

# A Study to Define a Linked Open Data Platform of Interoperable Repositories to Enable Open Science

R. Barbera, S. Bianco, M. Maggi,  
D. Menasce, L. Patrizii - INFN - Italy  
M. Saccone, L. Trufelli, L. Reggiani - CNR - Italy  
R. Smareglia - INAF - Italy  
P. Budroni - University of Vienna - Austria

L. Alagna, A. Masiero  
Representatives of Italy to the OECD Global Science Forum

Version: 0 - Created: 20171118

Version: 2

Revision: 20180322 20180323

## 1. Introduction and context

1.1 - In its pioneeristic report *“Making Open Science a Reality”* [1] published in 2015, OECD defined Open Science as *“a means and not an end in itself and [...] much more than just open access to publications [...]”*. Among the various Open Science aspects and enablers discussed in that document, opening research data was considered key not only to accelerate scientific discovery and improve science reproducibility, but also to speed up innovation and improve citizen engagement with research, hence benefiting the whole society.

1.2 - On October 18, 2017, an OECD-GSF “Workshop on Open Data” discussed recent activities on Open Science repositories and concluded on the existence of two areas of interest:

- Quality and certification of Open Data: trust building, re-fitness or use and reusability (certification and reproducibility);
- Open Data repositories’ issues: multi/trans-disciplinary use of data, including interoperability and standards.

1.3 – In order to address interoperability and federability of Open Data repositories, the concepts of FAIR (Findable, Accessible, Interoperable, Reproducible) data [2] and “5 stars” data [3] have recently been coined and they are actively pursued at various political [4,5] and policy making levels [6-10]. However, large scale, cross-country, cross-disciplinary and multi-purpose implementations of FAIR and “5 stars” compliant Linked Open Data (LOD) infrastructures are still missing.

1.4 - The importance of Open Data was also extensively underlined throughout the book *“Opening Science”* [11]. Indeed, Open Data as a concept and Data Repositories (DRs) as the

implementation of that concept, are indicated as key elements of the Democratic and the Infrastructure “schools of thought” of Open Science, respectively (see chapter “*Open Science: One Term, Five Schools of Thought*” in [11]).

1.5 - Complexity, variability and heterogeneity are all defining attributes of modern Big Data Science and the concept of “openness”, in such a multi-dimensional data universe definitely requires the establishment of a common framework to insure its interoperability among all stakeholders and users. It is believed that such a goal can only be reached by adopting a shared, unitary approach taking into account both the large phase space of the problem along with the specificities of the individual communities involved.

## 2. Objectives and coherence

The main objective of the present project is **to carry out a study and deliver a report on enablers, obstacles, policies and recommendations for the setup of a cross-country federation of standard-based, multi-domain Linked Open Data repositories leveraging as well as addressing the concept of Open Science.**

The main output of the proposed study would be a multi-dimensional, multi-level reference framework to allow researches (and users in general) to openly and easily share, retrieve, locate data, resources, instruments, services and results connected to data made publicly available.

To do that, the report will:

- Identify different stakeholders (researchers, policymakers, ICT specialists, academic institutions, citizens, industries, etc.);
- Identify and analyse a multiplicity of different cultural contexts, practices, approaches and the relative technical languages.

Furthermore, it will address several important questions of Open Data, such as:

- In the context of data handling, what are the lessons learnt from Big Science experiments that can be applied to “Long Tail of Science” groups?
- What are the technical/political challenges to achieve full interoperability among Open Science repositories?
- Who are the key actors for the deployment of Open Science repositories?
- What is the role of Data Stewardship?
- Who owns the data?
- Who is entitled to provide the certification of the quality of an Open Science repository?
- Is it possible to broadly categorize the type of information in Open Science repositories (Big-Science/Research, Citizen demographics, Social and Human Sciences, Law Studies, Cultural Heritage, Health, Environment, etc.)?
- Can the above disparate communities use a common infrastructure?
- What is the correct granularity needed to define metadata semantics to satisfy the different community requirements?
- What is the role of private companies and societal organisations?
- Etc.

In conclusion, the study is meant to be the natural continuation/evolution of those that led to [1] and [12] and it is fully coherent with OECD [13] and GSF [14] missions.

## References

1. OECD, "Making Open Science a Reality" (<http://dx.doi.org/10.1787/5jrs2f963zs1-en>).
2. <http://www.force11.org/group/fairgroup/fairprinciples>
3. <http://5stardata.info>
4. European Commission, "Open innovation, open science, open to the world - a vision for Europe" (<http://dx.doi.org/10.2777/061652>).
5. European Open Science Cloud (<https://goo.gl/qzYpNo>)
6. High Level Expert Group on the European Open Science Cloud (<https://goo.gl/hEt61N>)
7. Workshops of European Open Science Cloud (<https://goo.gl/97TYoK>)
8. European Commission, "Guidelines on FAIR Data Management in Horizon 2020" (<https://goo.gl/MwiHqX>)
9. EOSCpilot Project, "D3.1: Policy Landscape Review" (<https://eoscipilot.eu/sites/default/files/eoscipilot-d3.1.pdf>)
10. EOSCpilot Project, "D6.3: 1st Report on Data Interoperability: Findability and Interoperability" (<https://eoscipilot.eu/sites/default/files/eoscipilot-d6.3.pdf>)
11. "Opening Science - The Evolving Guide on How the Web is Changing Research, Collaboration and Scholarly Publishing", Sönke Bartling and Sascha Friesike (Eds.), Springer 2014 (<http://book.openingscience.org>)
12. OECD, "Business models for sustainable research data repositories" (<https://goo.gl/AX8e69>)
13. <http://www.oecd.org/about/>
14. "THE OECD GLOBAL SCIENCE FORUM - Strategic Directions 2015-2019" (<https://www.oecd.org/sti/sci-tech/STI-GSF-brochure.pdf>)