under an open licence, which does not impede its reuse for free.

LOD Repository (or Semantic repository) A semantic repository is an engine similar to a database management systems (DBMS) that permits the storage, querying and handling of structured data. In addition, a semantic repository uses **ontologies** as **semantic schemata** to

automatically reason about the queried data. Semantic repositories make use of generic and flexible physical data models, such as graphs. This permits them to quickly read and implement new **metadata schemata** or **ontologies**. As a result, semantic repositories provide better incorporation of assorted data as well as more analytical power. However, these kinds of repositories are still in the early stages of their development.

Master data (or Common data) Master data represents the business objects which are agreed on and shared across the enterprise. It can cover relatively static **reference data**, transactional, unstructured, analytical and hierarchical metadata. It is the primary focus of the Information Technology (IT) discipline of Master Data Management (MDM).

Metadata Metadata is "data [information] that provides information about other data". Three distinct types of metadata exist: descriptive metadata, structural metadata, and administrative metadata.

- Descriptive metadata describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
- Structural metadata is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.
- Administrative metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

Metadata enrichment Similar to Automatic annotation, this is the process that consists to automatically add new metadata to digital content.
Cf. Automatic annotation

Metadata schemata A metadata standard is a requirement which is intended to establish a common understanding of the meaning or semantics of the data, to ensure correct and proper use and interpretation of the data by its owners and users. To achieve this common understanding, a number of characteristics, or attributes of the data have to be defined, also known as **metadata**.

Named Entity Recognition (NER) Named-entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify **named entities** in text into pre-defined categories such as the names of persons, organisations, locations, expressions of times, quantities, monetary values, percentages, etc.

Natural Language Processing (NLP) Natural Language Processing (NLP) is a series of techniques to "understand" **relevant keywords** based on the semantic value inside a text.

Ontology Ontology is the hierarchical structuring of knowledge about things by subcategorising them according to their essential (or at least relevant and/or cognitive) qualities.
This is an explicit formal specification of how to represent the objects, **concepts**

and **other entities** that are assumed to exist in some area of interest and the relationships that hold among them.

Open Data

Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. The goals of the Open Data movement are similar to those of other "open" movements such as open source, open hardware, open content, open government and open access. Paradoxically, the growth of the Open Data movement is paralleled by a rise in intellectual property rights. The philosophy behind Open Data has been long established (for example in the Mertonian tradition of science), but the term "Open Data" itself is recent, gaining popularity with the rise of the Internet and World Wide Web and, especially, with the launch of open-data government initiatives such as Data.gov and Data.gov.uk.

Reasoning techniques

A semantic reasoner, reasoning engine, rules engine, or simply a reasoner, is a piece of software able to infer logical consequences from a set of **asserted facts** or **axioms**. The notion of a semantic reasoner generalises that of an inference engine, by providing a richer set of mechanisms to work with. The inference rules are commonly specified by means of an ontology language, and often a description logic language. Many reasoners use first-order predicate logic to perform reasoning; inference commonly proceeds by forward chaining and backward chaining. There are also examples of probabilistic reasoners, including Pei Wang's non-axiomatic reasoning system, and probabilistic logic networks. Rule Interchange Format (RIF) is a W3C recommendation that allow to formalise the triple patterns used to generate the new facts in the knowledge base.

Reference Data

Reference data are data that define the set of permissible values to be used by other data fields. Reference data gain in value when they are widely re-used and widely referenced. Typically, they do not change overly much in terms of definition, apart from occasional revisions. Reference data are often defined by standards organisations, such as country codes as defined in ISO 3166-1. Examples of reference data include: Units of measurement, Country codes (NUTS), Corporate codes, Fixed conversion rates (e.g., weight, temperature, and length), Calendar structure and constraints, etc.

Relevant keywords

The relevant keywords are terms or set of terms that were identified with **NLP techniques**. Contrary to the relevant terms, not only the frequency of the terms is taken into account in the weighting calculation. It can consider different layers of semantics (e.g. if the keyword is present in the title, relation to acronyms, etc.) to enrich the keywords weighting.

