3.3 DEVELOPMENT OF AN OPEN DATA SERVICE, SUPPORT AND TRAINING PACKAGE IN THE AREA OF LINKED OPEN DATA, DATA VISUALISATION AND PERSISTENT IDENTIFICATION (2016.18)

3.3.1 IDENTIFICATION OF THE ACTION

Type of Activity	Common Services and Reusable Generic Tools
Service in charge	Publications Office of the European Union
Associated Services	JRC, ESTAT[35]

3.3.2 EXECUTIVE SUMMARY

The action aims at providing data providers (institutions, agencies and other bodies) of the Open Data Portal (ODP) and other stakeholders of the Open Data Portal (Pan-European Data Portal including affiliated MS portals) with an open data service package in order to enable them to further open up their data and increase data interoperability in view of better data reusability and data visualisation.

It has been triggered by the needs expressed by EC DGs, institutions and agencies during workshops and individual meetings with ODP teams and additionally brought to evidence by the interinstitutional survey on data visualisation needs (carried out between May and July 2015), namely

- o for **a structural approach for sharing knowledge about open data management** (Guidelines, consultancy, assistance, training, awareness raising, etc.), reusable data structures, data visualisation and the development of shareable tools.

- o for significantly **mutualising and rationalising efforts** towards the identification, development and sharing of modern **data processing and visualisation tools** (i.e. a corporate visualisation solutions pool).

- o to **raise awareness** about the dependency of easily reusable visualisation tools on interoperable data structures, and to put in place structural measures in order to achieve this alignment.

- o To support further opening of data namely by assistance in transforming data into Linked open data (LOD) and persistent identification.

---

[35] Services involved in the elaboration of this proposal. The exact nature of the collaboration with these services is still subject to formal agreement.

Through its different action strands (implementing guidelines, consultancy and hands-on assistance, workshops and training sessions, development of toolsets, etc.), the project will deliver an

- o interinstitutional **knowledge base** on open data publishing, open data treatment and processing, data visualisation
- o interinstitutional **pool of reusable** data processing and data visualisation **tools**
- o governance structure, **implementation guidelines** and development of a toolset for unique and **persistent identification based** upon the sub-domain data.europa.eu and accompanied, where suitable, by the Digital Object Identifier (DOI).

Moreover, it will contribute to building up an intra-institutional community of practice in the domains of open data and data visualisation (a network of open data and data visualization stewards) and an increased data publishing maturity of participating institutions, agencies and bodies.

The project will be implemented with existing key partners of the Open Data Portal (data providers, stakeholders) as well as with new partners to be identified during the follow-up phase of the interinstitutional survey on visualisation needs (September - December 2015)

3.3.3 OBJECTIVES

The project aims at providing data providers (institutions, agencies and other bodies) of the Open Data Portal and other stakeholders of the Open Data Portal (Pan-European Data Portal including affiliated MS portals) with an open data service package in order to enable them to further open up their data and increase data interoperability in view a better data reusability and data visualisation.

3.3.4 SCOPE

The project covers

- o open data publishing, open data treatment and processing, data modelling in view of linked open data, data visualisation (setup of an interinstitutional knowledge base)
- o the set-up of an interinstitutional pool of reusable data processing and data visualisation tools
- o the implementation of a toolset for unique and persistent identification based upon the sub-domain data.europa.eu and accompanied, where suitable, by the Digital Object Identifier
- o community building in all abovementioned areas

It entails activities in the areas of:

- o consultancy and conceptual assistance,

- o development of training material, delivery of training sessions/workshops,

- o development of toolsets (including catalogues of tools)

- o development of interoperability guidelines and documentation

3.3.5 PROBLEM STATEMENT

a) EU institutions, agencies and other bodies need a structural approach for sharing knowledge about open data visualisation and a pool of reusable data visualisation tools

Raw data can be hard to interpret even by professionals with advanced computational skills. One of the best ways to convey a message is to use visualisations to quickly draw attention to key findings. Furthermore, by visually representing data it is possible to uncover patterns and observations that are not apparent from looking at the raw data alone or even at a few summary statistics.

Given the growing amount of data produced, gathered, processed and published by the EU institutions, there is a need to develop effective tools for visualising data. This has to be done in a cost-effective and mutualised way, building where possible on existing experiences, interoperable standards and tools used in the institutions.

