Knowledge Exchange Discussion Workshop

Main Drivers for Successful Re-use of Research Data

Berlin, September 23–24, 2009
Europäisches Haus / Hotel Park Inn Berlin-Alexanderplatz

Organised by the Knowledge Exchange Initiative and the Knowledge Exchange Working Group on Primary Research Data

Programme Committee: John Doove (SURFfoundation), Simon Hodson (JISC), Christoph Kümmel (DFG), Laurents Sesink (DANS)

Preliminary Report
by Christoph Kümmel

Contents

Abstract
1 Background and goal of the workshop
2 Programme structure, keynotes and participation
3 Summary of general aspects addressed in the presentations and discussions
4 Impact and follow-up activities

Abstract | On 23-24 September 2009 an international discussion workshop on "Main Drivers for Successful Re-Use of Research Data" was held in Berlin, prepared and organised by the Knowledge Exchange Working Group on Primary Research Data. The main focus of the workshop was on the benefits, challenges and obstacles of re-using data from a researcher’s perspective. The use cases presented by researchers from a variety of disciplines (13 presentations) were supplemented by two keynotes and selected presentations by specialists from infrastructure institutions, publishers, and funding bodies (national and European level, 8 presentations). By choosing this design the workshop was able to provide a critical evaluation of what lessons have been learned concerning sharing and re-using research data from a researcher’s perspective and what actions might be taken to encourage and facilitate more successful re-use. Despite the individual differences characterising the diverse disciplines, it became clear that important issues are comparable. Apart from several technical challenges such as metadata exchange standards and quality assurance it was obvious that the most important obstacles to re-using research data more efficiently are socially determined. It was agreed that in order to overcome this problem more effort should be made 1) to raise awareness and 2) to encourage stakeholders to combine forces in order to support and stimulate sharing and re-use of research data on all levels (researchers, institutions, publishers, funders, governments).
1 Background and goal of the workshop

Today, there is increasing recognition that the organisations and individuals involved in publicly-funded research have a responsibility to ensure the availability of the outputs of that research, both to underpin the scholarly record, and to enable such outputs to be the basis for future research. In such a context, it is timely to ask how researchers are already benefiting from this development in practice. What are their experiences by re-using data and what are the main drivers for success in that context?

The discussion workshop was planned to bring together successful examples from a whole range of scientific disciplines in order to make a case for the particular benefits of re-using data from a researcher’s perspective. But the contributions and following discussions were not only intended to show the potential data-sharing may have in practice but also to demonstrate research-specific challenges and obstacles that still hinder the most effective re-use of research data. The use cases presented by researchers were supplemented by selected presentations by specialists from data centres, information management divisions, publishers and funding bodies in order to give insights in what has been achieved so far and what still has to be done to provide the appropriate infrastructure and legal framework for successful re-use of data.

In this way the workshop was designed to show based on practical experiences from researchers, what lessons have been learned concerning the sharing and re-use of research data and what actions might be taken by the different stakeholders to improve the conditions for a successful re-use of research data.

2 Programme structure, keynotes and participation

The two-day programme was structured in three main sections: (i) the researchers’ perspective on the benefits and challenges of re-using data (“data re-use in practice”, 13 presentations); (ii) contributions by infrastructure institutions on the current state of data related scientific infrastructure (“preparing the grounds for re-use”, 4 presentations); and, (iii) statements given by national and European funding bodies and an international publishing company (Nature Publishing Group) on the opportunities to improve the re-use of research data by policy regulations and funding schemes (“Induce re-use”, 4 presentations). On each day the programme started with a keynote presentation.

Ulf Leser from Humboldt University (Institute for Computational Sciences), kicked off the first day with his general considerations about “Social Issues in Scientific Data Exchange” strongly stressing the importance of the mental, organisational and methodological pre-conditions for successful re-use of scientific data. On Day Two, John Wood (Imperial College London) gave an overview of “The Future of Integrated Research” from a broad European perspective. He most convincingly argued that “you can no longer separate the data management issue from the planning and running of research infrastructures”. Concluding remarks were given by Liz Lyon (UKOLN, Digital Curation Centre) who appealed to all participants to join forces to raise awareness and to foster collaborative activities across borders.

