**11th e-Concertation meeting European e-Infrastructure projects –**

**Discussion and contributions received on Digital4Science**

## Results of Social Sciences and Humanities discussion – Steve Brewer

I was invited to act as the rapporteur for the breakout session on Social Sciences and Humanities during the 11th e-Concertation meeting European e-Infrastructure projects.

**Introduction – Wim Jansen (chair), Jarkko Siren (co-chair)**

* The session started with an introduction from Wim to the packed room.
* The objective of this grouping is to identify and capture the synergies and opportunities between the projects.
* No presentations were made: two sentence introductions from each of the 15 partcipants representing the projects.
* Goal: what can we contribute to the communities of social scientists and humanities?
* Traditionally this community is less interested in e-Infrastructure – but this is the long tail for e-Infrastructure.
* We need to focus on this group and support them.
* What can we do to bring e-Infrastructure providers closer to these communities in the very near future?
* Which elements should be taken into consideration?

**Raporteur’s summary of the session as presented to whole group**

**Topics:**

1. There exists a **large community** of researchers across humanities and Social Science
2. The challenge is to communicate with them, we need **brokers**: libraries, e-Humanities centres, etc.
3. Many sections of this community maintain a healthy **scepticism about data**
4. History is now starting to be based on **digital records** eg. end of 20th Century
5. Long-term vision needed but **short-term gains and success stories** – focus on the medium tail – eg. Collections not individuals
6. Important factors for succesful solutions were identified: **co-design and co-evolution**
7. Examples of **success stories** needed: Greek social scientists wanted to interrogate newspaper archives and ILSP (Institute for Language and Speech Processing) enabled them to look at which groups have participated in social confrontations (strikes, occupations, etc.) over time by searchig across collections.
8. EUROPEANA has been successful, but their focus is on **metadata**. Researchers need to look at full-data. But many **legal, social and regulatory** issues to be overcome to simplify this in practice.
9. Researchers will not be interested in Cloud per se, that is just a box with an API. **They will want solutions**.

**Actions**

1. **Share knowledge**: THOR has a role to raise awareness, engagement will identify needs as a side project and share this, others can capture information about needs
2. **Training projects could work together across this scattered landscape** – join forces and deliver coherent message. (Humanities presence is currently light within RDA). However, we need to go to Digital Humanities conferences and talk to them directly.
3. Lost resource in publishing – many are small scale, lacking DOIs. They are beholden to shareholders. **We need to analyse the market**. We need to speak to **publishers** and understand their needs. Roles for SMEs etc.
4. Where is the future? Sharing? How do we enable this? Institutions are interested in this. **We need to go to Digital Humanities conferences.**

**Further conclusion:**

We should all look at Europeana and what they are doing, how they are doing it, how can we contribute?

#### [**EDISON project support for Social Science and Humanities**](https://ec.europa.eu/futurium/en/comment/5786#comment-5786) **- Steve Brewer**

We in the EDISON project are certainly keen to support the Social Science and Humanities communities. We would be interested to see how they see the role of data scientists and what skills they will need from data scientists working in this field. Therefore, we would be interested to hear from those involved in either training data specialists for this domain or employers looking to recruit specialists.

Top of Form

#### [**Humanities/Social Science conferences to target?**](https://ec.europa.eu/futurium/en/comment/5787#comment-5787) **– Steve Brewer**

Do people have suggestions for humanities and/or social sciences conferences that we should target? We could propose a workshop to bring together domain specialists together with e-Infrastructure specialists. What skill-sets are needed to accelerate progress in the adoption of e-Infrastructure across the Social Science and Humanities community.

# Notes from the Transversal Meeting during the e-Concertation event – Lucia Florio

### Attendees/projects represented

* PRACE, Jules Wolfrat involved in the operation of PRACE and on the AAI aspects. Interested in looking at AARC solutions that PRACE can adopt.
* AARC/GEANT, Licia Florio AARC coordinator, but also representing the Trust and Identity work in GEANT. Interested in meeting potential user communities interested in using federated access.
* EGI-ENGAGE, Ludek Matyska representing EGI-ENGAGE. Interested in learning what do is missing in the big picture as  opposed to what we have
* SESAMenet, Karen Padmore  project coordinator of SESAMenet which aims to promote the take up to High Performance facilities and eInfra to SMEs.
* MUG,  Modesto Orozco and Anna Montras
* VRE4EIC, Keith Jeffery interested in understand what other project are doing
* POP project looking for synergies in the area of high performances
* VI-SEEM, Lazarou Constantinos
* EUDAT - Per Oster lots of work done on the AAI, interested in synergies with other projects
* THOR -  Adam Farquhar
* EoCoE, Edouerd Audit project looking for synergies in HP area.

