Research Data e-Infrastructures: Framework for Action in H2020

How policy makers and funders can target their limited resources at so many points of the data sharing ecosystem for maximum social and economic benefit is an enormous question to which there are no simple answers.

But two things are clear: that investment at all these points is necessary to create a fully realised data sharing system; and that gaps and redundancies in investment can best be avoided by a co-ordinated approach on the part of all agencies – governmental and non-governmental – that make research policy and fund research activities.[1]

European Commission

Directorate General CNECT

Directorate C: Excellence in Science

Unit C1 - e-Infrastructures

[1] From the project *Opportunities for Data Exchange* (ODE).

(table of contents)

Introduction

The emergence of **data driven science** reflects the increasing value of a range of observational, sensor, simulation, streaming and experimental data in every field of science. Data e-infrastructures link *knowledge territories* blurring geographical and disciplinary boundaries.

The present **European and global research data landscape is highly fragmented**, by disciplines or by domains (oceanography, life sciences, health, agriculture, space, climate, etc.). A variety of institutions, some national, some international, strive to deal with some aspects of data, but no effort exists where some degree of coherence is achieved or even sought.

The **stewardship of research data remains uncertain**: beyond the scientist(s) or laboratories that produce data, no specific group or profession (such as librarians for publications) is in charge of the preservation and organization of data. This makes data very vulnerable whenever institutional shifts occur (without mentioning broader concerns such as political or economic instability).

Some **research domains are experiencing exponential growth of data produced** with doubling rates that can be as short as a few months (seven months in the case of second generation sequencing of genes), while others plan new instruments that will suddenly produce enormous amounts of data. This is sometimes referred to as a "data tsunami" and presents major challenges in terms of storage, computation and long term preservation [10].

It is not only about the volume but also **scale and complexity**. Research is increasingly undertaken through global collaborations, using very large and heterogeneous data collections, huge computing resources accessible through high speed networks.

The European strategy in this area addresses the challenges of access and long term preservation recommending that results should become widely and openly accessible, preserved and curated in a cost effective way in order for citizens to **trust the scientific enterprise** as generator of the future knowledge and wealth.

Open data e-Infrastructures increase scope, depth and economies of scale of the scientific enterprise; they are catalysts of new and unexpected solutions to emerge by global and multidisciplinary research. They bridge the gap between scientists and the citizen and are enablers of trust in the scientific process.

In order to develop an **open, interoperable e-infrastructure for scientific data** one needs to find the right balance between standardisation and invention, control and freedom, performance and cost, public and private, international and local. It requires the engagement of several players: **(a)** data generators that can be big research infrastructure installations (telescopes, spallation centres, …) or medium size scientific laboratories, simulation centres or individual researchers, etc **(b)** technically able teams which define the disciplinary standards and the community service architectures **(c)** e-infrastructure and technology providers of common data services and **(d)** researchers, using the data for their scientific discoveries, who are at the centre of the whole scientific enterprise.

(empty)

Data infrastructures for Excellence in Science

Directorate "C" in DG CNECT is responsible for the Excellence in Science portfolio. It supports the *Digital Agenda for Europe* aiming at exploiting the vast potential of Information and Communication Technologies (ICT) to the fulfilment of the Europe 2020 Strategy to make the EU a smart, sustainable and inclusive economy.

The work of unit CNECT.C.1 on e-infrastructures is of key relevance to the development of the *on-line European Research Area* which, in a globalised world, is a pre-condition for an increased European competitiveness and an innovation driven economy. It supports the development and deployment of e-infrastructures that will make every European researcher digital and contribute to a seamless and open European space for online research. The actions focus on data-enabled science and engineering, computational infrastructure, research and education networks such as GÉANT, virtual research communities, and e-science software environments. A successful implementation will make Europe a leader of a global movement towards open, interconnected, data-driven and computer-intensive science and engineering.

The European Commission proposes here a set of actions that, if taken with appropriate resources and critical mass, can project Europe into the new world of data driven science.

The approach is to combine the expertise of scientific communities that know best their needs and the meaning of the data produced in their fields with the expertise of ICT communities capable of exploring the limits of high bandwidth communication, high-performance computing, open scientific software and virtual research environments.

The majority of the proposed actions will be implemented through the H2020 work programmes. The available H2020 instruments would bring together the combination of expertise ensuring the best balance of competences, critical mass and European dimension.