Resource Description Framework (RDF)

Resource Description Framework (RDF), is a family of World Wide Web Consortium (W3C) specifications to allow the interchange of data among machines.
RDFS Ontology Resource Description Framework Schema (RDFS)

Semantic annotation	<p>The semantic annotation is a technique that permits to associate metadata to the content in order to express its meaning.</p> <p>More complex than the automatic annotation, it usually requires human validation to ensure the accurate meaning of the content.</p>
Semantic Web	<p>The Semantic Web is an extension of the World Wide Web through standards by the World Wide Web Consortium (W3C). The standards promote common data formats and exchange protocols on the Web, most fundamentally the Resource Description Framework (RDF).</p> <p>According to the W3C, "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries". The term was coined by Tim Berners-Lee for a web of data that can be processed by machines—that is, one in which much of the meaning is machine-readable. While its critics have questioned its feasibility, proponents argue that applications in industry, biology and human sciences research have already proven the validity of the original concept.</p>
SKOS	<p>SKOS (Simple Knowledge Organisation System) is a format based on RDF that permits to describe the basic structure and the conceptual models such as lists, taxonomies, classification models, thesaurus, etc.</p>
SPARQL	<p>SPARQL (pronounced "sparkle", a recursive acronym for SPARQL Protocol and RDF Query Language) is an RDF query language, that is, a semantic query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (RDF) format. It was made a standard by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium, and is recognised as one of the key technologies of the semantic web. On 15 January 2008, SPARQL 1.0 became an official W3C Recommendation, and SPARQL 1.1 in March, 2013.</p>
Taxonomy	<p>Taxonomy is the organisation of a particular set of information for a particular purpose. The information is represented and organised into categories, sub-categories and final leaves. This a hierarchical tree of categories.</p>
Thesaurus	<p>A thesaurus is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms), in contrast to a dictionary, which provides definitions for words, and generally lists them in alphabetical order. The main purpose of such reference works is to help the user "to find the word, or words, by which [an] idea may be most fitly and aptly expressed" – to quote Peter Mark Roget, architect of the best known thesaurus in the English language.</p>
Uniform Identifier (URI)	<p>Resource In information technology, a Uniform Resource Identifier (URI) is a string of characters used to identify a resource.</p>
Uniform Locator (URL)	<p>Resource A Uniform Resource Locator (URL), colloquially termed a web address, is a reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it. A URL is a specific type of Uniform Resource</p>

	<p>Identifier (URI), although many people use the two terms interchangeably. URLs occur most commonly to reference web pages (http), but are also used for file transfer (ftp), email (mailto), database access (JDBC), and many other applications.</p>
Extensible Markup Language (XML)	<p>In computing, Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. The W3C's XML 1.0 Specification and several other related specifications—all of them free open standards—define XML.</p>
XSD (or XML schema)	<p>XSD (XML Schema Definition), a recommendation of the World Wide Web Consortium (W3C), specifies how to formally describe the elements in an Extensible Markup Language (XML) document. It can be used by programmers to verify each piece of item content in a document. They can check if it adheres to the description of the element it is placed in.</p>
XSLT	<p>eXtensible Stylesheet Language Transformations, language for transforming XML documents into other formats (e.g. HTML, PDF, etc.).</p>
Web of Data	<p>The Data Web transforms the Web from a distributed file system into a distributed database system.</p> <p>Rather than webpages, pieces of data (RDF triples) and records formed from them (sets, trees, graphs or objects). Some of these could even come from databases.</p> <p>Tim Berners-Lee has suggested that Data Web may be a more appropriate name for the Semantic Web. Tim O'Reilly, who coined the term Web 2.0 has mentioned that the long-term vision of the Semantic Web as a web of data, where sophisticated applications manipulate the data web.</p>

Table 138: Glossary

ACRONYMS AND ABBREVIATIONS

Term	Description
AIP	Archival Information Package Source: OAIS
APEnet	Archives Portal Europe Network
ARP	Administrative Retention Period Source: IR SEC(2009)1643
COMREF	Platform related to the human resource management of the European Commission, to which Hermes is integrated.
CRL	Common Commission level retention list Source: IR SEC(2009)1643
CSV	Comma-Separated Values

