Priority Initiative Digital Information of the

Alliance of German Science Organisations

"Research Data at Your Fingertips"

A Position Paper by the

Research Data Working Group

[February 2015]

I. Vision 2025

"Research data at your fingertips"

Researchers in all disciplines can access all research data quickly and easily in a straightforward process in order to carry out research at the highest level and produce excellent results. They can collaborate with other researchers and securely store their research findings. Research data are made available in a form that enables and facilitates interdisciplinary and international research.

The publication of research data and software enhances the academic reputation. Researchers are supported in the collection, gathering, entry and management of their data.

Easy-to-use digital infrastructures and scientific and technical information specialists support the complete research cycle.

The following statements and recommendations are focused on the topic of research data. In this paper, the Research Data Working Group of the Digital Information Initiative of the Alliance of German Science Organisations describes the current situation and identifies a number of challenges to lay the foundations for the necessary discussion of and reflection on the future handling of research data:
http://www.allianzinitiative.de/en/core-activities/research-data.html

Authors: Michael Franke, Stefan Heinzel, Reiner Mauer, Dr. Janna Neumann, Dr. Heike Neuroth, Dr. Hans Pfeiffenberger, Henriette Senst, Andrea Siegberg, Dagmar Sitek, Claus Spiecker, Prof. Dr. Joachim Wambsganß, Dr. Danny Weber, Dr. Stefan Winkler-Nees

For the online version of this publication see:
http://doi.org/10.2312/allianzfd.002

II. Current Situation: Initial Findings

The field of research data is extremely dynamic and diverse. In a description of the current situation, the following key aspects must be taken into account:

1. Disciplines and structural policy

All stakeholders are confronted by a broad, complex range of tasks, many of which are still unclear and not yet specified. Specific responsibilities, for example related to the long-term archiving of research data, are often not known or have not been designated. There are two particularly important questions: the distinction between discipline-specific and cross-disciplinary tasks and whether these tasks should be bundled at the national or even international level. Some successful initial approaches have been made, for example cross-disciplinary data infrastructures[1] and international research data infrastructures[2] for specific disciplines.

In Germany, current organisational structures are strongly influenced by the federal structure and the heterogeneous landscape of research organisations, which operate to some extent on a (scientific) competitive basis. This results in overlapping or mutually incompatible developments.

Nevertheless, important progress has been made in a number of areas, for example the introduction of services that issue persistent identifiers to allow research data to be permanently referenced and cited, and the introduction of university advisory services to provide support with research data management. These are valuable building blocks in the development of research data infrastructures.

2. Awareness of the professional handling of research data

In some disciplines, for example the geosciences, environmental sciences and social sciences, there is already a strong awareness of how to handle research data and professional data management is employed with the use of suitable tools. For the most part, a subject-specific culture of long-term data management has emerged in line with good scientific practice.

At the same time, the reuse of research data has generated valuable results in the scientific knowledge process. In some areas the growing availability of data also allows research data to be reused for socially relevant questions outside the domain of research. For example, epidemiology data have provided vital information about the spread of diseases, their risk potential and suitable methods of prevention.

Data generated in non-scientific contexts are becoming increasingly important in scientific research. This includes public sector information from a legal or administrative context as well as data from social networks, the healthcare sector and citizen science initiatives.

In established, publicly funded information institutions[3] there are different degrees of awareness with regard to the handling of research data, but also an emerging willingness to engage with this issue. Many commercial companies[4] have also recognised the economic value of data.

---

[1] European Data Infrastructure (EUDAT), http://www.eudat.eu/ (02.02.2015)
[2] E.g. EMBL-EBI, http://www.ebi.ac.uk/ (02.02.2015) and ELIXIR, http://www.elixir-europe.org/ (02.02.2015); ESFRI-ERICs such as CESSDA, CLARIN, DARIAH, SHARE, etc.
[3] Archives, libraries, museums, statistical offices etc.
[4] So-called big data firms, research-driven companies, publishing houses, etc., as well as business consultancies [McKinsey Global Institute: Big data: The next frontier for innovation, competition, and productivity, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (02.02.2015)]

3. Research organisations

Many research organisations have begun to address the issue of handling research data but mostly implement it in a heterogeneous, uncoordinated way. The Alliance's Research Data Working Group therefore developed the ***Principles for the Handling of Research Data***, which were approved by the executive boards of all Alliance organisations in 2010[5].