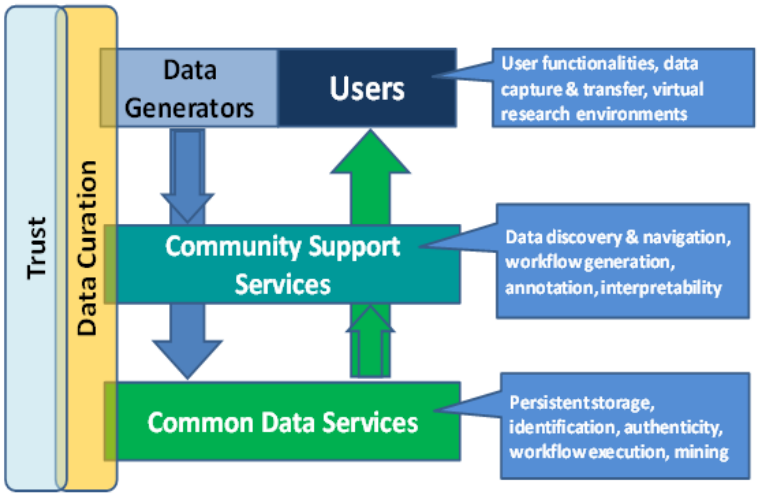
The implementation in a context of capacity building lifecycle, will balance service-driven and community-driven data e-infrastructures initiatives. They will respond to the e-Infrastructure requirements of researchers and research organisations and, at the same time, support European policies on Innovation and Societal Challenges.

(empty)

Background and references

The Framework for Action results from extensive consultations and reports, in particular:

(1) The **High Level Expert Group on Scientific Data** presented in October 2010 the **Riding the Wave** report [2] with a forward looking vision for a "data e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure a valuable asset on which science, technology, the economy and society can advance".



framework suggested by the High-Level Group on Scientific Data

(2) The European Commission is chairing the **G8+O5 working group on scientific data** infrastructures [3] that presented the final report during the group of senior officials (GSO) meetings in Cape Town (November 2011) and in Hamburg (April 2012). The GSO supported the conclusions of the report in particular the need to work together at global level to develop a "Global Collaborative Data and Knowledge Infrastructure".

(3) The European Commission in collaboration with the Association of Academies of Science (ALLEA) organised in April 2012 a **consultation event to prepare for H2020** [4]. It tackled the emergence of data driven science in very insightful workshops with the participation of experts coming from Europe, United States, Australia, Canada, China, India and South Africa.

(4) Knowledge Exchange published in November 2011 the report "**A Surfboard for Riding the Wave - Towards a four country action programme on research data**" [5] outlining an action programme to realise a collaborative data infrastructure. The report is a response to the Riding the Wave report.

(5) The European Commission adopted in July 2012 a package consisting of a **Communication and a Recommendation on aspects of open access, preservation and e-infrastructures for scientific information** [6,7]. It outlines a framework to optimize the incentives for scientific discovery and support collaboration across disciplinary and geographical boundaries, and to further develop the European innovation capacity.

(empty)

Europe's capacity building efforts

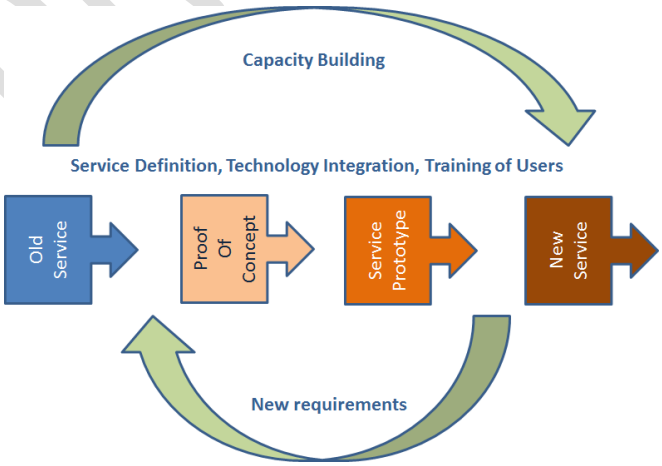
To create a competitive European Research Area, Europe has already invested a significant amount of resources in modernizing the European landscape of Research Infrastructures and facilities of excellence.

an immense global data factory: the forthcoming Square Kilometre Array (SKA) radio telescope will generate more data per day than the entire internet when it comes online in 2020; it will detect radio waves using dishes with a collecting area of 1 million square metres, distributed over a distance of 3000 km; SKA will drive innovation in query and knowledge creation on large databases, energy efficient computing (as very large computers like the SKA massive energy demands), and improved communications and networking technologies.