Term	Description
DUA	Duration of administrative usefulness (FR: Durée d'utilité administrative)
DIP	Dissemination Information Package Source: OAIS
EAC	Encoded Archival Context (format for exchanging archival data)
EAD	Encoded Archival Description
EC	European Commission
EDG	Enterprise Data Governance
EEAS	European External Action Service
EGAD	Expert Group on Archival Descriptions
EIF	European Interoperability Framework
ENA	Estonian National Archives
EP	European Parliament
EU	European Union
FRBR	Functional Requirements for Bibliographic Records
GDPR	General Data Protection Regulation See http://eur-lex.europa.eu/eli/reg/2016/679/oj for full text.
HAEU	Historical Archives of the European Union
HAN	Hermes/Ares/NomCom
HAS	Historical Archives Service
HPS	Hermes Preservation Services
HRS	Hermes Repository Services
HTML	Hypertext Markup Language
ICA	International Council on Archives
IP	Information Package (Source: OAIS)
ISAAR	International Standard Archival Authority Records (archival standard)
ISAAR(CPF)	International Standard Archival Authority Record For Corporate Bodies, Persons and Families
ISAD	International Standard Archival Description (archival standard)
ISAD(G)	General International Standard Archival Description
JSON	JavaScript Object Notation
LOD	Linked Open Data
LOGD	Linked Open Government Data
NARA	National Archives and Records Administration (USA)
NLP	Natural Language Programming

Term	Description
NOMCOM	IT application managing the European Commission filing plan.
OAIS	Open Archival Information System Source: OAIS
OCLC	Online Computer Library Centre
PHP	Hypertext Preprocessor
RDF	Resource Description Framework
RiC	Records in Contexts
RiC-CM	Records in Contexts Conceptual Model
SAM	Sampling Source: IR SEC(2009)1643
SEL	Selection Source: IR SEC(2009)1643
SGML	Standard Generalised Markup Language
SPARQL	SPARQL Protocol and RDF Query Language)
SIP	Submission Information Package Source: OAIS
SRL	Specific DG/Service level retention list Source: IR SEC(2009)1643
TNA	The National Archives (UK)
TSV	Tab-Separated Values
URI	Uniform Resource Identifier
XML	eXtensible Markup Language
XSD	XML Schema Definition

Table 139: Acronyms and abbreviations

REFERENCES

DESK RESEARCH SOURCES

The following sources have been consulted during the desk research:

Type	Name	Link
Organisations and related publications	International Council on Archives ICA	https://www.ica.org/en
	DLM Forum	http://www.dlmforum.eu/
	Society of American Archivists SAA Publications/Magazines	https://www2.archivists.org/ https://www2.archivists.org/publications/book-publishing/trends-in-archives-practice#.VxZ6yHErJhE
	Archives & Records Association ARA Best practices guidelines ARC Magazine online	http://www.archives.org.uk/ http://www.archives.org.uk/publications/best-practice-guidelines.html http://www.archives.org.uk/publications/arc-magazine.html
	Association of Canadian Archivists ACA	http://archivists.ca/ http://archivists.ca/content/archivaria
	Australian Society of Archivists	https://www.archivists.org.au/
	OCLC Research	http://www.oclc.org/research/publications/reports.html
	Library of Congress Digital Formats	https://www.loc.gov/preservation/digital/ https://www.loc.gov/preservation/digital/formats/ https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml
	W3C	https://www.w3.org
Standards and specifications	APEX project	http://apex-project.eu/images/docs/EAG_2012_guide_2013-05-27.pdf
		http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First
		http://core.vraweb.org/vracore_presentations.html
	Jenni Ley website	http://jennriley.com/metadatamap/seeingstandards_glossary_pamp_hlet.pdf
		http://dublincore.org/specifications/
		http://eac.staatsbibliothek-berlin.de/

