

Background Speaking Points

Scientific Data e-Infrastructures in the European Capacities Programme

Krystyna Marek

European Commission Activities in e-Infrastructures

The European Commission and EU Member States are investing in different domains of **e-Infrastructures**. The aim is to connect researchers, scholars, educators and students through high speed research networks (GÉANT project), provide access to cooperative grid-computing platforms (EGI, e-Science grids projects), develop supercomputing capacity for very demanding applications (PRACE) and help the emergence of a robust platform for access and preservation of scientific information (Scientific Data e-Infrastructure - SDI).

e-Infrastructures for e-Science

Striving for world leadership in **e-Science**, establishing e-Infrastructures as a sustainable utility and exploiting them to promote innovation are the three vectors of a renewed European strategy to support the ground-breaking science of 2020 and beyond. This strategy was presented in the European Commission communication to the European Parliament and Council of Member States.

e-Science is generally defined as the combination of three different developments: the sharing of computational resources, distributed access to massive datasets and the use of digital platforms for collaboration and communication. By its very nature, e-Science is broad in scope, covering the complete spectrum of modern research and education from the arts to the traditional physical science, from the theoretical to the experimental, from the commercial to the academic, amongst numerous other dimensions.

Until recently, the concept of e-Infrastructure has usually been defined minimally, to include networks, authentication and authorisation mechanisms, middleware, computational resources (in particular high-performance computers), and those which enable collaborative working, including Grid technologies.

We adopt a wider interpretation and include technologies of various kinds for creating, collecting, annotating, manipulating, storing, finding and re-using information and services such as those to provide user support, and training, preservation. Further, we include information resources and associated tools such as vocabularies, ontologies, rights management and privacy protection systems, and curation. Several of these resources depend upon manual human input.

An e-Infrastructure for e-Science digital repositories is taken to be the technical and administrative framework and facilities underlying e-Science digital repositories.

Scientific Data e-Infrastructure in Europe

The European data infrastructures landscape as a whole remains heterogeneous and scattered. The approaches to the most pressing problems related with storage, access, quality assurance, and preservation are not yet well coordinated and the integration with e Science computational environments has not been achieved. It is estimated that only a very limited part of European research output is managed in digital repositories.

Capacity building in the Scientific Data domain is therefore much needed. In launching the SDI area in FP7 several issues were taken into account:

- the need to gain insight into the practices and requirements of major stakeholders in research: scientists, libraries and data centers, publishers and research funders;
- the importance to coordinate with EU Member States and leverage national e-Science initiatives;
- the need to bring together dynamic scientific communities with their most pressing data management problems;
- the objective to provide e-Infrastructures capable to fulfil European policies for widespread access and long term preservation;
- the need to exploit synergies with e-Infrastructures for networking and computing in order to offer integrated services supporting e-Science;
- the objective to place Europe in key global scientific partnerships acknowledging that science is mostly an international endeavour which requires bridging national and institutional policy and funding structures;

Investments in Scientific Data e-Infrastructure in Europe

The **Scientific Data e-Infrastructure** (SDI) domain was launched in 2007 under the Capacities programme of FP7. It complements the investment done by the European Commission, in a coordinated effort with EU Member states, in other domains of e-Infrastructures.

First two calls for proposals in FP7

Having in mind the emergence of e-Science and the need for Europe to be an active and inspiring partner in **global scientific collaborations**, the European Commission involved the wider community of interested stakeholders in several preparatory workshops to discuss and provide recommendations for the SDI area. It also launched several projects and studies to analyse the landscape of repositories in Europe.

Issues like promoting widespread access to scientific information resources, supporting multi-disciplinary research, sharing practices and software between communities and exploiting synergies of investments were explicitly set as some of the desirable features for the European e-infrastructure for scientific repositories. The heterogeneity of requirements and lack of maturity of the running initiatives resulted in splitting the available resources for the SDI area in two open calls for proposals.

The first call gave the opportunity for communities with well identified data challenges to propose projects addressing their

most pressing needs. The second call was designed to encourage communities to identify and tackle common (or boundary) problems, which cut across different scientific domains, to promote multi-disciplinary collaborations integrating with other e-infrastructures domains of networking and computing.