The ESFRI roadmap stretches across a range of scientific disciplines in different European nations and includes recommendations for a suite of ambitious initiatives in areas such as biological and medical sciences (ELIXIR, BBMRI), environment (LifeWatch, ENES, EPOS), social sciences and humanities (CLARIN, CESDA, DARIAH), geophysics and astronomy (SKA, EISCAT-3D, EPOS), physical and engineering (WLCG and ISIS).

A large number of data e-infrastructures, mixing competences of scientific communities and technology providers, have been launched in domains of astronomy (Euro-Virtual Observatory, …), earth and ocean observation (SCIDIP-ES, GeoSEAS, iMarine, DRIHM, …), climate (METAFOR, ESPAS), environment and biodiversity (4D4Life, ViBRANT, PESI, agINFRA, transplant, …), etc. Other e-infrastructures initiatives were launched cutting-across disciplinary domains: OpenAIRE, providing a participatory network of Open Access repositories at European scale and the EUDAT initiative filling the gap between user-application and generic e-infrastructure layers for high-volume storage, data interoperability, high-performance computing and connectivity layers.

All these initiatives have a common aspect: they are the **biggest research data factories** of the present and the future. Some are led by large research infrastructures undertakings (ESA, EBI, ESO, CERN, EMBL,…) and others by collaborative undertakings of e-infrastructure service providers (university and national libraries, data-centers, super-computing centers, etc).

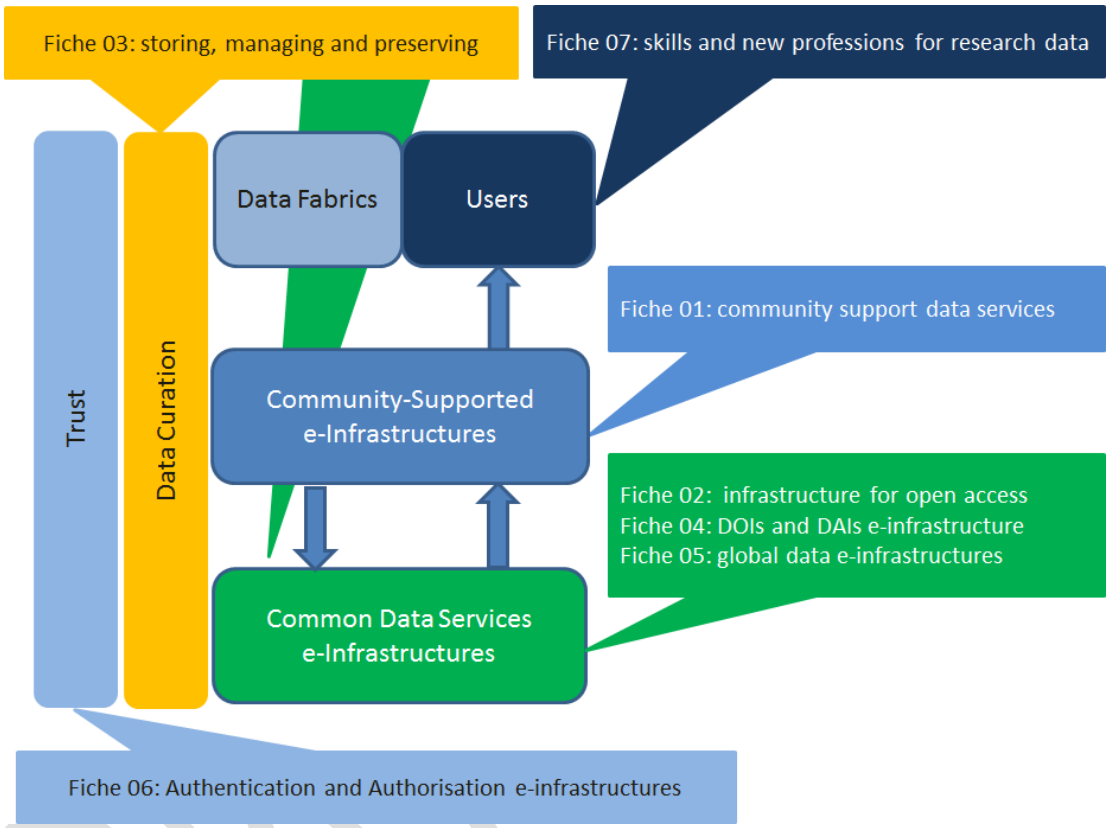


capacity building lifecycle

(empty)

e-Infrastructure Action Fiches

We present here a set of "Fiches" that translate a broad framework, inspired in extensive consultations, into concrete domains of action. They aim at the fulfilment of the European vision for a global interoperable data e-infrastructure supporting open, digital-driven, science. In the figure below the fiches are mapped against the framework structure suggested by the High-Level Group on Scientific Data.



