Minutes of the Parallel Session 1a "Data Management and EOSC involvement"

1st ERIC Forum

16-17 November 2017

BBMRI-ERIC and Medical University Graz, 8010 Graz, Neue Stiftingtalstrasse 2-4

Chair: *Petr Holub* BBMRI-ERIC Senior IT/Data Protection Manager and CTO

Rapporteur: Dr. *Juan Miguel González-Aranda* LifeWatch ERIC CTO

Petr Holub made an excellent presentation on EOSC & Data Management, putting in value main Data in Science issues, followed by a very interesting and productive discussion among session participants:

- Data curation is more costly than storage itself, about all at long term.
- Importance of developing common core components to guarantee the Data Reproducibility for further researches.
  - In fact, Data Reproducibility is an extra to consider-complmenet the Data Life Cycle stages.
  - Data Reproducibility is also associated to Data Provenance concept.
- Assurance of Data "FAIRness" (Findable Accessible Interoperable Reusable): A very nice example on Raw Data from Apollo missions was presented: Not-readable, because they lack of a proper metadata framework.
  - But, is it FAIR enough ?: In fact, only 50% from Pharmacological data comply it and associated (opensource or not) repositories need a proper documentation of what was done.
- Importance of defining proper Data Management Plans (DMP) from the early stages, in order to, amongst other issues, to calculate costs of Data Management associated to all Data Life Cycle stages.
- Digital Object Identifiers (DOIs) & Persistent Data Identifiers (PIDs) are essential, although it is difficult to define them (particularly PIDs) in order to guarantee interoperability mechanisms among data coming from different Organizations, mainly because of the underlying Complexity associated to the nature of Scientific Communities themselves.

- Certifications and Service Legal Agreements (SLAs) mechanisms: Some opinions, points of view from the Biodiversity and Cultural Heritage (Humanities in general terms) ERIC representatives were also presented during the session.

- The problem is not the standards to follow but the protocols.

  - Data Provenance and Change Management tools. It is important to mantain scientific consistency. A pragmatical approach is necessary to define common procedures under not obligation to fit 100% of standards.

- Data Intellectual Property Rights (IPR) as an additional and key factor, above all considering Service Legal Agreements (SLAs) from both Data Providers' and further Consumers' perspectives. Radioastronomy ERIC representatives emphasized on this issue.

- General Data Protection Regulation (GDPR) allows derogations for Member States, as ERIC are mainly subject to Hosting Countries' national legislations.
  I
  - Country to country data transfer mechanisms are still costly. Recent experience shows even difficulties in finding interoperable mechanisms by using, e.g., Edugain mechanisms among Organizations (Universities, Research Centres, etc.).

- Importance of Authentication and Identification (AAI) systems: A nice conceptual framework based on AAI applied to Life Science ERICs was also introduced during the presentation.

  - What roles ?: By properly defining-considering the (1) Service providers/consumers; (2) Domain-specific bodies for FAIR data certifications; (3) Project participants.

- Focusing on EOSC, it is not clear if final "Data Consumers" are going to directly sign their corresponding Service Legal Agreements (SLAs) with EOSC, when it is obvious-clear that many of its different components are designed by Community Stakeholders, in turn directly engaged with ERICs.

  - Important: Who are in charge of custody ?. EOSC Governance Scheme is not clear at all. "EOSC in the service of ERICs or viceversa ?". A good solution should be the creation of a proper Marketplace.

- Moreover, it seems very likely that EOSC Pilot initiative is going to be "toppled"-serously reformed by Member States, but however and in contrast, it seems EOSC Hub initiative is doing quite well and it is expected to deliver reasonable solutions mainly by providing Thematic e-Services.

- ✓ *In view of these circumstances and therefore, session participants agreed that it is necessary to establish a Data Working Group among existing ERICs in order to define and agree Data Policy Commons.*

  - Even though some ERICs Data Managers-CTOs are engaged (occasionally or not) with existing and relevant initiatives (such as, e.g., the Group of European Data Experts-GEDE dependent on Research Data Alliance Working Groups which usually advice to EU Commission on this regard), that is not enough as it is essential to perform an exercise of good practices identification based on a previous benchmarking-assessment analysis among different ERICs activities in this regard.

  - In fact, ERICs should play a more active role the definition of high-level EOSC policies ("Their voice must be heard and heeded in relation to EOSC decisions").

  - *These should be reflected in the associated INFRASSUP proposal Tasks and Deliverables, so that it is proposed to share their associated co-leadership between LifeWatch ERIC and BBMRI-ERIC CTOs above mentioned, who will also serve as spokespersons in those Fora where it properly applicable.*