November 2021

Rat für Informations Infrastrukturen

Towards Common Scientific Data Infrastructures in Europe: Which Road To Take?

REPORT ON AN INTERNATIONAL HIGH LEVEL EXPERT WORKSHOP OF THE GERMAN COUNCIL FOR SCIENTIFIC INFORMATION INFRASTRUCTURES (RAT FÜR INFORMATIONSINFRASTRUKTUREN – RFII)

On May 21, 2021, the working group on international monitoring of the German Council for Scientific Information Infrastructures (RfII) organised an expert workshop to compare and discuss the ongoing developments regarding Open Science as well as e-infrastructures and data infrastructures in the **Netherlands**, **Great Britain**, **France** and **Germany**.[1] After his words of welcome, councillor and chair of the working group (WG) Klaus Tochtermann outlined the WG's activities and introduced the objectives of the High Level Expert Meeting. In 2017, the RfII published a first international comparison on research data infrastructures and policies related to scientific data infrastructures with the aim to accompany the initial design phase of the *National Research Data Infrastructure* (NFDI) in Germany. In the meantime, the NFDI has been launched and the RfII has been turning to new topics, especially in the field of Open Science in Europe and at different subnational levels. The workshop addressed the current state of affairs of the national research and innovation landscapes of the Netherlands, Great Britain, France and Germany in terms of the different scientific systems, institutional policies and recent roadmaps in the area of Open Science and research infrastructures. The overarching goal was to identify common challenges, shared positions and proven practices – also with regard to policies for the evolving ecosystem of the *European Open Science Cloud* (EOSC). The invited experts and the RfII WG members discussed key questions and strategic priorities in three different sessions.

PANEL 1: NATIONAL PLAYERS, DEVELOPMENTS AND CHALLENGES

The first session started with kick-off talks on the Open Science landscape of each country. **Suzanne Dumouchel** outlined the situation in France, **Karel Luyben** explained the approach to Open Science in the Netherlands, **Matthew Dovey** portrayed Great Britain and **Klaus Tochtermann** introduced the development in Germany. Moreover, **Sarah Jones** (Great Britain) and

[1] For contributors and their institutional affiliations please see the list at the end of the report.

Volker Beckmann (France) as well as Stefan Decker and Ramin Yahyapour from the RfII, i.e. Germany, complemented the picture of the development of Open Science and scientific information infrastructures in the respective countries.

The exchange on the current conditions of digital science and research was also concerned with the political environment of the national scientific systems and the question of how both the scientific and the research (data) infrastructure communities can be addressed and more strongly involved in the joint shaping of Open Science. Furthermore, the session focused on concrete policies for the further progress of Open Science. Though the Netherlands, Great Britain, France and Germany take different ways in digital policy efforts, they share similar goals for the FAIRification and the sharing of research data, for Open Access and Open Data. An important topic of the first session was the question of how data policies between science and industry or publicly funded and private-sector research can be moderated, also with a view to the public large-scale projects EOSC and Gaia-X. In this context, it also became clear that the term "open" can lead to misunderstandings, especially among companies, with regard to a supposed lack of security and trust in data quality and data services, and that more information should be provided on this.

PANEL 2: INTEGRATING NATIONAL AGENDAS AND CONVERGENCES OF DATA ECOSYSTEMS

In view of the establishment of the EOSC, the second session first discussed whether the evolving scientific data infrastructures should be managed by already existing organisations and structures. Here, for example, *SURF* or *GÉANT* came into view, but also the *European Research Infrastructures Consortia* (ERIC) or *European Stakeholder Forum on Research Infrastructures* (ESFRI) cluster projects, which have so far covered mainly the area of e-infrastructures or storage, connectivity and computing.[2] Secondly, the participants discussed the extent to which further or new organisations and independent institutions particularly beyond the EOSC ecosystem are needed in order to further expand the infrastructures for managing data in science and research of the respective countries.

Whatever country the discussion focused on and regardless of talking about the EOSC or the German NFDI, there was agreement that information and access points as well as support structures or offices, help desks and trainings – such as the French (EOSC) "collèges" – are time-critical actions to be taken to introduce scientists from the higher education institutions to research data management, FAIR data practices and the various facets of Open Science. Data Competence Centres and Contact Points such as those in the Netherlands or the Digital Curation Centres in the UK are further effective means of organising and shaping the digital transformation in science and research on a broader scale, as is the installation of trusted intermediaries for data exchange. Especially both the aim to install a countrywide expert network of around 1000 data stewards and the designation of about 40 data stewards in the Netherlands

[2] For the explanation of organisation acronyms please see the brief glossary at the end of the report.

who are in charge of aligning the Dutch national landscape of data services received high attention.