"action fiches" outlook

The fiches address a variety of challenges such metadata and semantics for discovery, understanding and operation of virtual environments where researchers can find and confidently use computing resources and software to mash-up data from different sources exploring multidisciplinary knowledge (**fiche01**); opening access to knowledge through reliable, distributed and participatory data e-infrastructures (**02**); cost effective infrastructures for preservation and curation for re-use of data that may be unique or very costly to replicate and coping with exascale volumes and complexity (**fiche03**); e-infrastructures for persistent availability of information and linking people and data through flexible and robust digital identifiers (**fiche04**); establish the interoperability layer for consistency of approaches on global data exchange (in domains like climate modelling, biodiversity, genomics, astronomy, high-energy physics, social sciences, etc) (**fiche05**); enabling trust through Authentication and Authorisation platforms that scholars and researchers can use to implement the Open Access policies taking into account privacy or other type of restrictions (**fiche06**); developing skills of users and producers of data to ensure a reliable "data life-cycle" (**fiche07**).

(empty)

e-Infrastructure fiche 01	community support data services
Scenario: Except for computer scientists an unfortunate legacy of computer history is that it was not really invented for them. It began with the military, and was developed for business. For instance, the standard, relational database was written to store relatively simple information about customers, not elementary physical particles or protein folding information. If researchers now want to share and store their data more easily, they will need better databases and ways of inter-operating them. They will also need new ways of handling metadata - the descriptive information about the various scientific attributes being studied, that can vary at present from one field or lab to another; how to translate them? There has been significant investment in semantic technologies. It is now the time to start deploying reliable services based on these powerful technologies in e-infrastructures, to help scientists know where another's data came from - its provenance – and to develop trusted collaborations.	
Actions address: Deploying community-driven research data e-infrastructures addressing societal challenges (e.g. Our Planet/Climate, Our Environment, Our Species, Our Oceans, Our Universe, Our Health, Our Education and Our Information) that define broad range of metadata and research data semantics requirements associated with the e-infrastructure services offered to the scientific communities (i.e. access, storage, discovery, analytics, preservation, mash-up of interdisciplinary data, …). The initiatives should include a balance set of competences on the different scientific domains and on e-Infrastructure technologies; the activities range from proof of concept to prototyping services that can assist researchers and educators in their activities by automation of processes, etc. The initiatives should bring together user communities to define the semantics, ontologies, etc (the so called "what" metadata) and the ICT experts to define the best computing models and level of abstraction to process the rich semantics at machine level (the so called "how" metadata); it is expected that some disciplines in particular will "push the envelope" technically. The community-led data e-infrastructures initiatives should also support proof of concept and service prototyping of advanced services for virtual research and education environments. They should provide the "researcher, educator, student, librarian" a toolset and desktop with easy to use functionalities and access to top-of-the range connectivity and computing e-infrastructures.	
Stakeholders	scientific communities as driving the requirements for research data services, data centres, computer centres, IT experts from research infrastructure facilities, IT labs in universities, ...

e-Infrastructure fiche 02	infrastructure for Open Access
Scenario: Europe needs a robust data e-Infrastructure supporting Open Access policies, including that of the Commission for Horizon 2020; researchers, educators and students are both producers and consumers of scientific information; the producers of data, both at individual and institutional level, should benefit from opening metadata and data to broad access so that they prefer to deposit their data with confidence in reliable repositories than to keep it in their own closed systems. A key element will be the capacity building to link literature and data to contribute to more transparent evaluation of research and reproducibility of results.	
Actions address: Supporting *service-driven data e-infrastructures* responding to general and specific requirements of researchers and research organisations for open access to scientific information (publications and data). The service-driven data e-infrastructures will further develop the research capacity through a coordinated and participatory architecture linking institutional and thematic repositories across Europe with scientific information to be used by humans and machines. Developing proof of concept and prototyping new services to link literature and data with technologies that can assist researchers and educators in everyday tasks while working in their lab or being engaged in remote on-line collaborations. Supporting the global interoperability of open access data e-infrastructures and linking with similar platforms across the globe in order to complement the physical access to unique research facilities with data access and to ensure that Europe plays a leading role in the international collaborations.	
Stakeholders	Libraries, IT labs in universities scientific communities as users,, computer centres, data centres, IT experts from research infrastructure facilities, …