About 65 invited participants from institutions in the partner countries (Denmark, the Netherlands, the United Kingdom, Germany), in France, the United States and from European institutions (European Science Foundation, European Commission) joined the workshop (see workshop programme for details) and actively discussed the issues that arose from the presentations.
3 Summary of general aspects addressed in the presentations and discussions

The presentations in section one covered a wide range of research areas (humanities, sciences, social sciences, explicit interdisciplinary approaches) and presented a variety of use cases showing what lessons have already been learned. Despite the individual differences characterising the diverse disciplines it became clear that certain challenges and obstacles are comparable on a more general level.

First of all it was repeatedly mentioned that basic technical questions are in principle solved but still need to be improved in practice. This applies above all to standards for metadata and metadata exchange but also to quality controlled repositories in general. But it also became clear that the most significant challenges to successfully running a data repository, organising a data service unit and ensuring the utmost compatibility of datasets are by no means technical only. Almost every presentation mentioned methodological implications such as harmonised ontologies and approved and acknowledged classification systems as well as other more disciplinary specific methodological problems to be solved in the respective research communities.

This shows that the most severe challenges for the successful re-use of research data are directly connected to the social background of research. As argued convincingly in the first of the keynote presentations by Ulf Leser, the willingness to share research data is primarily a social matter, a sharing mentality. And, of course, increased opportunities to re-use data can only be generated by encouraging such a cultural shift and by creating particular incentives for single researchers as well as the research community as a whole. Hence it was recommended to give further thought to:

- The education of researchers regarding data re-use (capacities and capabilities/skills)
- Community building and raising awareness (on all levels: government, management, researchers, publishers)
- Developing best-practice guides according to the needs of every discipline
- Promoting research projects’ re-use of data and providing funding opportunities especially designed to improve data re-use
- Providing advanced and long-term funding opportunities for the improvement of data centres and data sharing
- Establishing legal certainty and working towards a legal framework explicitly beneficial to researchers and their interests

This would first of all require incentives for researchers, both to share and re-use data. These incentives can be provided by funding opportunities, but also by incorporating data sharing and re-use in the research assessment. The issue of data re-use would be a lot more relevant to researchers if publishers, for example, asked authors for data to underpin their publications. Support for the technical aspects is important and could be improved by increased funding of data storage and sharing facilities, by instructing and recruiting specialised staff but also by establishing agreed standards (metadata, ontologies) which will allow different datasets to be evaluated, compared and combined. Researchers will need to be trained, but awareness will also need to be created at a much higher level. Bottom-up and top-down approaches need to work in unison.

Given the challenges and hindrances that are still to be overcome if we are to benefit on a broader scale from sharing and re-using data, the participants of the workshop agreed on the need for an intensified engagement of the stakeholders: the researchers and research
institutions themselves but also policy makers, such as national and European funding and governmental bodies. As a general recommendation the participants agreed that it is time to cooperate in more ambitious international activities to establish reliable and sustainable support for initiatives in the field of data related research infrastructure.

4 Impact and follow-up activities

From the reactions of the participants it became clear that the discussion workshop was a welcome event which allowed the timely exchange of ideas about the most urgent challenges to be overcome in order to unlock the potential of shared and re-used research data. The Knowledge Exchange initiative will publish a workshop report including the abstracts of the presentations and a summary of the discussions and conclusions as soon as possible on the internet (www.knowledge-exchange.info) and in printed form. The Knowledge Exchange Working Group on Primary Research Data will, in its next meeting, evaluate the outcomes and prepare a position paper with recommendations for future international engagement and potential activities on an international level.

Christoph Kümmel
2009 November 12
(christoph.kuemmel@dfg.de)

Special thanks for contributions go to Simon Hodson, Hans Pfeiffenberger, Keith Russell and Stefan Winkler-Nees.