### Introduction

The projects gathered for the trasversal group covered different perspectives (some focused more on SMEs, some others more on end-users and data driven science and some others offers facilities and services outside the project remit). All representatives had a common understanding on the value of AAI. EUDAT noted that they already support federated access and they are also working to join eduGAIN with some of their services.  Whilst there was consensus that existing AAIs work within their community, all agreed that seamless sharing of resources across e-Infrastructures is not really possible to date. We should enable SSO across e-infrastructure, implement better authorisation mechanism whilst preserving security and in line with the data protection. Regarding security, all current e-Infrastructure could probably make an effort and highlight the security aspects they cover.  This may help both users understanding what exists and how increasing the level of security adds complexity, costs and challenges easy of use.  The main goal for those delivering AAI should be to remove fragmentation and improve the user experience.

### Requirements

Participants were asked to share some their requirements.  VRE4EIC presented some concrete examples on researchers that need access to services that are scattered among different e-infrastructures. Although each e-Infrastruture offers a way to access services, these approaches varies among  infrastructures. The user experience is not homogenous, due to different technologies used (federated access, X.509 certificates and in same cases a mix of them)  as well as different procedures. AARC is looking at improving this aspect by enabling users to access services across the e-infrastructures regardless of the credentials they own and by harmonising policies. However the process will need time as the adaption of common policies cannot be a top down process.   Further requirements identified covered:

* Research entitlement management.  This poses a new requirement on the AAI where a user can delegate somebody else to access data (i.e. data generated from earthquake sensors in Italy made available by the owner to a user in the UK).
* group management to access facilities
* Role based access control and multi-factor authentication
* persistent IDs and how to use them in existing AAIs
* Scalability vs specific requirements
* Support for Citizen scientist; openAire noted there already users in the Horizon2020 do not user federated access.

The discussion moved to open access - are researchers always happy to share? Do they benefit of an AAI that support this? The MuG project representatives noted that the trend is changing and more researchers particular on the medical side, are happier to share data and to use third parties to hold their data. In many cases they only share the data after the research is published. Not all researchers however have a proper AAI in place.  The representative of the EoCoE project noted that in France strong national regulations are in place to access machines in super computing centres; this more in general shows how national laws have an impact on research and sharing of data.

### Summary

Whilst the group did not come up with a concrete set of actions to carry over, the discussion however highlighted a few opportunities. Namely:

* Cross sector is an interesting case both in terms of privacy (particularly when dealing with medical data, eIDAS and open government data) and technology. eID could be used as an alternative to support citizen scientists.
* Long term preservation, what is that standard pipeline? All agreed on the importance of the AAI to manage access to the data, although it was agreed that the  general preservation aspect should be dealt by each community. Is there an opportunity to engage with the public sector (particularly with national data centres)?
* data protection: health sector was mentioned as very critical for data protection. To this extend it was noted that sharing information on common practices would be helpful (safe harbour was mentioned).
* Terms of usage in different e-Infrastructures are different; even more so when using commercial facilities. Should we provide a comparisons? Is this in scope?

 A discussion followed on the lack of mechanisms to monetise the results of the research; in many cases sharing the results of a research makes it more difficult to exploit it with SMEs. The EC has a policy to encourage the funded projects to make data available after the research is done, but to also support patents Lastly we talked about how do we ensure that users can still access their research data when they move. Licia noted that some projects to look at this space already exists, namely eduKEEP (led by the Swiss NREN SWITCH).  It was agreed to use the digital4Science platform to share information among the projects in this group.