Another focus of the second session was how the harmonisation of the open science landscapes as well as the data ecosystems for science and research in and between the Netherlands, the UK, France and Germany can be organised and which connecting lines between the national data infrastructures are advisable. In addition to this, there was a discussion about which forums and structures exist or are needed (in addition to those of the EOSC) in order to consolidate cooperation between the countries and consultation among transnational organisations in the area of research data and open science. The focus here was on the *Council of Open Science Coordination* (CoNOSC), which is increasingly attracting interest from various European countries. The discussants regretted that up to now Germany has neither a national Open Science coordination nor a single person who represents the countries' landscape at the CoNOSC; it was argued that maybe the NFDI association could take that part. But the experts at the workshop also took the example of *Data Together*, an alliance for coordinating actions and activities towards the goals of CODATA, WDS, RDA and GO FAIR, to discuss how to deal with the increasing number of transnational organisations and initiatives as well as the diversity of stakeholders in a productive and participatory way. It became clear that the division of tasks and joint coordination ought to strengthen mutual support and complementarity in order to avoid duplicate structures both as sensible as possible and as far as doable.

PANEL 3: PROMOTING A COMPREHENSIVE VISION: WHAT NEEDS TO BE DONE?

The third session asked about the current priorities of the transition to openness. One of the topics here was the need for widely available, easily accessible and uncomplicated tools and services for everyday scientific work, both for research data management and for Open Science in all its forms. At the same time, for more complex requirements such as research data management, there is a need for access to specific solutions that is as barrier-free as possible, as well as options for contacting competent staff ("customised help"). As mentioned above, particularly staff and experts like the Dutch data stewards and information infrastructure personnel is critical for a broad roll-out of Open Science – be it within the NFDI, the EOSC or the national data infrastructure landscapes in general.

In order for the transition to openness to succeed in the long term, the workshop participants also felt it would be advisable not only to approach the various research funders specifically about more intensive support for research data management and high-quality data, but also to promote the benefits of Open Data and Open Science. In addition to professional associations, it is organisations such as the *European University Association* (EUA) or the *Conference of European Schools for Advanced Engineering Education and Research* (CESAER), as well as national university associations, whose involvement is urgently needed to promote the cultural turn towards digital science and to reinforce Open Science in the scientific communities in Europe.

CONCLUSION: FOSTERING COLLABORATIONS FOR OPEN DIGITAL STRATEGIES

One of the many insights from the RFII international monitoring WG workshop was that regular exchange such as roundtables, cooperation and collaboration across the EU and Europe should be continued and intensified in order to move towards the goal of a shared data ecosystem for science and research. Another insight was that researchers identify first and foremost with their subject and their scientific communities and less with the research institutions or universities they belong to.

At the national level information infrastructure initiatives in France, the Netherlands, Great Britain and Germany share a lot of features, while they differ in others: all countries are committed to a common data approach in Europe, Open Access and Open Data have become standards in research policy. Indicators and platforms have been developed in all countries mentioned above to assess their progress with regard to open access publications. France and the Netherlands have established their own development plans for the transition to openness. These, like the plans of Great Britain and Germany, are broadly pointing in the same direction: publication systems are being digitally transformed, often in close cooperation with major publishers. Implementing the FAIR criteria may not directly be part of the objectives of Open Science, but it is part of the countries‘ approaches to data.

In terms of international cooperation and how countries align themselves in the development of common cross-national information infrastructures and Open Science policies, it is apparent that the Netherlands and France as well as Germany are heavily involved in the ongoing EOSC processes, not least in terms of governance through the EOSC Board of Directors. Organizations from Great Britain are currently not allowed to participate in the EOSC Association as full members. With the Netherlands and France, two countries without an initiative comparable to the NFDI are involved in European cooperation beyond EOSC. On the other hand, Gaia-X, which is being launched by France and Germany, is foremost a business and industry oriented project rather than a science-oriented one. How Gaia-X and EOSC fit together and find reliable interfaces for data exchange has to be seen.