e-Infrastructure fiche 03	storing, managing and preserving research data

Scenario: Diversity is likely to remain a dominant feature of research data – diversity of formats, types, vocabularies, and computational requirements – but also of the people and communities that generate and use the data. Europe needs to develop an integrated and service-driven approach to e-infrastructures for the data of a wide range of research communities. Europe should therefore step-up the available capacity to cope with extremely large, heterogeneous and complex datasets incorporating advanced computing and software. Furthermore, costs of storing and preserving data can be significant if one needs to keep it well managed for long periods to be used by others. Different institutions archive their research data in different ways - making access difficult from outside the institution - whereas storage and computing media evolve and become quickly obsolete. How will we preserve and maintain future access to priceless research data? Data management plans are intended to help researchers, funders and data repositories to get maximum value from research data at minimum cost.

Actions address:

Deploying *service-driven data e-infrastructure* to provide a tiered architecture for data management and long term preservation accommodating the need for replication and the reusability of information in different research and education contexts. To achieve cost efficiency in preservation infrastructures - tier zero will be developed by a limited number of data centres operating at the European level, associating the function of traditional storage preservation of information with the advanced curation of information for re-use, complemented by national, institutional or thematic networks of data centres.

Services to ensure the quality and reliability of the infrastructure, including through the use of certification mechanisms for repositories; e-infrastructure for certification services to test and benchmark capabilities in terms of resilience and service continuity of the infrastructure; and measures to ensure high levels of trust on the scientific content produced.

Supporting, in a coordinated way, Data Management Plan (DMP) tools and services used to plan the data lifecycle, identify open data sources and data collected with sensitive or restricted access features. The e-infrastructure should have an open architecture to federate data management services available on a wide institution-based network.

Deploying services to maintain and persistently link scientific software, models and algorithms that embed valuable information and knowledge. This includes the software and models that were used for generating, processing and correlating data so that the reproducibility and accuracy of the data can be verified.

Proof of concept and prototypes of databases and associated services for extremely large or highly heterogeneous data sets. Clean slate approaches to data management targeting 2020+ "data factories" requirements of research communities or facilities are encouraged.

Stakeholders	computer centres, data centres, IT experts from research infrastructure facilities, IT labs in universities scientific communities as users, libraries…

e-Infrastructure fiche 04	discovery and provenance of research data

Scenario: In a fully distributed, decentralized and heterogeneous environment of data production and consumption, digital identifiers (DIs) are expected to provide global keys for information access, and as such form a critical enabling infrastructure. Just as physical networks connect physical machines, digital identifiers connect the digital objects (data and software data producers), and those that work with the data. DIs have the potential to connect contents to authors and authors to their institutions, research projects and potentially any other relevant entity in the production value chain of scholarly content. In this respect, DIs could tighten the threads of a network of connections within the research ecosystem, opening new prospects for advanced information services for science and education. The successful implementation and sustainability of DIs is essential for an open interoperable data e-infrastructure.

Actions address:

Developing a Digital Identifier e-infrastructure (for digital objects, datasets, etc and authors) which cuts across geographical, temporal, disciplinary, cultural, organizational and technological boundaries, without relying on a centralized system; promoting an e-infrastructure for DIs that federates locally operated systems, each based on a single local authority, to a system where these authorities work in coordination to support the interchange and entire lifecycle of DIs. The e-Infrastructure initiatives should aim at systems and service interoperability through open standards and best practices maximising connectivity of the overall e-infrastructure for data. The solutions should respect social, organizational and technical requirements set in the agreed policies.