#### [**Requirement update and action added**](https://ec.europa.eu/futurium/en/comment/5762#comment-5762) **-** [**Jules Wolfrat**](https://ec.europa.eu/futurium/en/users/jules-wolfrat?flavour=digital4science)

Under the requirements is mentioned the use of persistent IDs and how to use them. Particularly this was mentioned in relation to the move of a person between organizations. Related to this is what happens with the ownership of the data of this person. Who is the owner, the person or the organization, or both?

About actions, I can add that at the moment PRACE is setting up a pilot with EUDAT where a PRACE user uses EUDAT services based on PRACE AAI information. However this will not be based on existing standards for exchanging credential information, at least as far as is foreseen now. Terms and Conditions should be clear before such a service can go in production, so it may be a showcase of what is possible.

[**Users moving and ownership**](https://ec.europa.eu/futurium/en/comment/5763#comment-5763) **-** [**Keith Jeffery**](https://ec.europa.eu/futurium/en/users/keith-jeffery?flavour=digital4science)

Jules, good point. It actually depends on the employment contract of the person. In my last employment all IP belonged to the employer. However, in many universities it is shared or owned by the person. There is plenty of opportunity for confusion here!  
Keith

#### [**General comments**](https://ec.europa.eu/futurium/en/comment/5764#comment-5764) **-** [**Keith Jeffery**](https://ec.europa.eu/futurium/en/users/keith-jeffery?flavour=digital4science)

Introduction: traversal ==> transversal  
Requirements: the real need is for the end-user to have homogeneous access (SSO) over heterogeneous AAAIs; X509 (in LDAP, AD etc), Federated access mechanisms, eduGAIN, ... VRE4EIC is (via EPOS) looking at UNITY. Where does AARC fit in? However the real requirement is much more complex than current AAAI - a matter of rights management (of users, datasets, software services, resources, end-to-end services) in all combinations. Implies RBAC and MFA - but MFA using many parameters describing the user, dataset, software etc - not simple 3-way as used now for users.

# HPC cluster & Centres of Excellence: meeting report – Dimitrios Axiotis

Dear participants,

Thank you for your participation in the eConcertation meeting. Please find below a short report of the HPC group for your information, comments or additions!

The discussions focused on the questions posed by the EC relating to the HPC strategy (e.g. who are the users, what are the applications, how close to exascale, co-design). A print-out summary of some of the responses of the projects was distributed by the EC in the meeting (see presentation attached). The projects were also encouraged to upload their responses in the D4Science platform.

* It was agreed that the projects should enter in a structured dialogue with PRACE and ETP4HPC.
* Projects shall try to raise user community awareness of their existence (good awareness within academia, lagging in industry).
* It is important that the projects are sustainable after the end of the EU funding.
* Projects shall participate in the European HPC Summit, in Prague in May 2016.
* Training and education was of interest; the EC informed about Marie Sklodowska Curie actions (see more info below). The CSA EDISON aims at accelerating the process of establishing the profession of Data Scientist.

Regarding synergies among the CoEs, a meeting took place between the CoE coordinators in the context of the EXDCI event in Rome in September 2015; it was agreed that the CoE projects will hold internal regular meetings to organise joint activities like dissemination but also technical level cooperation.

**Annex: Marie Skłodowska-Curie Actions**

The topics in WP 2016-2017 Marie Skłodowska-Curie Actions which address staff exchange between research and market are:

**MSCA-RISE-2016 & MSCA-RISE 2017: Research and Innovation Staff Exchange**

Objective:  The RISE scheme will promote international and inter-sector collaboration through research and innovation staff exchanges, and sharing of knowledge and ideas from research to market (and vice-versa).

**MSCA-ITN-2016 & MSCA-ITN-2017: Innovative Training Networks**

Objective:  The Innovative Training Networks (ITN) aim to train a new generation of creative, entrepreneurial and innovative early-stage researchers, able to face current and future challenges and to convert knowledge and ideas into products and services for economic and social benefit.

#### **INCO Group -** [**Minutes of the tel conference for the preparation for e-concertation**](https://ec.europa.eu/futurium/en/comment/5753#comment-5753)

#### Attendees

* + Dale Robertson, GÉANT
  + Damien Alline, IRD, TANDEM (<http://www.tandem-wacren.eu/>)
  + Ognjen Prnjat, GRNET, VI-SEEM
  + Tiziana Ferrari, EGI
  + Florencio Utreras, CLARA, MAGIC

Simon Taylor, Coordinator of SciGAIA sent his apologies for not being able to attend due to previous commitments.