		http://eurovoc.europa.eu/drupal/
		http://id.loc.gov/authorities/subjects.html
		http://id3.org/
		http://pbcore.org/schema/
		http://publications.europa.eu/mdr/authority/index.html
		http://references.modernisation.gouv.fr/archivagenumerique
		http://rahvusarhiiv.ra.ee/en/principles-standardsguidelines/
		http://www.3dicons-project.eu/
	UK Archives standards sheets	http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_article_ISAAR.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_article_ISDF.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_articleEAC-CPF.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_for_archives_ead.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/Standards for Archives ISAD.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/Standards for Archives ISDIAH.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_for_archives_iso_15489.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_for_archives_iso_23081.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/standards_for_archives_paimas.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/unesco.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/web_standards_for_archives_dacs.pdf http://www.archives.org.uk/images/documents/DSG_docs/DSG_Standards/web_standards_for_archives_dublinc.pdf
	French archives	http://www.archivesdefrance.culture.gouv.fr/static/2822 http://www.archivesdefrance.culture.gouv.fr/static/2823 http://www.archivesdefrance.culture.gouv.fr/static/7109
		http://www.archivschule.de/uploads/Forschung/Retrokonversion/Vortrag_Kolloquium/13_Desantes.pdf
		http://www.cdncouncilarchives.ca/archdesrules.html http://www.cdncouncilarchives.ca/RAD/RAD_GeneralIntro_July2008.pdf
		http://www.digitisationguidelines.gov/guidelines/TIFF_Metadata_Final.pdf
		http://www.dnb.de/EN/Netzpublikationen/Ablieferung/ablieferung_n

		ode.html
		http://www.getty.edu/research/publications/electronic_publications/cdwa/introduction.html http://www.getty.edu/research/tools/vocabularies/tgn/index.html
		http://www.iptc.org/std/photometadata/documentation/userguide/index.htm#!Documents/introduction.htm
		http://www.moreq.info/ http://moreq2.eu/attachments/article/189/MoReq2_typeset_version.pdf
		http://www.nationaalarchief.nl/sites/default/files/docs/Toepassingsprofiel_metagegevens_rijksoverheid.pdf
		http://www.nationalarchives.gov.uk/documents/information-management/egms-metadata-standard.pdf
		http://www.oclc.org/en/viaf.html
		https://mpeg.chiariglione.org/standards/mpeg-1 https://mpeg.chiariglione.org/standards/mpeg-7
		https://iptc.org/standards/photo-metadata/iptc-standard/
		https://jpeg.org/jpeg
		http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html
		http://www.ukat.org.uk/
		https://public.ccsds.org/pubs/650x0m2.pdf
		https://www.aoc.cat/content/download/13501/32409/file/Cond_espec%C3%ADfiques_iARXIU_amb_annexos.pdf
		https://www.nlm.nih.gov/class/nlmclassintro.html
		https://www.iso.org/obp/ui/#iso:std:iso:30300:ed-1:v1:en https://www.iso.org/standard/53732.html https://www.iso.org/standard/54534.html
		https://www.ica.org/en/egad-ric-conceptual-model https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf
		https://www.iasa-web.org/sites/default/files/downloads/publications/TC03_English.pdf https://www.iasa-web.org/tc03/ethics-principles-preservation-strategy https://www.iasa-web.org/tc04/audio-preservation https://www.iasa-web.org/tc05/handling-storage-audio-video-carriers
	German archives	federal https://www.bundesarchiv.de/archivgut_online/standards_werkzeuge/02255/index.html