Stakeholders	Information managers in IT labs in universities scientific communities computer centres, data centres, IT experts from research infrastructure facilities,…

e-Infrastructure fiche 05	towards global data e-infrastructures
Scenario: Since 2010, a consensus has started to emerge among the countries leading the world's scientific production on how to exchange data without limiting flexibility or hampering competition. A forum is being established under the name of Research Data Alliance (RDA), modelled partly on a technical committee, the Internet Engineering Task Force that successfully guided the Internet's early evolution. The RDA follows from reference documents that informed the Commission position namely the *Riding the Wave* report and the report of the *G8+5 working group on data e-Infrastructures*. This new group would be guided by a council of experts, while embracing all stakeholders: research groups, institutions, non-governmental organizations, industry and citizens. Its main work would be organizing debate to create consensus on solutions to specific problems hampering data exchange and interoperability. It would operate bottom-up, building on existing projects and resources. In essence, it could be a clearinghouse for standards, ideas, projects and other initiatives to make global scientific e-infrastructure a reality.	
Actions address:	
Supporting the definition and monitoring of the governance structure of Research Data Alliance (RDA). The supporting project should ensure that RDA serves its main aim, namely to foster research data interoperability at global level; that RDA activities and deliverables reach out to industry and promote innovation; and that there is coordination with international organisations dealing with standardisation, research data and education issues (IETF, W3C, CODATA, OECD, UNESCO, …);	
Stakeholders	Scientific data producers, users, technology providers, funders of e-infrastructures, computer centres, data centres, IT experts from research infrastructure facilities, IT labs in universities scientific communities…,.

e-Infrastructure fiche 06	Authentication and Authorisation e-infrastructures

Scenario: Authentication and Authorisation Infrastructures (AAIs) play a crucial role in creating trust in distributed research virtual environments where scientific resources can be stored, accessed and shared. A European-wide single sign-on service will enable researchers to share their work within a secure and trusted environment while maintaining seamless access to their entitlements. For this, managing the authorisation granularity is essential. With such a service, authorisation could be granted to a secure personal environment for an individual's own research, to secure spaces for sharing within a closed research group, or more widely to large research communities. As more, often national, AAI infrastructures are developed the technical challenges of making them work together in a seamless and user-friendly way will need to be addressed.

Actions address:

Deployment of a pan-European identity federation for researchers, educators and students, in compliance with the EU's eID. Supporting a medium size pilot involving institutions in all EU Member States that are ready to move to define concrete prototype services with federated identities in the context of access to scientific data and scholarly communication resources. In a second phase the pilot would become a permanent service supporting the vision of mobility of researchers in the ERA.

Stakeholders	funding authorities in EU Member States, universities, Research Infrastructures and facilities, IT labs in universities specialized in distributed computing (grid-like) scientific communities as users, libraries computer centres, data centres, IT experts from research infrastructure facilities;

e-Infrastructure fiche 07	skills and new professions for research data
Scenario: In a fast developing data-intensive world of scientific and scholarly research, what kind of skills are needed for creating, handling, manipulating, analysing and storing for re-use of large amounts of data by others? Some researchers in data-intensive research areas have acquired considerable skills in handling and managing data themselves or have a colleague who has these skills, but in many cases researchers turn to the institutional IT services or library for assistance and advice. Current data scientists usually end up in their roles accidentally as formal education hardly exists.	
Actions address: Defining or updating university curricula and sharing best practices across Europe; developing training programmes for data scientists working as part of a team of researchers or in close collaboration with them as responsible for computing facilities, storage and access; developing training programmes for data librarians from the library community who are specialised in the curation, preservation and archiving of data.	
Stakeholders	e-Infrastructure providers; software designers education tools designers, associations such as LIBER, LERU, EAU, ScienceEurope, Universities, Libraries,…