The Meeting starts at 14 UTC on October 7, 2015.

#### **1. Presentation by the projects**

From the agenda, it is clear that the main idea is that this group meet from 13:30 to 15:15 in the afternoon and the issue some presentation on possible answers to the questions raised by the EC.

#### **3. Discussion on the possible answers to the questions raised by Aniyan Varghese**

Aniyan Varghese sent the following questions:

* + How to achieve synergies and co-operation of projects in this group and beyond?
  + How to involve local players and stakeholders including decision makers? How ICT skill development and training can be applied locally and transferred to local trainers for sustainability?
  + What are the common methods and tools that can be applied by all or most of the projects?
  + How to maximise the use of lessons learned in the past?
  + How to collect and disseminate local lessons learned effectively and measure KPIs? How to give visibility to local/regional innovation and explore wider take-up?
  + How to establish an international co-operation strategy for eInfrastructure, and what role the projects can play?

There is consensus that these questions are very large to be answered one by one and maybe the EC will summarize the contributions during the e-Concertation Meeting to get a better picture of possible answers to them.

The following possible cooperation issues are discussed:

* + share applications across the projects; applications of common interest should be promoted beyond the project for broader adoption
  + exchange approaches towards service catalogues
  + harmonize of training programmes (TBD)
  + cloud federation and cloud standardization
  + groupware standards
  + research community building

The participants share their views on this.

GEANT cloud service catalogue will form a blueprint for MAGIC cloud catalogue. EGI maintains an Application Database where different types of application services / cloud images are available. VI-SEEM will be developing a service catalogue, while also EUDAT and PRACE service catalogues should be taken into account. TANDEM will be promoting the existing NREN Services and elaborate a roadmap for the West and Central Africa for implementation.

The EGI Platform is being built by the EDISON project <http://www.egi.eu/services/training_marketplace/>

There is also an application database for sharing of tools: <https://appdb.egi.eu/> It is agreed to continue contacts in order to reach more concrete collaboration points.

**The meeting is adjourned at 12:15 UTC.**

[**INCO slides**](https://ec.europa.eu/futurium/en/comment/5759#comment-5759) **– Yannick Legré**

The slides presented by Aniyan and discussed during the session are available here: <https://documents.egi.eu/document/2649>

**Bottom of Form**

[**Actions and Next steps: Meeting summary and conclusions**](https://ec.europa.eu/futurium/en/comment/5761#comment-5761) **- Aniyan Varghese**

Actions and next steps - what, when, who, how?  
1. Global Policy: Including How to bring the right Stakeholders  
Who: RDA, GEANT, RED CLARA, IRD, EGI  
What: 2-3 page discussion paper  
When: February 2016  
2. Synergy (working together):  
Who: All INCO projects  
What: 3 slides to present the expected outcome of International co-operation activity; share project url on D4Science  
When: Before end of 2015  
3. Common methods and tools:  
Who: REDCLARA  
What: Collaborative Platform  
When: asap  
4. ICT for Development  
Who: GEANT, RED CLARA, IRD  
What: Identify opportunities  
When: tbc  
5. KPIs  
Who: Hilary (RDA), Ognjen (GRNET)  
What: 2-3 page discussion paper  
February 2016  
Other topics for actions: (did not have time to discuss, to be followed up on-line)  
Sustainability, Lessons learned, Wider take-up, Networking, Catalogue of Service,  
Dissemination, Applications, Federation.