In general, the expert workshop showed that the vision of “Open Science” is often also interpreted by science policy actors and scientific initiatives without regard for the country they come from. In other words, when it comes to Open Science it seems that academics more intensely refer to their disciplinary background, their scholarly identity, experiences and scientific workflows rather than on national or institutional affiliations. Last but not least different stakeholders and both scientific and infrastructure communities practice a variety of “Openness”-approaches and cultivate their own understandings of Open Science. Likewise, both the strategies and the specific understanding of “openness” within the national policies and coordination efforts of the Netherlands, Great Britain, France, and Germany differ. Hence the different actors of science, research and their underlying data infrastructures are looking for ways to negotiate and shape the digital transformation nationally and internationally in such a way that the project of developing information and data infrastructures, which was initially primarily technology-driven, covers much more aspects than Open Access and Open Data only.

The workshop served as a forum on Open Science strategies and also implementations of Open Science practices in the Netherlands, Great Britain, France and Germany. As an opportunity to talk about the challenges and chances of openness in science and research, the workshop was a productive occasion to discuss how the four countries are contributing to making Open Science the new normal, said host Klaus Tochtermann at the end of the event. To the present, it remains a challenge to transform open sciences into a comprehensive program for the future.

LIST OF WORKSHOP PARTICIPANTS

- Dr. Volker Beckmann, French Ministry of Higher Education, Research and Innovation (MESRI)
- Prof. Dr. Stefan Decker, Fraunhofer FIT, RWTH Aachen & RfII member
- Matthew Dovey, Head of e-infrastructure strategy, Jisc
- Suzanne Dumouchel, CNRS, Huma-Num & Board of Directors EOSC Association
- Dr. Barbara Ebert, German Federation for Biological Data e.V. (GFBio) & guest of the RfII international monitoring WG
- Andrea Herdegen, Federal Ministry of Education and Research
- Dr. Dirk Hommrich, RfII Head Office
- Sarah Jones, GÉANT & Board of Directors EOSC Association
- Prof. Dr. Karel Luyben, Rector Magnificus Emeritus of the Delft University of Technology and President of the Board of Directors EOSC-Association
- Prof. Dr. Otto Rienhoff, Department of Medical Informatics at the University Medical Center Göttingen & guest of the RfII international monitoring WG
- Prof. Dr. Klaus Tochtermann, ZBW – Leibniz Information Centre for Economics, Board of Directors EOSC-Association & RfII member
- Prof. Dr. Joachim Wambsganß, Centre for Astronomy of Heidelberg University & RfII member
- Prof. Dr. Ramin Yahyapour, GWDG Göttingen & RFII member
- Daniel Zdun, RfII Head Office

BRIEF GLOSSARY

CESAER is the *Conference of European Schools for Advanced Engineering Education and Research*. As an European association of over 50 leading research-intensive specialised and comprehensive universities of science and technology in Europe and beyond CESAER was founded in 1990 as not-for-profit association under Belgian law (AISBL). CESAER is an acknowledged stakeholder organisation for the European Research Area, was part of the Open Science Policy Platform (2016 to 2020) and is a founding member of the EOSC association. – https://www.cesaer.org

CODATA is the *Committee on Data of the International Science Council*. It was established in 1966. The Committee helps to realise the International Science Council's vision of advancing science as a global public good by fostering international collaboration to advance Open Science and to improve the availability and usability of data for all areas of research. CODATA runs Task Groups and Working Groups but also supports the Data Science Journal and major data conferences like SciDataCon and the International Data Week. – https://codata.org

CoNOSC, the *Council of Open Science Coordination* is a forum for different actors and European Member States dealing with the Open Science discussion. In autumn 2019 France, the Netherlands and Finland invited representatives of the ERAC[3] countries to discuss the creation of a network for Open Science coordination. Twenty-one countries were present and expressed interest, as well as the European Union. The objectives and organisational principles of the initiative are specified in a joint Memorandum of Understanding which defines three main goals: CoNOSC helps to fill in the gaps in national Open Science coordination, CoNOSC aims to provide valuable information and detailed knowledge about the field of Open Science in Europe through the dialogue with other international partners, and CoNOSC membership is in principle open to all countries within the *European Research Area* (ERA). – https://conosc.org/