		https://www.bsigroup.com/en-GB/bs-10008-electronic-information-management/
		https://www.biodiversitylibrary.org/
		https://www.aoc.cat/content/download/6657/24722/file/estructuraPitMets.pdf
		https://www2.archivists.org/groups/standards-committee/australian-government-locator-service-metadata-standard-agls https://www2.archivists.org/groups/standards-committee/library-of-congress-classification-lcc https://www2.archivists.org/groups/standards-committee/library-of-congress-subject-headings-lcsh https://www2.archivists.org/groups/standards-committee/metadata-encoding-and-transmission-standard-mets https://www2.archivists.org/groups/standards-committee/mpeg-multimedia-content-description-interface-mpeg-7 https://www2.archivists.org/groups/standards-committee/open-archival-information-system-oais https://www2.archivists.org/groups/standards-committee/open-digital-rights-language-odrl https://www2.archivists.org/groups/standards-committee/preservation-metadata-implementation-strategies-premis https://www2.archivists.org/groups/standards-committee/public-broadcasting-core-metadata-dictionary-pb-core https://www2.archivists.org/groups/standards-committee/visual-resources-association-core-categories-vra-core https://www2.archivists.org/standards/handling-and-storage-of-audio-and-video-carriers https://www2.archivists.org/standards/safeguarding-the-audio-heritage-ethics-principles-and-preservation-strategy
		https://www.w3.org/2003/01/21-RDF-RDB-access/#Introduction https://www.w3.org/2004/02/skos/ https://www.w3.org/standards/techs/owl#w3c_all https://www.w3.org/TR/2004/REC-rdf-primer-20040210/ https://www.w3.org/TR/2012/REC-owl2-overview-20121211/ https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/ https://www.w3.org/TR/webarch/#identification
		https://www.loc.gov/ead/eadschema.html https://www.loc.gov/ead/index.html https://www.loc.gov/marc/marc.html https://www.loc.gov/preservation/digital/formats/fdd/fdd000005.shtml https://www.loc.gov/preservation/digital/formats/fdd/fdd000062.shtml https://www.loc.gov/preservation/digital/formats/fdd/fdd000105.shtml#identification https://www.loc.gov/preservation/digital/formats/fdd/fdd000115.shtml

		<p>ml https://www.loc.gov/preservation/digital/formats/fdd/fdd000155.shtml ml</p>
IT Tools & Solutions		<p>https://preservica.com/resources/knowledge-centre/archivesspace-connector-overview http://archivesspace.org/developers/integrations-with-archivesspace/ https://preservica.com/resources/press-releases/preservica-simplifies-digital-preservation-for-calm-catalogue-users https://libraries.mit.edu/digital-archives/integrating-tools/ http://www.eark-project.com/resources/project-deliverables/93-d34-1 http://www.eark-project.com/resources/project-deliverables/92-d54/file https://www.archivematica.org/en/docs/archivematica-1.6/user-manual/access/access/ https://www.accesstomemory.org/es/docs/2.3/user-manual/import-export/oai-pmh/ http://amlad.es/amlad/ https://www.archivematica.org/ https://preservica.com/ http://www.oclc.org/en/contentdm.html http://www.omeka.net/ http://neatline.org/ http://www8.hp.com/us/en/software-solutions/file-analysis-dark-data-cleanup/index.html https://www.ibm.com/support/knowledgecenter/SS8NLW_10.0.0/com.ibm.discovery.es.nav.doc/iypofnv_prodovert_cont.htm https://www2.deloitte.com/content/dam/Deloitte/us/Documents/about-deloitte/us-ibm-watson-client.pdf http://everismoriarty.com/ https://www.sas.com/en_us/software/visual-text-analytics.html https://www.sas.com/en_us/solutions/analytics.html#text-analytics https://www.blancco.com/products/file-eraser/ https://www.opentext.com/file_source/OpenText/en_US/PDF/opentext-po-redact-it-enterprise-en.pdf https://joinup.ec.europa.eu/collection/common-assessment-method-standards-and-specifications-camss https://joinup.ec.europa.eu/document/interoperability-quick-assessment-toolkit</p>

Furthermore, several reports and market studies have been consulted during the research:

Name	Content
Archival Management Software_Spiro_2009 Lisa Spiro	Comparative market study. Contains interesting tables as appendix.