In view of these aspects, in its February 2015 meeting, the Commission's Information Systems Project Management Board (ISPMB) asked the Publications Office to conduct a survey in view of producing a global overview of visualisation needs and visualisation tools of the Commission services. Another goal was to assess the possibilities of sharing data visualisation tools and of developing new tools which would enable data consumers to better explore EU data. Considering its interinstitutional role, the Publications Office carried out the survey between April and July 2015 across EU institutions, agencies and other bodies. The survey was developed and implemented by the EU Open Data Portal. Participants were selected based on job function and/or unit: webmasters, information and/or communication specialists, IT support specialists, data specialists, publication officers, and heads of sector/unit. In total, 99 respondents from 26 DGs, 25 agencies and 9 other EU institutions and other bodies participated in the survey.
The report will be made available as of September.

The survey results indicate that:
- more than 50% of the participants in the survey use visualisation tools, equally across Commission DGs, Institutions and agencies.

- approximately 50% of the respondents which use visualisation tools indicated that they have tools which are not shared yet, but which can be shared. Approximately one-fifth indicated that they already share all of their visualisation tools.

- almost 70 % of the participants in this survey indicated that they publish data for which they do not have appropriate visualisation tools yet. 37% have already carried out research on new data visualisation tools that may be suitable for their data. 41% have special categories of data that need to be visualised. Approximately two-thirds would like to visualise linked data or geospatial data. There is also an interest in visualising big data, data illustrating processes or flows and visualising networks.

- Some respondents are noticeable for their initiative to develop in-house tools[36] and/or to share the associated source code. Several DGs and agencies, including DG CNECT, DG DIGIT, DG ESTAT, EEA and DG JRC indicated that they have developed in-house tools or tools that are a combination of open-source tools and developed in-house tools ([37]) and that they publish or intend to publish the source code of the data visualisation tools that they use ([38]).

- The most important requirements for data visualisation tools in the EU institutions are tools that are platform-independent and accessible via phone, tablet and personal computer, tools that are reusable, tools that are easy to be handled internally, and tools enabling data export in various formats.

In conclusion, the key message conveyed by the survey responses is that the participating DGs, institutions and agencies have common visualisation needs, that currently there is little sharing of knowledge and tools, and that several participants have started working independently on the development of data visualisation solutions.

b) EU institutions, agencies and other bodies need a structural and reuse-oriented approach for sharing knowledge about open data management and publishing, data models, data identification, and the development of shareable tools

The survey and repeatedly expressed requests by EC DGs, institutions and agencies during workshops and individual meetings with the ODP brought to evidence that **there is a need:**

- for a structural approach for sharing knowledge about open data management (Guidelines, consultancy, assistance, training, awareness raising, etc.), reusable data structures, data visualisation and the development of shareable tools.

- for significantly mutualising and rationalising efforts towards the identification, development and sharing of modern data processing and visualisation tools (i.e. a corporate visualisation solutions pool).

- for **raising awareness** about the dependency of easily reusable visualisation tools on interoperable data structures, and to put in place structural measures in order to achieve this alignment.

- for further opening of data namely by assistance in transforming data into Linked open data (LOD) and the set-up of a framework service for unique and persistent identification via URIs.

Preparatory work to be done via ISA 1.1 in view of the ISA[2] project
The results of the visualisation survey and the actions to be derived from it will be ready to be discussed with ISPMB and the stakeholders' community not earlier than late September 2015. This

---

[36] Approximately two-thirds of the respondents using data visualisation tools indicated that they use commercial tools, while approximately 45% use open source visualisation tools. Approximately two-thirds indicated that they have tools developed in-house or tools that are a combination of open source tools and tools developed in-house.

([37]) DG AGRI, DG CLIMA, DG CNECT, DG DIGIT, EACEA, EASME, ECDC, EEA, EIGE, EFSA, DG EMPL, EU-OSHA, Eurofound, EP, DG COMM, ESTAT, FRONTEX, JRC, OHIM.
([38]) DG CNECT, DG DIGIT, EAC, ECB, ESTAT, EU-OSHA, EEA, DG JRC, REA, SG.

process will require a structured approach over 2-3 months in order to properly define the follow-up strategy and measures to be taken.

Against this background, it is proposed to use an action within ISA 1.1. as a platform for bringing together the stakeholders and define the requirements, priorities and workplan for a subsequent implementation within an ISA[2] action as of May 2016.

- Definition – in close cooperation with the key stakeholders – of the priorities and actions of the envisaged corporate approach for sharing knowledge about data visualization.
  Based on the defined requirements (i.e. in terms of consultancy, assistance, training, development of reusable tools, etc.), this activity is expected to deliver a detailed work plan for a follow-up activity to be implemented within ISA[2] as of May 2016. Special attention in this preparatory work shall be given to an alignment with the needs of the Pan-European Portal (via DG CONNECT), in order to make the envisaged outcomes fully reusable by the Member States.