References

1. Richard L. Hudson, Carlos Morais Pires: “Open Infrastructures for Open Science - Horizon 2020 consultation report
2. European Commission. ‘Riding the Wave: How Europe can gain from the rising tide of scientific data.’ October 2010. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf
3. Report of the G8+O5 working group on scientific data; Cape Town (November 2011) and Hamburg (April 2012)
4. Open Science for the 21st century - declaration of ALL European Academies (ALLEA) presented at a special session with the Vice-President of the European Commission, and Commissioner in charge of the Digital Agenda, Mme Neelie Kroes on occasion of the ALLEA General Assembly held at Accademia dei Lincei, Rome, on 11-12 April 2012
5. Knowledge Exchange. ‘A surfboard for riding the wave: towards a four country action programme on research data.’ November 2011. http://www.knowledge-exchange.info/default.aspx?id=469
6. COM(2012) 401 final, Brussels, 17.7.2012: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Towards better access to scientific information: Boosting the benefits of public investments in research
7. C(2012) 4890 final, Brussels, 17.7.2012: Commission Recommendation on access to and preservation of scientific information
8. Vision on Open Science and Open Education in Europe, February 16, 2012 by the Royal Netherlands Academy of Arts and Sciences (KNAW), the Netherlands Organisation for Scientific Research (NWO) and the Association of Universities in the Netherlands (VSNU)
9. The LERU Roadmap towards Open Access http://www.leru.org/files/publications/LERU_AP8_Open_Access.pdf
10. Beyond Sharing and Re-using:Toward Global Data Networking (aka as the DAITF paper) by Jean-Claude Guédon, Fred Friend and Herbert Van de Sompel, October 2011
11. Alan Blatecky, Chris Greer position paper "The DataWeb Forum (DWF)", October 2011
12. Hey, Tony; Stewart Tansley and Kristin Tolle, Eds. “The Fourth Paradigm: Data-Intensive Scientific Discovery.” Microsoft Research. Redmond, Wash: 2009. PDF at http://research.microsoft.com/en-us/collaboration/fourthparadigm/
13. OpenAIRE report on lessons learned from the FP7 OA pilot
14. Norber Lossau, An overview about Research Infrastructures in Europe –recommendations to LIBER
15. Pollack, Andrew. “DNA Sequencing Caught in Deluge of Data.” The New York Times, Nov. 30, 2011. http://www.nytimes.com/2011/12/01/business/dna-sequencing-caught-in-deluge-of-data.html?pagewanted=all
16. SKA Program Development Office. “The Square Kilometre Array: Factsheet for Scientists and Engineers.” April 2010. http://www.skatelescope.org/PDF/100420_SKA_Factsheet-Scientists-Engineers.pdf
17. Thanos, Costantino, GRDI2020. ‘Global Research Data Infrastructures: Final roadmap.’ March 2012. www.grdi2020.eu.
18. St. Vincent Millay, Edna. Huntsman, What Quarry? Harper & Bros.: New York. 1939.
19. Stallings, William. High-Speed Networks: TCP/IP and ATM Design Principles. Prentice Hall, Upper Saddle River, N.J. 1998.
20. Desmond, Adrian and James Moore. Darwin. Penguin Books, London. 1992.

21. Benson, Dennis A., Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and Eric W. Sayers. “GenBank,” Nucleic Acids Res. 2011 January; 39(Database issue): D32–D37. 10 November 2010. DOI:10.1093/nar/gkq1079.
22. Karklins, Janis (Assistant Director-General for UNESCO’s Communication and Information Sector), at ‘Open Infrastructures for Open Science’ conference of the European Commission, 11-12 April 2012, Rome.
23. www.galaxyzoo.org.
24. Morais Pires, Carlos. ‘Scientific Data e-Infrastructures in the European Capacities Programme’, European Commission.
25. Max Planck Institute for Psycholinguistics. “The Language Archive - Data Archive.” http://www.mpi.nl/research/research-projects/the-language-archive/data-archive.
26. CERN. “Worldwide LHC Computing Grid.” http://public.web.cern.ch/public/en/lhc/Computing-en.html.
27. European Commission. ‘Knowledge without Borders: GÉANT 2020 as the European Communications Commons.’ October 2011. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/geg-report.pdf
28. Winkler-Nees, Stefan (Programme Officer, Deutsche Forschungsgemeinschaft), at ‘Open Infrastructures for Open Science’ conference of the European Commission, 11-12 April 2012, Rome.
29. Knowledge, Networks and Nations: Global scientific collaboration in the 21st century The Royal Society 2011
30. Science as an open enterprise , The Royal Society Science Policy Centre, June 2012
31. Clear Green And Pure Gold: Open routes to open access for EU-funded research outputs, by Dr. Fred Friend presentation at the ALLEA General Assembly, April 2012, Rome
32. Surfer sur le tsunami de données de recherché, Le Monde, Science&Techno Samedi 9 Juin 2012
33. University of Trento et al. – DIGOIDUNA: Digital Object Identifiers and Unique Authors Identifiers to enable services for data quality assessment, provenance and access, December 2011
34. Terena, LIBER et al– "Study on Authentication, Authorization and Accounting (AAA) Platforms For Scientific data/information Resources in Europe", October 2012
35. A. Swan, S. Brown; The skills, role and career structure of data scientist and curators: an assessment of current practice and future needs; JISC, 2008