https://www.clir.org/pubs/reports/spiro/spiro_Jan13.pdf	<ul style="list-style-type: none"> • How to Select Archival Management Software • Criteria for Choosing Archival Software • Types of Software
Archival Management Software (wiki) http://archivalsoftware.pbworks.com/w/page/13600254/FrontPage	The above study in a wiki format. Although this wiki was created with the aim of regularly maintaining the published information, our investigation has shown that it has never been really updated since 2009.
Collection Management_pros_cons_2009	Comparative market study of Collection Management software.
Council of Nova Scotia Archives. Archives Management Software Review (2009): http://www.councilofnsarchives.ca/resources/SoftwareReview.htm	Reviews of archives software
http://collectionstrust.org.uk/software/	Web-site about technical data about collection scope products
Gartner and Ovum Market studies	The research focus was in the search of comparative market studies in the scope of archives management systems and related tools ⁵⁷ .
https://goascribe.com/wp-content/uploads/2015/05/Vendor_Landscape_Big_Dat.pdf	

Table 140: Main sources

The table hereunder lists the references and deliverables of this project:

Name	Link
E-ARK Project	http://www.eark-project.com/
D3.4 Records export, transfer and ingest recommendations and SIP Creation Tools Deliverable This document is a document which describes a software deliverable which delivers a number of E-ARK tools.	http://www.eark-project.com/resources/project-deliverables/93-d34-1
D5.1 E-ARK GAP report between requirements for access and current access solutions Deliverable This report is a GAP analysis between current access services and user requirements for Access.	http://www.eark-project.com/resources/project-deliverables/3-d51-e-ark-gap-report
Deliverable The purpose of this document is to describe the tools for accessing	http://www.eark-project.com/resources/project-deliverables/3-d51-e-ark-gap-report

⁵⁷ No current studies available.

archival material and to provide links to their documentation and the code. The tools are based on specifications that have been partially or fully created in the E-ARK project.	deliverables/92-d54/file
---	--

Table 141: E-ARK deliverables for IT tools inventory

BIBLIOGRAPHY

- (n.d.). In J. W. James Lowry, *Integrity in Government through Records Management: Essays in Honour of Anne Thurston* (p. 139). Routledge.
- (2017). In A. J. Gilliland, *Research in the Archival Multiverse* (p. Chapter 23).
- Archives, U. N. (2018). *Best practice guide for appraising and selecting records for the National Archives*. Retrieved from <http://www.nationalarchives.gov.uk/documents/information-management/best-practice-guide-appraising-and-selecting.pdf>
- Berners-Lee, T. (2009, February). The next web. TED. Retrieved October 15, 2017, from https://www.ted.com/talks/tim_berners_lee_on_the_next_web
- Berners-Lee, T., Bizer, C., & Heath, T. (2009). , Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May 17). The Semantic Web. *Scientific American*, 284(5), 28-37. Retrieved October 15, 2017, from <https://www.scientificamerican.com/magazine/sa/2001/05-01/#article-the-semantic-web>
- Coeckelbergs, M. (2018). *GitHub*. Retrieved from <https://github.com/MathiasCoeckelbergs/Concepts-within-Topics>
- Cook, T. (1994, November). Electronic records, paper minds: the revolution in information management and archives in the post/ custodial and post/ modernist era. *Archives and Manuscripts*, 22(2), 300-328. Retrieved from <http://search.informit.com.au/documentSummary;dn=950302854;res=IELAPA>
- Digital Preservation Coalition. (2017). *Digital Preservation Handbook*. Retrieved November 06, 2017, from Digital Preservation Handbook: <http://dpconline.org/handbook>
- Ding, L., Peristeras, V., & Hausenblas, M. (2012). Linked Open Government Data [Guest editors' introduction]. *IEEE Intelligent Systems*, 27(3), 11-15. Retrieved November 2, 2017, from <http://ieeexplore.ieee.org/stamp/stamp.jsp?reload=true&tp=&arnumber=6237454>
- DLM Forum. (2017). *The E-Ark Work Programme*. Retrieved November 06, 2017, from DLM Forum Foundation: <http://www.dlmforum.eu/index.php/resources/e-ark-programme>
- DREG. (2017). Retrieved October 01, 2017, from <https://www.ica.org/en/our-professional-programme/expert-group-digital-records-dreg>
- Dunia Llanes-Padrón, J.-A. P.-S. (2017). *emeraldinsight*. Retrieved from <http://www.emeraldinsight.com/doi/abs/10.1108/PROG-03-2017-0021>
- E-ARK Project. (2017). Retrieved October 10, 2017, from <http://www.eark-project.com/>
- EGAD. (2017). Retrieved September 5, 2017, from <https://www.ica.org/en/about-egad>
- EG-MDPR. (2017). Retrieved September 15, 2017, from <https://www.ica.org/en/expert-group-on-managing-digital-and-physical-records-eg-mdpr>