- Definition of the requirements for the elaboration of:
  o guidelines for data providers (Open Data Portal, Pan European) on data management with reuse and visualization in mind. Special attention shall be given to the visualization and data exploration opportunities provided by linked data. This is a recurrent request by data providers and the final deliverable is expected to cover – in the light of the "Reflection Paper on Information Management" of July 2015 – hands-on guidelines on formats, data structure definitions for different content types, data and metadata management, data quality, dataset identification, etc.
  o related toolsets (methodology, develop libraries for data conversion and for visualisation, data-mining and comparative analysis that enables presentation of data with different granularity and modality (spatial, temporal, organizational, etc.)
  o provision of related consultancy and training.

The implementation of those recommendations by data providers are expected to significantly facilitate visualizations via common libraries, aggregation and sharing of data through common tools, interoperability between platforms, and reusability by data consumers. This activity within ISA 1.1 is expected to deliver detailed requirements for a follow-up activity to be implemented within ISA[2] as of May 2016.

- Definition of the requirements for the creation of a catalogue of shareable visualisation tools (and hosted by the EU Open Data Portal), as a first step towards a corporate visualisation solutions pool. The basic input concerning tools in use and their owners is being provided by the abovementioned survey.

3.3.6 EXPECTED BENEFICIARIES AND ANTICIPATED BENEFITS

Beneficiaries	Anticipated benefits
Commission, all other institutions, agencies and bodies interested in participating in the projects	o Single point of access to a catalogue of shareable visualisation tools (hosted by the EU Open Data Portal), as a first step towards a corporate pool of visualisation solutions. o Interoperability guidelines in the domains of data visualisation, data treatment for reuse and visualisation, open data publishing, in particular as Linked Open Data o Possibility to obtain training, share knowledge, obtain conceptual support o Increased data publishing maturity of participating institutions, agencies and bodies o Building up an interinstitutional community of open data and data visualisation experts, evangelists and quasi-contact points
Users of the Open Data Portal and the Pan-European Portal (diverse reuse communities)	o Better data quality, including better documentation o More and better visualisations o More open datasets published on the Open Data Portal
Member States Portals and other portals	o Reuse of the developed guidelines and toolsets

3.3.7 RELATED EU ACTIONS / POLICIES

Action / Policy	Description of relation, inputs / outputs
Open data policy	The project will directly foster open data portals via improved open data publishing and portal services (preview, visualisation, etc.), enhanced reusability, etc.
Commission corporate Information management	Fostering interoperability of metadata and data through guidelines, promotion of standards and common authority list; internal reuse of data; unique and persistent identification;
Better Regulation	Enhanced transparency, improved data communication and interpretation,
Open Science	Fostering the publishing of more research related datasets

3.3.8 REUSE OF SOLUTIONS DEVELOPED BY ISA, ISA[2] OR OTHER EU / NATIONAL INITIATIVES

- o DCAT AP, DCAT AP STAT, DCAT AP GEO (2015)

- o Methodologies of modelling linked open data and outcomes developed in projects TED as LOD and EU BUDGET as LOD (2014-2015)

- o Identification of high-value datasets (2013)

- o Unique and persistent identification (2014)

3.3.9 EXPECTED RE-USABLE OUTPUTS (solutions and instruments)

Output name	**Knowledge base and the reuse-oriented data management and data treatment (in particular towards LOD) and data visualization**
Description	Knowledge base consisting of: o Implementation guidelines o Training material, footage of training sessions, o briefings on conceptual assistance, o development of interoperability guidelines and documentation
Reference	
Target release date / Status	As of project inception, continuously enriched and improved

Output name	**Common toolset in the domains data visualization and the reuse-oriented data management and data treatment (in particular LOD)**
Description	Toolset (presented most probably as a catalogue of applications, libraries, etc.) consisting of reusable tools for data visualization and data treatment and conversion (methodologies, develop common libraries for data conversion and for visualisation, enabling the presentation of data with different granularity and modality (spatial, temporal, organizational, etc.).
Reference	
Target release date / Status	To be defined

Output name	**G**overnance structure, implementation guidelines and development of a toolset for unique and persistent identification based upon the sub-domain data.europa.eu
Description	• Setup of URI governance model • Development of a toolset for decentralised URI generation and management • Development of a central URI Register
Reference	
Target release date / Status	To be defined

Output name	**Data models, ontologies and content of pilots treated within the project**
Description	• Transformation of some data into LOD (CORDIS, TED, ESTAT, BUDGET, other interested DGs) and publishing them on ODP • Visualizations produced and listed on ODP
Reference	
Target release date / Status	To be defined