The FP7 Open Access Pilot: the third call for proposals

A third call for proposals was organised to address the European Open Access (OA) policy initiative.

Launched in August 2008 as a result of an extensive consultation with EU Member States the FP7 OA pilot follows the EU Competitiveness Council conclusions inviting the Commission "to experiment with open access to scientific data and publications resulting from projects funded by the EU Research Framework Programmes in order to assess the appropriateness of adopting specific contractual requirements".²

A key objective of the FP7 OA pilot is to establish fast and reliable access to EU-funded research results, specifically peer reviewed research articles, in order to drive innovation, advance scientific discovery and support the development of a strong knowledge-based economy.

Under the OA pilot, grant recipients in seven areas of FP7 (energy, environment, health, parts of information and communication technologies, research infrastructures, science in society, and social sciences and humanities) are required to deposit peer

¹ Seventh Framework Programme for research and technological development.

reviewed research articles or final manuscripts resulting from their FP7 projects into an online repository and make their best efforts to ensure open access to these articles within either six or twelve months after publication (the so-called embargo period).

The OpenAIRE project was selected as a result of the third call of proposals, involving 38 partners from all EU Member States plus key European research laboratories. The project is supported on the idea of federating national, institutional and thematic repositories that can share resources and information.

This initiative puts Europe in a world leadership position leveraging the high-quality work done in EU Member States and providing an open platform for researchers, students, educators to access scientific information.

The current picture: projects in the area of Scientific Data e-Infrastructure

Overall, there was a very high response and interest to all calls with oversubscription in the order of 6:1. It was possible to launch 15 projects corresponding to an overall investment of 45 million euro of EC contribution.

Many scientific disciplines responded to the calls. As can be seen from the description of the running projects (Annex) many scientific domains (bioinformatics, astronomy, space-physics, biodiversity, earth observation, chemistry, etc) succeeded in launching

relatively small projects (average duration of 2,5 years). In many cases user communities and technology developers address issues of common interest. Some projects are linked with national initiatives that could become European scale collaborations.

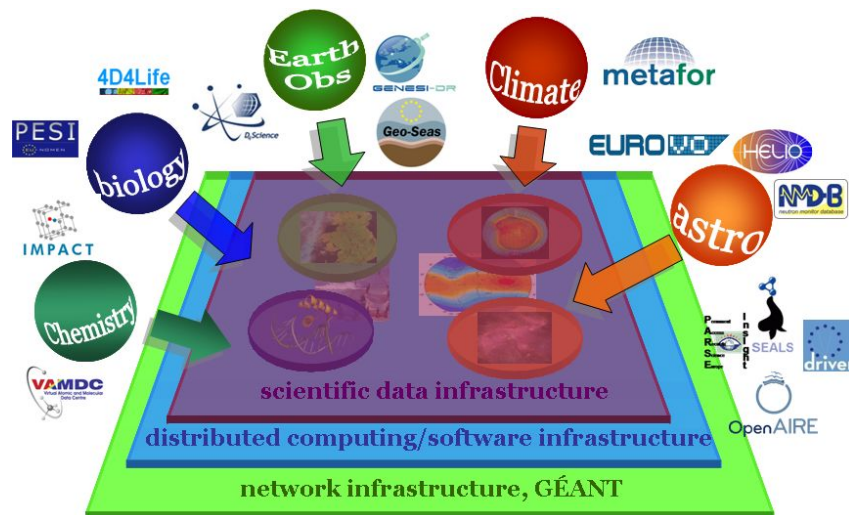


Figure 1 – Landscape of e-infrastructure projects in the SDI area.

Figure 1 shows the current landscape of e-infrastructure projects in the SDI area. It highlights the centres of gravity of the different projects the fact that data e-infrastructures build on functionalities made available by computing and network infrastructures. Some SDI projects are of horizontal nature like OpenAIRE, DRIVER or PARSE. Insight while others are more focused on (and triggered by) specific needs of scientific user communities. Some of these projects, in particular those coming from the second call (VAMDC, Geo-Seas), have partnerships of different (boundary) disciplines.

In conclusion

Incorporating **e-Science** digital repositories and their holdings into an **open information ecosystem** will deepen and broaden

Europe's Single Information Space. It will support new scientific methods and paradigms, improving the efficiency of the scientific process and its impact.