- Elastic*. (2018). Retrieved from <https://www.elastic.co/products/elasticsearch>
- European Commission. (2018, February 07). *CEF Programme 2018 - New Building Blocks*. Retrieved from CEF Programme 2018: <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2018/02/07/CEF+Telecom+Work+Programme+2018%3A+New+Building+Blocks+and+Grant+Funding>
- European Interoperability Framework. (2017).
- European Law Blog*. (2015). Retrieved from <https://europeanlawblog.eu/tag/c-61513-clientearth-and-pan-europe-v-efsa/>
- Gartner. (2017, October 30). *Magic Quadrant for Master Data Management Solutions*. Gartner.
- Google Inside Search*. (2018). Retrieved from The Knowledge Graph: <https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html>
- Government, O. o. (2014). *Automated Electronic Records Manahement Plan*.
- Harvey, R. (2008). *"Appraisal and Selection"*. Edinburgh: Digital Curation Centre.
- HAS. (2017). *Overview of Metadata and Digital Objects Managed by the Historical Archives Service of the European Commission*. Brussels: European Commission.
- Hengchen, S., Coeckelbergs, M., Hooland, S. V., Verborgh, R., & Steiner, T. (2016). Exploring the Archives with Probabilistic Models: Topic Modelling for the Valorisation of Digitised Archives of the European Commission. *IEEE International Conference on Big Data*, (pp. 3245-3250).
- Highland, & Wood. (2011). *The Joy of Data - Cookbook for Publishing Linked Government Data on the Web*. W3C.
- Horsman, P. (2002, Fall). The Last Dance of the Phoenix or The De-Discovery of the Archival Fonds. *Archivaria, The Journal of the Association of Canadian Archivists*, 54. Retrieved from <https://archivaria.ca/archivar/index.php/archivaria/article/view/12853>
- ICA Expert Groups. (2017). Retrieved September 7, 2017, from <https://www.ica.org/en/our-professional-programme/expert-groups>
- ICSTI. (2009). *ICSTI Annual Conference - "Managing Data for Science"*. Ottawa. Retrieved from <http://www.icsti.org/IMG/doc/ProceedingsOTTAWAV-F.doc>
- ICSTI. (2017). Retrieved September 11, 2017, from <http://www.icsti.org/>
- InfoCuria - Case-law of the Court of Justice*. (2001). Retrieved from <http://curia.europa.eu/juris/liste.jsf?language=en&num=C-615/13%20P>
- ISO. (2012). *www.iso.org*. Retrieved from <https://www.iso.org/standard/57284.html>
- JISC. (2017). *Assessment of UKDA and TNA Compliance with OAIS and METS Standards*.
- Library of Congress*. (2018). Retrieved from <http://www.loc.gov/bibframe/implementation/>
- Lucene*. (2018). Retrieved from <https://lucene.apache.org/>
- OIB. (2017). Retrieved August 20, 2017, from http://ec.europa.eu/historical_archives/index_en.htm
- Open Link Software*. (2018). Retrieved from <https://virtuoso.openlinksw.com/>
- Pederson, A. (1987). *Keeping Archives*.
- Popovici, B. F. (2016). Records in Contexts: Towards a New Level in Archival Description? *Technical and Field Related Problems of Traditional and Electronic Archiving* (pp. 13-31). Radenci: Regional

