



Bundesministerium
für Bildung
und Forschung

European Cloud Initiative

Key Issues Paper of the Federal Ministry of
Education and Research

Berlin, March 2016

1. The Data Challenge

Advanced technologies together with data-intensive research are multiplying the volumes of data in all scientific disciplines and forms of research. The processing, analysis and distribution of large data volumes is no longer a specific feature of basic research in data-intensive disciplines like physics, environmental sciences or astronomy but has become an everyday routine in many other areas of science.

At the same time, research into key technologies in particular is continuously providing new opportunities for the acquisition, storage, archiving, analysis, reproducibility and distribution of digital data.

Data and the knowledge they generate are increasingly becoming a factor for success in scientific competition and for the societal and economic benefit of research.

We are aware that the falling cost and widespread use of sensors and mobile equipment, high-performance computer systems, available software and modelling platforms and complex ways of online communication are facilitating scientific and technical cooperation across national borders.

These newly emerging technological possibilities and social conventions are ensuring the availability of as yet unused research data and results which generate new scientific knowledge or validate existing results. This applies in particular to joint international research projects ranging from the flagships of global research infrastructures to major research initiatives such as the Human Brain Project. Societal challenges like climate change are increasingly being addressed in transnational projects.

New findings and major innovations can be generated by linking as yet unlinked data. Suitable data management is indispensable for fully exploiting the potential of digitalization. This requires an information infrastructure which is adapted to the life cycle of data, that is to say which stores large amounts of data over very long periods of time, adapts to new data formats and changed search algorithms and ensures that state-of-the-art systems are able to deal with data stored in older formats.

What conclusions must we draw from this development? What material and political conditions are needed to ensure the provision of high-performance information infrastructures?

How must national and European measures be interlinked to create optimum conditions for data-based transborder cooperation in research and innovation?

In Germany, the Joint Science Conference (GWK) entrusted the Council for Scientific Information Infrastructures with drafting recommendations for these issues. Many institutions such as the *Deutsche Forschungsgemeinschaft*, the German Rectors' Conference, the Helmholtz Association (HDF: Helmholtz Data Federation) and Fraunhofer (Industrial Data Space) as well as some *Länder* (including Baden-Württemberg, Bavaria, Schleswig-Holstein) are currently developing solutions in their own fields of responsibility.

The European Commission is planning to issue a Communication on the European Cloud Initiative this spring which will also cover the development of a European Open Science Cloud. We want to make an early contribution to this process in order to ensure that the interests of users in science and research are adequately taken into account in the Commission's considerations.

2. Success Factors

The following factors are of key importance in considerations concerning the design, organization, implementation and sustainability of the European Cloud Initiative:

a. Orientation to user interests and needs of research stakeholders

The central point of reference for all considerations concerning the design of a European digital knowledge infrastructure and particularly of a European Open Science Cloud are the different interests and needs of users in the various science disciplines and forms of research. This includes consideration of the requirements to be met in terms of storage, processing capacities, data organization and quality management in the acquisition, archiving and dissemination process. It also includes the question of which services the infrastructure should offer and to what extent the data can and should be linked efficiently across disciplinary borders.

Researchers can already use a broad range of instruments and services as well as model applications to record, process, analyse, archive and make accessible digital material. Scientific value creation from research data and the research information gained therefrom involves the key challenge of making such data not only available for the research for which it was originally generated but also ensuring smart links with other data and information. In addition, the data should be accessible and usable across disciplinary, institutional and national borders. In this context it is important that joint (European and international) standards are developed and agreed, for example to describe the data or link data systems. The long-term availability and usability of the stored data must be ensured. Possibilities for funding the required sustainable information infrastructures including viable architectures and software to ensure interoperability must also be considered and guaranteed at an early stage.

b. Career prospects for research data managers

From the point of view of the Federal Ministry of Education and Research, it would make sense to establish a system which allows the dynamic integration of knowledge. This should be done in a cooperative effort involving researchers, information infrastructure, IT and software experts and other experts with practical experience as well as science administrators and policy-makers. This system must not be planned at administrative level but should result from best practice solutions. Sustainable research data management means implementing knowledge infrastructures and ensuring compatibility and avoiding isolated solutions. Research data management must be based on existing practice. This is why experts with practical experience must be involved in the process of establishing a European Open Science Cloud.