However, a number of items still remain unclear, mainly the funding of suitable research data infrastructures. Funding organisations have recognised this need and responded by developing a range of funding measures. All these measures apply only within the framework of time-limited projects and do not therefore permit the necessary long-term operation of infrastructures. At the same time, the various funding opportunities are not coordinated between funding organisations.

Research funding providers must recognise that data management is an essential component of research projects and needs financial support.

An essential prerequisite for the handling of research data are accompanying sets of rules (policies) and formal requirements, such as those currently being tested in the pilot project for open research data in Horizon 2020, which includes a requirement to use data management plans and to share research data after the end of the project.

The successful implementation of professional research data handling also depends on political support. Some positive first steps have already been taken, for example the start of coordination efforts between the federal and state governments (the Joint Science Conference (GWK[6])) and activities at an international level[7]. However, an institutionalisation of research data infrastructures is still a far-off goal. In many areas the legal framework, for example, remains to be clarified.

Information specialists and information service providers are indispensable to scientific data management. In current practice these tasks are mainly performed by people trained in other specialisations; there is no systematic establishment of new professional fields.

4. Operating in the international arena

Research data are not restricted by national borders. Numerous researchers are involved in international collaborations, which builds up both their own knowledge and their professional reputations. This must be supported by appropriate international infrastructures. Some initial positive approaches can already be seen[8], but face numerous challenges. For example, the availability of long-term funding models is patchy[9].

---

[5] Alliance of German Science Organisations - Principles for the Handling of Research Data, http://www.allianzinitiative.de/en/core-activities/research-data/principles.html (02.02.2015)
[6] Cf. Rat für Informationsinfrastrukturen (Council for Information Infrastructures), http://www.gwk-bonn.de/index.php?id=205(available in German only) (02.02.2015)
[7] E.g. G8+O5 Global Research Infrastructure, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=international_level (02.02.2015); Research Data Alliance (RDA), https://rd-alliance.org/ (02.02.2015)
[8] European Strategy Forum on Research Infrastructures (ESFRI), http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri (02.02.2015); Science Europe, http://www.scienceeurope.org/ (02.02.2015)
[9] For example, a unique legal form was developed for the ESFRI projects: European Research Infrastructure Consortium (ERIC), http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric3 (02.02.2015)

III. Recommendations

The implementation of the following recommendations will help to provide an adequate research data infrastructure for researchers in Germany in the near future. This is dependent on the scientific recognition of the publication of research data for re-use (in a citable and accessible form) and the consideration of this factor in the evaluation of individuals and institutions.

1. Nationwide discipline-specific coordinating committees are needed to gauge requirements, identify the necessary infrastructure components, promote their development and expansion, and define relevant policies and standards. Ideally, these committees should have members representing academia, government, research funding and infrastructures. The German Data Forum (RatSWD)[10] may be cited as a good model.

2. As with libraries and their expertise in the provision of literature, the long-term provision of research data requires the establishment of trustworthy, legally binding organisational structures. Ideally, specialist research data centres should be established and developed.

3. Stable long-term funding models should be developed and implemented to cover development and investment costs and ongoing operating costs. An appropriate level of basic funding for infrastructure is a key component.

4. Researchers need a clear, unambiguous legal framework for the use and publication of research data. For example, clear rules are needed as to the legal scientific use of personal data and copyrighted or licensed material.

5. To allow global and networked research, technical, semantic and syntactic standardisation is required. Emphasis should be given to international coordination and implementation, with successful examples being DFN/GEANT[11] (network infrastructure), INSPIRE[12] (geo-referenced data) and DDI[13] (social sciences).

6. The digital transformation process calls for additional researcher training. New professions and career paths should be established through the creation of more specific curricula for information specialists.

~~~

[10] German Data Forum (RatSWD), http://www.ratswd.de/en/ (02.02.2015)
[11] German National Research and Education Network (DFN) https://www.dfn.de/en/ (02.02.2015) and Pan-European Data Network, http://www.geant.net/ (02.02.2015)
[12] Infrastructure for Spatial Information in the European Community, http://inspire.ec.europa.eu/ (02.02.2015)
[13] Data Documentation Initiative, http://www.ddialliance.org/ (02.02.2015)
~~~