

e-IRG Document

e-IRG
e-Infrastructure
Reflection Group

Paving the way towards a general purpose European e-Infrastructure

Summary of Policy Recommendations Drawn from the e-IRG Blue Paper on Data Management

February 27, 2013

<http://e-irg.eu>

Acknowledgments

This document was produced by the e-IRGSP4 project for e-IRG.



e-IRGSP4 is supported by the FP7 Capacities Programme under contract nr RI-632688.

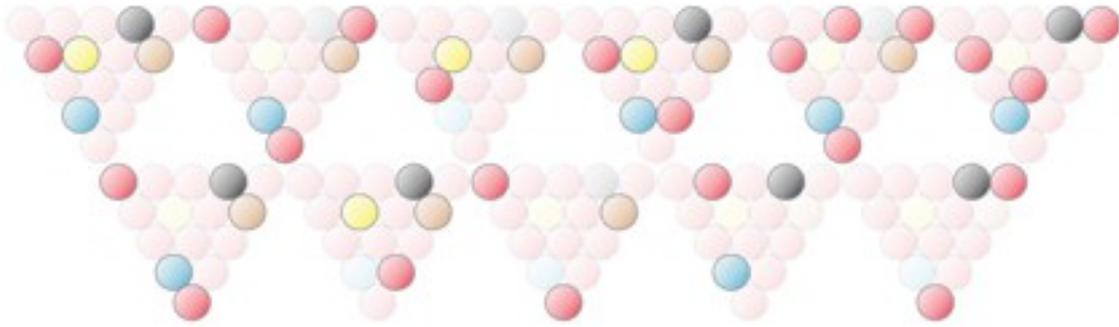
Version 1.0

e-IRG secretariat

P.O. Box 93575,
NL-2509 AN The Hague,
The Netherlands
Phone: +31 (0)70 344 0526
E-mail: secretariat@e-irg.eu

Visiting address

Java Building, Laan van Nieuw Oost-Indië 300
NL-2593 CE The Hague
The Netherlands



Summary of Policy Recommendations Drawn from the e-IRG Blue Paper on Data Management

Gelsomina Pappalardo

Sverker Holmgren

Norbert Meyer

Bjørn Henrichsen

Lorenza Saracco

Research infrastructures (RIs), such as the initiatives on the ESFRI roadmap, produce and are dependent on rapidly increasing amounts of data. For research and society to take full benefit of the major investments in RIs the data needs to be made openly and easily available for researchers, over wide spans of fields, in sustainable settings. To enable this, the data needs to be managed, stored and preserved in a cost-efficient way, with appropriate quality and safety assurances. Also, access to the data across borders and domain boundaries must be secured. e-infrastructures provide the versatile services and tools needed for both data management and access, but the development of such infrastructure must be complemented with an effort on RI policy and coordination to accomplish the goals above in practice. This effort must be driven by the needs of the researchers, but many stakeholders needs to be orchestrated to complete the build-up of data infrastructures rapidly while still arriving at sustainable and cost-efficient solutions.

In 2012, e-IRG presented a Blue Paper on this topic with a focus on the needs of large RIs such as the ESFRI Roadmap initiatives. This summary contains the main policy recommendations of the 2012 Blue Paper, extracted to provide a comprehensive list of actions that should be taken to arrive at situation where research and society can reap the full benefits of major RIs.

As a fundament for RIs, sustainable e-Infrastructure services for enabling access to, storing, preserving and curating large amounts of data need to be in place. Policy makers are recommended to take action to ensure that

- Roles (e.g. end users, data owners, infrastructure providers, service providers, and researchers on data management) are identified and , when appropriate, partitioned between different actors to ensure effective and cost-efficient solutions, fulfilling the needs of the end users and data owners.

- Governance and mandates for different actors are clarified and their way of interacting is sufficiently formalised. Actors specializing on different tasks ensures that synergies can be exploited, leading to cost-efficient implementation of services. Clear responsibilities and formalised relations ensure that the relevant quality of services can be maintained. Funding paths are defined and sustainability for all parts of the e-Infrastructure is secured.
- Costs for different services and procedures are made transparent and that different options for implementing them are investigated.

Also, to ensure that data will be available across borders and disciplinary domains, RIs and e-infrastructure providers are recommended to take appropriate steps to

- Ensure that data formats are standardised and contain sufficient information on the data (metadata) to enable global usage within the discipline, across disciplines, and in new research settings that could possibly not be envisaged at the time of creation of the data.
- Build e-infrastructure solutions consisting of multiple layers, successively adding more specialised higher-level services using standardised interfaces. Here, different layers can be provided by different actors.
- Adopt a global, standardised lowest-level data infrastructure layer, including e.g. authorisation and authentication and persistent data identifiers. Here, federative approaches could be used to include existing solutions.
- Define and successively move towards a common second-level data storage layer where cross-related requirements between different RIs are identified and utilised to enhance cost-efficiency and quality. Also here, standardised interfaces and federative approaches should be used to include existing solutions.
- Ensure that quality of the e-infrastructure services and the data security is delivered at a level which is relevant for the data at hand.

In this process it is important to take into account already existing and used e-infrastructures and services, which should be integrated with the potentially new infrastructures using well-defined and standardised interfaces. The overall goal is clear: adopt existing e-infrastructure solutions whenever available, remove isles, and build standards for new challenges which are at the beginning of the development and can be easier integrated in the multi-layer structure.

For the ESFRI initiatives now under implementation some of the mentioned points are partly followed up in work packages in the cluster projects funded by the European Commission. These projects can be seen as a first step towards more coordinated European data infrastructures for some of the new and hopefully also some of the previous established Pan-European ESFRI RIs. The present cluster projects are mainly covering one or two scientific areas (for instance SSH or medicine). There is a need, however, also to secure synergies between the different scientific areas for instance by having closer cooperation between projects within social sciences and humanities on the one hand and environmental projects and/or the medical sciences. Only by coordinating the e-infrastructure activities and building a multi-layer data structure can we succeed in establishing an RI system in Europe which will avoid separated isles of infrastructures for a specific scientific domain.

The RIs are recommended to

- Continue the work in the cluster projects and to include new incoming ESFRI initiatives under one of the umbrella of one of them.
- Coordinate horizontal activities between ESFRI cluster projects.

This will allow that requirements are synchronised, and that services developed within the clusters are deployed in the early stage of implementation of ESFRI initiatives.

One way for ESFRI to follow up the ideas from the e-IRG Blue Paper is to give advice both to the Commission as well as national funders concerning the need for coordination of data infrastructure activities. This is necessary both of economic and scientific reasons. Another and perhaps parallel way could be to ask e-IRG to play a more active part in organising or give advice based on the many initiatives now coming from different bodies concerning not only data but e-infrastructure activities in common. By a better coordination of the activities within the area it will be easier also for ESFRI to follow up on the policy level.

References

e-IRG Blue Paper on Data Management

http://e-irg.eu/documents/10920/238805/e-irg-blue_paper_on_data_management_v_final.pdf



<http://e-irg.eu>