The professional handling of research data requires qualified scientific and technical staff whose work focuses on data within and beyond their own disciplines as well as on the storage, organization, archiving, processing, linking and long-term availability of such data. They link the areas of IT infrastructures and research and must be given an opportunity to build a reputation based on infrastructure performance and to identify and pursue new career paths.

c. Improvement of the technical environment for implementing a European digital knowledge infrastructure

Efforts should be made to further develop digital infrastructures for data generation and dissemination and for data storage and analysis with the aim of ensuring ideal conditions for excellent research. Sustainable operation must be guaranteed.

For this purpose, the best technologies available on the market must be used in each case to meet relevant needs. Limiting the focus to European technologies and related industrial policy objectives is not always fully compatible with this research policy objective, particularly in the area of high-performance computing. The initiatives supplying research with basic computing time in internationally competitive top-quality systems (PRACE) and promoting technology development towards European pre-exascale and exascale systems (ETP4HPC, cPPP) must continue to be considered as parallel but complementary initiatives. Linking or even merging these initiatives would not serve the objectives of science policy. Close cooperation between individual partners in PRACE and stakeholders in technology development is possible, however.

d. Development of strategies for use and governance structures

Open access to state-funded publications and research data is to be increased and organized in an optimal way in order to make better use of the potential of data-intensive and networked research activities. It is therefore not only necessary to develop technical infrastructures but also to provide for governance structures which give open access to research data and results in such a way as to ensure that "Open Science" can lead to "Open Innovation" which encourages optimum value creation and societal innovations. These governance structures must be able to operate smoothly and respond flexibly. Consideration must be given to the related costs and any restrictions that may be required due to privacy rights and confidentiality interests as well as other major legal interests. Rules are needed for participation by private stakeholders, for example regarding the conditions for access by private enterprises to a European Open Science Cloud. Suitable conditions must be created for R&D cooperation involving public and private stakeholders.

Efforts must be made to guarantee that research data can be efficiently and reliably accessed, analysed, archived and used from almost anywhere. Innovative processes including new scientific workflows and training opportunities must be developed to support interdisciplinary and transborder cooperation between users. European coordination of these efforts is of key importance to ensure the dynamic release of data for interdisciplinary and transborder research in a fair, effective and cost-efficient way. Any European solution must aim to ensure compatibility with similar global projects from the outset.

e. Integration of existing initiatives and digital knowledge infrastructures

Existing initiatives must be considered and integrated as appropriate following a decentralized approach in the establishment, development and governance of a European digital knowledge infrastructure. In a first step, it is necessary to find out where investments in new infrastructures are required and where linking existing infrastructures is sufficient. We must first of all consider the ESFRI process, including its CESSDA, CLARIN, DARIAH ESSsocial and SHARE initiatives in the humanities and social sciences or EU-Openscreen and EuroBioImaging in the life sciences, but also the use and integration of existing expertise at the DFN communications network for science and research in Germany and at CERN, for example. In addition, other initiatives such as the science-driven international Research Data Alliance or GEOSS in the geosciences and environmental sciences should be considered.

A challenge to be addressed is the development of options for the internationally coordinated funding of these infrastructures to promote stronger cooperation while paying attention to individual funders and national priorities. For this purpose, support should also be given to the work of various initiatives such as the Research Data Alliance. The cooperation already

established between these networks and relevant government initiatives, including the Group of Senior Officials on Research Infrastructures set up by the G7 Science Ministers, is necessary and welcome to ensure compliance with national and international requirements.

The special feature of this effort will be to move away from the great number of current bilateral funding models for specific topics or disciplines and focus on a more general, multilateral and cooperative funding approach. Plans for additional investments to establish and develop the technical infrastructure of a European Open Science Cloud must not be detrimental to the funding of research projects under Horizon 2020. Instead use should be made of other funding sources of the European Union such as the Structural and Investment Funds and the European Fund for Strategic Investments.

Thought must also be given to the ongoing efforts of all European partners for a second phase for PRACE to ensure the excellence-based provision of computing time for European researchers. From the point of view of the Federal Ministry of Education and Research, coordination would only be needed for the European part of the PRACE systems. National parts must remain a matter of national responsibility and governance.

3. Close Cooperation between Commission and Member States

Close cooperation between the European Commission, the Member States and the scientific community is required from the outset as different user interests are to be considered and various ongoing initiatives and existing infrastructures integrated. This applies to both the development of technical solutions and the discussion about suitable governance models. The development of a European Open Science Cloud and the use of the emerging European digital knowledge infrastructure are closely related to the planned European Open Science Agenda. This involves issues of open access to data, evaluation of infrastructures and services and scientific integrity. An early exchange is necessary between the European Research Area Committee or groups focusing on the European Research Area and the Open Science Policy Platform as well as other bodies relevant in this context.