3.3.10   ORGANISATIONAL APPROACH

3.3.10.1 Expected stakeholders and their representatives

Stakeholders	Representatives
Publications Office	Open Data Portal (Norbert Hohn, Agnieszka Zajac) Common repository (Cellar) : Peter Schmitz Metadata registry : Willem van Gemert
Eurostat	Marc Vanderperren (Renovation of the ESTAT dissemination chain)
CONNECT	Daniele Rizzi (Pan-European Data Portal)
JRC	Andrea Perego, Anders Friis Christensen
DIGIT	TBD
Others	Other key stakeholders identified further to the presentation of the visualisation survey and of the ISA[2] project layout within September-December 2015

3.3.10.2 Communication plan

The results of the visualisation survey and the actions to be derived from it will be ready to be discussed with ISPMB and the stakeholders' community not earlier than late September 2015. This process will require a structured approach over 2-3 months in order to properly collect input, and to define the follow-up strategy and the measures to be taken.

For this, it is proposed to use an action within ISA 1.1. as a communication platform for bringing together the stakeholders and define the requirements, priorities and work plan for a subsequent implementation within an ISA^2 action as of May 2016.

During the lifetime of the ISA^2 project, communication of the project will be ensured through the following channels:
- o Open Data Portal network and the Publications Office's publishing officers network towards institutional data providers
- o Pan-European Data portal towards affiliated MS portals
- o Open Data Portal social media and other channels towards reuser communities

3.3.10.3 Governance approach

The Publications Office will ensure the chairmanship and the project management.
All key stakeholders shall be members of the project steering committee.

3.3.11 TECHNICAL APPROACH

The nature of the works to be executed will require different approaches:

- o the knowledge transfer and capacity building tasks (consultancy/conceptual support, assistance, training, and the production of guidelines for data providers) will be done by domain experts on-demand. It will not contain technical implementation tasks.
- o the tasks related to elaborating the toolset will require technical analysis of data and tools, their categorization, and the development of required adaptations

The action will deliver:

1) Implementation guidelines for successful open data management and publishing with reuse and visualization in mind. In the light of the "Reflection Paper on Information Management" of July 2015, it shall provide a blueprint for organisation, governance and processes and hands-on guidelines on formats, Identification and definition of domain specific data structure definitions, data and metadata management, data quality, etc.
A specific project strand will be dedicated to developing the governance structure, the implementation guidelines and a toolset for unique and persistent identification based upon the sub-domain data.europa.eu

2) Actions to further opening-up data, in particular towards Linked Open Data, in terms of:

o providing training and tailored project oriented workshops
o conceptual counselling
o support in the identification and creation of ontologies and data models
o support in data conversion, in particular towards RDF (via pilots in cooperation with associated project partners)

3) Provide support in the visualisation of data:
   o Create a catalogue of existing shareable visualisation tools (hosted by the EU Open Data Portal), as a first step towards a corporate pool of visualisation solutions. The crucial challenge in this activity will consist in designing the catalogue around the institutional reuse needs, in order to ensure easy shareability and fitness for purpose. It is expected to cater for different categories of data (e.g. numerical/statistical, geospatial, network/graph, process/flow).
   o Develop/adapt a toolset of reusable tools for data visualization (methodology, develop common libraries, etc.)
   o Provide training and tailored project oriented workshops
   o Conceptual counselling and support

The actions will lead to the following results:
o Interinstitutional knowledge base on open data publishing, open data treatment and processing, data visualisation (via training materials, guidelines and project briefings)
o Interinstitutional pool of reusable data processing and data visualisation tools (toolset)
o Intra-institutional community of practice in the domains of open data and data visualisation (a network of open data and data visualization stewards).
o Increased data publishing maturity of participating institutions, agencies and bodies
o Governance structure, implementation guidelines and development of a toolset for unique and persistent identification based upon the sub-domain data.europa.eu

3.3.12 COSTS AND MILESTONES

3.3.12.1 Breakdown of anticipated costs and related milestones

Phase: Inception Execution Operational	Description of milestones reached or to be reached	Anticipated Allocations (KEUR)	Budget line ISA^2/ others (specify)	Start date (QX/YYYY)	End date (QX/YYYY)
	Project Management	140	ISA^2	Q2/2016	Q4/2017
	Development of documents	30	ISA^2	Q2/2016	Q4/2017

	Consultancy, training, assistance	100	ISA^2	Q2/2016	Q4/2017
	Development of toolsets	150	ISA^2	Q2/2016	Q4/2017
	Total	420			

3.3.12.2 Breakdown of ISA^2 funding per budget year

A more precise breakdown can be established only on base of the elaborated requirements and workplans; (around January 2016).

Budget Year	Phase	Anticipated allocations (in KEUR)	Executed budget (in KEUR)
2016		300	
2017		120	
2018			
2019			
2020			