#### [**Notes “Skills and Training” breakout session - 9 Nov 2015**](https://ec.europa.eu/futurium/en/comment/5780#comment-5780) **-** [**Yuri Demchenko**](https://ec.europa.eu/futurium/en/users/yuri-demchenko?flavour=digital4science)

**Notes from the “Skills and Training” breakout session - 9 November 2015, Brussels - As part of the 11th e-Infrastructure Concertation meeting – 9 November 2015, Brussels**

**SUMMARY of discussions and agreed actions**

Presented to Plenary session; Rapporteur – Yuri Demchenko (EDISON)

1. Active discussion on tasks, existing and planned activities that identified many synergies

* + Common understanding that there is a need to recognise a new profession.
  + Many names and definitions exist: agreement on the common name “Data Scientist” that correctly reflects the new profession which combines research methods on working with data with a focus on specific scientific domains or industry sectors
  + The new profession definition needs to consider the views and needs of industry as they are currently driving demand and will consume a large number of future professionals
  + Despite the demand for Data Scientists growing, the profession is still not formally recognised which creates problems with competences and skills definition and building career paths. Defining the profession will help with job alignment and establishing career paths.
  + New profession definitions typically include a definition of the required competences and skills, taxonomy, Body of Knowledge, model curriculum, and certification scheme and certification authority. Best practices and experience of European e-Competence Framework and ACM/IEEE should be re-used.
  + Currently many training activities co-exist across many projects. There is a need for exchange of information about available courses and schedule. As a first step, compatible description should be agreed using common metadata. Use of and compatibility with existing taxonomies and classifications is seen as beneficial, however the problem is that Data Science is still not supported.

The following actions have been agreed as a result of networking session. 1. Create inventory of existing activities Content: Develop questionnaire and run survey and interviews Lead by EDISON Date: January 2016

2. Promote and advertise existing training resources between projects and for target communities Content: Coordinate and develop a synergy approach in the short term. Lead by Transverse projects: EUDAT, EGI-Engage, PRACE, SESAMENET Date: On-going

3. Define the profession of Data Scientist Content: Share EDISON working documents for discussion and collect external contribution Lead by EDISON Date: Mid-January 2016

4. Recognition of the Data Scientist profession Content: Specific focus on interoperability with industry and gender balance. Lead: ECAM Date: On-going

**A. General discussion: topics discussed**

1) Professional designation of Data Scientist 2) Need for new profession of Data Scientist 3) Primarily demand from industry 4) Gap in industry with experienced staff 5) Key thing is employability across research and industry

**1) Professional designation of Data Scientist**

There are number of terms used or linked to the new profession such as data researcher, data analyst, Big Data specialist, research infrastructure engineer, data manager, data archivist, data librarian, etc. The term “research technologist” has become quite well respected but covers a slightly different role and range of responsibilities. Data Scientist is recognised as best reflecting the specifics of the new profession in bringing scientific methods to data analysis related to different scientific and industry domains. The Data Scientist definition should not be limited to research as the industry demand is an order of magnitude higher. Research and industry should be considered as possible career paths – there will be benefits for both.

**2) Creating a new profession of Data Scientist is a multi-dimensional problem and requires a structured and systematic approach.**

**3) Data Scientist is cross-domain and cross-organisational role.**

An important feature of the profession and environment is the openness that requires and allows collection and processing of wide range of data as well as communication with multiple organisational roles and bridging multiple stages in organisational workflow and data lifecycle. This attitude is demonstrated by leading data driven companies such as Google, Facebook, Twitter, Booking.com. In contrast Apple have Data Scientists but behind closed doors, and they struggle to recruit as they cannot do open science.

**4) What are components of the profession definition?**

Typically the following components: competences and skills, taxonomy, Body of Knowledge, model curriculum, and certification scheme and certification authority. Data Scientist skills and knowledge must be linked to background ICT and HPC skills and knowledge. This is also beneficial for defining re-skilling and building career path. Soft skills need to be addressed but require more practical training.

**5) Data management knowledge is becoming important and should be part of the background knowledge required for all Data Scientists.**

Data management literacy should include: - Metadata - Data lifecycle - Change management Data validation and curation - Quality - Research reproducibility Data management should reflect all stages of the data lifecycle

**6) Education and training as a need for existing e-Infrastructures, current technologies and sustainability of the future profession.**

Training the trainers is important, especially for the RDA

**7) Gender issue**

Gender stereotyping must be addressed in the profession definition, developing training materials and conducting training sessions. 8) Shortage of experienced Data Scientists can be addressed by using Marie Curie initiative. • How to engage with SMEs that are driven by profit? • Need to present them with success stories, e.g. productivity matters, access to technologies, information and (new) markets. • Training needs to lead directly to results.