Data Together is a relatively new collaboration of the four major international data organisations RDA, CODATA, WDS and GO FAIR. The joint commitment (published in spring 2020) is to optimise the global research data infrastructure and to identify opportunities that will trigger federated information infrastructures to service the new reality of data-driven science. These infrastructures can be defined at a high level as forming a global trusted ecosystem that provides seamless access to high quality interoperable research outputs and services. Addressing the aim to support cross-geographical and cross-disciplinary Open Science, Data Together pays special attention to the idea of Global Open Research Commons (GOSC) and the evolvement of a Global Open Science Cloud (GORC) infrastructure. – https://www.rd-alliance.org/data-together-fostering-cooperation-among-open-science-platforms-march-2021

ERIC is the *European Research Infrastructure Consortium*, a specific legal form that facilitates the establishment and operation of Research Infrastructures with European interest. This has the advantage that new or existing Research Infrastructures can operate on a non-economic basis. – https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/eric_en

ESFRI is the *European Strategy Forum on Research Infrastructures*. It was founded in 2002 by the EU Council of Research Ministers (Competitiveness Council), to coordinate the effort and support the development of a comprehensive European Research Infrastructure policy and corresponding roadmaps. – https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/esfri_en

[3] European Research and Innovation Area Committee

GÉANT, the *Gigabit European Academic Network*, is a membership organisation acting with and for its members to further research and education networking in Europe and globally. GÉANT connects Europe's national research and education networks to each other and to similar networks worldwide. More than 50 million users in over 10,000 research and education institutions around the globe can work together by using GÉANT's e-infrastructure. Through its integrated catalogue of connectivity, collaboration and identity services, it provides 39 national research and education partners and their users with reliable, unconstrained access to computing, analysis, storage, applications and other resources. – https://www.geant.org/

EUA, the *European University Association*, represents more than 800 universities and national rectors' conferences in 48 European countries. EUA plays a crucial role in the Bologna Process and in influencing EU policies on higher education, research and innovation. Through continuous interaction with a range of other European and international organisations, EUA ensures that the independent voice of European universities is heard and provides members with opportunities to shape European policies and initiatives affecting higher education and research. – https://www.eua.eu

GO FAIR is a stakeholder-driven and self-governed initiative that aims to implement the FAIR data principles, making data Findable, Accessible, Interoperable and Reusable (FAIR). It offers an open and inclusive ecosystem for individuals, institutions and organisations working together through Implementation Networks (INs). – https://www.go-fair.org

RDA is an abbreviation for the *Research Data Alliance*, which is an international community-driven initiative building social and technical bridges to enable the open sharing and re-use of research data. RDA was launched by the European Commission, the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation in 2013. RDA serves as a global platform for scientific communities and both e-infrastructure and data infrastructure communities to capture and share good practices and standards for data management, sharing and analysis, and to facilitate the uptake of those goals. Similarly to CODATA, RDA runs different kinds of Interest Groups and Working Groups. – https://www.rd-alliance.org/about-rda and https://www.rd-alliance.org/groups/rda-europe-national-nodes

SURF is a Dutch cooperative in which universities, colleges of higher education, university medical centres, secondary vocational education institutions and research institutions work together to develop information & communication technology (ICT) innovations. More than one hundred institutions from five sectors in the Netherlands are members of SURF. – https://www.surf.nl/en

WDS stands for the *World Data System*, which is an interdisciplinary body of the *International Science Council* established in 2008. WDS objectives are to preserve quality assured scientific data and information, to facilitate open access, and promote the adoption of standards. To this end, WDS coordinates and supports trusted scientific data services for the provision, use

and preservation of relevant datasets and strengthens links with the research community. – https://www.worlddatasystem.org

Imprint

German Council for Scientific Information Infrastructures (RfII) – Head Office
Papendiek 16, 37073 Göttingen, Germany
Phone +49(0)551-392 70 50
E-mail info@rfii.de
URL www.rfii.de

SUGGESTED CITATION

RfII – German Council for Scientific Information Infrastructures: Towards Common Scientific Data Infrastructures in Europe: Which Road to Take? Report on an International High Level Expert Workshop, Göttingen 2021, 8 p.

The German National Library lists this publication in the German National Bibliography; detailed bibliographic data are available on the Internet at urn:nbn:de:101:1-2021090825.