- Archives Maribor. Retrieved October 15, 2017, from http://www.pokarh-mb.si/uploaded/datoteke/Radenci/radenci_2016/013-031_popovici_2016.pdf
- RMEG. (2017). Retrieved September 11, 2017, from <https://www.ica.org/en/expert-group-records-management-rmeg>
- Society of American Archivists. (2017). *Glossary Terms*. Retrieved November 06, 2017, from <https://www2.archivists.org/glossary/terms>
- Society of American Archivists*. (2018). Retrieved from Glossary: <https://www2.archivists.org/glossary>
- Spencer. (2016). *Youtube*. Retrieved from Binary Trees? Automatically Identifying the links between born digital records.: <https://www.youtube.com/watch?v=Ked9GRmKIRw>
- Stardog*. (2018). Retrieved from <https://www.stardog.com/>
- TechTarger Network. (2017). *Taxonomy*. Retrieved November 06, 2017, from TechTarger Network: <http://searchcontentmanagement.techtarget.com/definition/taxonomy>
- The Digital Preservation Coalition. (2017). Retrieved October 20, 2017, from <http://www.dpconline.org/>
- The European Library*. (2007). Retrieved from <http://www.theeuropeanlibrary.org/archive/projects/telplus/index.php>
- The National Archives. (2016). *The application of technology-assisted review to born-digital records transfer, inquiries and beyond*. Research Report, The National Archives. Retrieved October 30, 2017, from <http://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf>
- Thibodeau, K. (2003, October 23). Personal communication.
- UK Data Archive. (2017). Retrieved September 14, 2017, from <http://www.data-archive.ac.uk/>
- UK National Archives. (2017). Retrieved September 12, 2017, from <http://www.nationalarchives.gov.uk/>
- Van Hooland, S. V. (2013). Evaluating the success of vocabulary reconciliation for cultural heritage collections. *Journal of the Association for Information, Science and Technology*, 464-479.
- van Hooland, S., & Verborgh, R. (2014). *Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata*. Facet Publishing.
- Vellino, A., & Alberts, I. (2016). Assisting the appraisal of e-mail records with automatic classification. *Records Management Journal*, 26(3), 293-313.
- W3C. (2013). Retrieved from <https://www.w3.org/TR/prov-overview/>
- W3C. (2015). Retrieved from <http://www.loc.gov/bibframe/implementation/>
- W3C. (2017). *Semantic Web*. Retrieved November 06, 2017, from <https://www.w3.org/standards/semanticweb/>
- W3C. (2018). Retrieved from <https://www.w3.org/community/archetypes/>
- Wikipedia*. (2018). Retrieved from https://en.wikipedia.org/wiki/Linked_data#Linked_open_data
- Wired*. (2008). Retrieved from <https://www.wired.com/2008/06/pb-theory/>

INTERESTING READINGS

- Millar, Laura A.* Archives. Principles and practices in records management and archives. 2017. 2nd Ed. Facet publishing
- Description. Innovative practices for archives and special collections. 2014 *Edited by Kate Theimer.* Rowman & Littlefield.
- Zeng, Marcia Lei.* Metadata. London: Facet Publishing, [2016]. Second edition
- Foulonneau, Muriel.* Metadata for digital resources: implementation, systems design and interoperability Oxford: Chandos, 2008
- Cole, Timothy W.* Using the Open Archives Initiative protocol for metadata harvesting. Westport (Conn.): Libraries Unlimited, cop. 2007
- Digital preservation metadata for practitioners: implementing PREMIS. Cham: Springer, [2016]
- Thomas, David.* The silence of the archive. London: Facet Publishing, 2017. 1st Ed.
- Engaging with records and archives: histories and theories. London: Facet Publishing, [2016]
- Crockett, Margaret.* The no-nonsense guide to archives and recordkeeping. London: Facet Publishing, 2016
- Appraisal and acquisition: innovative practices for archives and special collections. Lanham: Rowman & Littlefield, 2015.
- Bastian, Jeannette A.* Archives in libraries: what Librarians and Archivists need to know to work together? Chicago: Society of American Archivists, cop. 2015
- Padfield, Timothy.* Copyright for archivists and records managers. London: Facet, 2015. 5th Ed.
- Van Hooland, Seth.* Linked data for libraries, archives and museums: how to clean, link and publish your metadata. London: Facet Publishing, 2014