

# German Data Forum welcomes the EOSC declaration and participates in the "coalition of doers" implementing EOSC principles

The German Data Forum<sup>1</sup> welcomes the European Commission's recently issued declaration<sup>2</sup> on the European Open Science Cloud (EOSC). The German Data Forum is committed to establishing an open European research data infrastructure and will continue to support the implementation process of the EOSC principles with its broad, interdisciplinary expertise.

As early as 2001, the Commission for the Improvement of the Informational Infrastructure between Research and Statistics (KVI), which preceded the German Data Forum, noted that "the performance of the data infrastructure is decisive for both the performance of a society and the innovation capability of social and economic research on an international scale."<sup>3</sup> Therefore, the German Data Forum considers it a necessity that European science organisations in the fields of social, behavioural and economic research actively pursue and support the development of a pan-European research data infrastructure.

## Commitment to action

Many aspects mentioned by the EOSC declaration are day-to-day practice at most research data centres (RDC)<sup>4</sup> accredited according to the standards of the German Data Forum. The German Data Forum will continue to advance and improve the research data infrastructure in close cooperation with the RDCs.

**[Data culture]** We are committed to raising awareness of the significance of data sharing, data provision and curation among researchers and funding organisations.

**[Skills]** To advance the recognition of data sharing and the RDC model, it is essential that RDCs provide trustworthy, reliable, and high-quality services. The German Data Forum established a working group to develop guidelines for skill requirements of data managers and data stewards employed at research data centres.

**[Data stewardship]** We are committed to ensuring the creation and provision of user-friendly data products, including high-quality standardised metadata, comprehensive documentation, training materials and support for data analysis.

**[Rewards and incentives]** We promote the academic recognition of data production, data archiving and data sharing as essential services for the scientific community. By anchoring incentives in the scientific reputation system, producers, archivists and sharers of research data receive credit in career assessments and in the evaluation of research projects.

**[FAIR principles]** The German Data Forum is committed to the implementation of the FAIR principles in the 31 accredited research data centres. The German Data Forum promotes the creation of additional RDCs to further improve data access for researchers. The principles of findability, accessibility and reusability have been guiding principles of the informational infrastructure and the establishment of RDCs since 2001. Our monitoring processes support the analysis and communication

of improvements of the research data infrastructure.<sup>5</sup> However, two conflicting priorities have emerged regarding the FAIR principles:

**1) Accessibility: Open Data Access vs. Data Protection**

The structural make-up of the social, behavioural, and economic sciences limits the degree to which data can be “accessible under well-defined conditions”.<sup>6</sup> Access to sensitive data is subject to various legal regulations. Some data stored by research data centres can be made available to researchers only under certain legally required precautions. In such cases, data analysis is only possible using a research data centre’s on-site workstation or secure remote execution without direct access to the microdata. The German Data Forum aims at making the use of data stored at accredited research data centres as flexible as possible in accordance with legal regulations.

**2) Interoperability: Specific disciplinary requirements vs. uniform metadata standards**

The research data centres accredited by the German Data Forum provide most data in formats, which can be processed in any statistics software. Regarding data linkage, different concepts (for example, varying definitions of the unit “household”) have to be considered and render the individual assessment of researchers indispensable.

From the long-standing experience of the RDCs, we know that achieving factual interoperability of the data holdings requires significant investments. In some cases, progress has been slowed down by a lack of long-term funding for harmonisation at the research data centres. Thus, improving interoperability will require additional resources for the development of metadata schemes and tools.

**[Standards]** The German Data Forum will continue to foster communication over technical and organisational standards for data archiving and data access. An outstanding example is the dissemination of definite, persistent identifiers (PID), which are used extensively by most research data centres.

**[FAIR Data governance]** We offer the broad, interdisciplinary experience in data sharing and data governance of eight elected members from social, behavioural, and economic disciplines and eight members of the most important national data producers as well as the expertise of 31 research data centres.

**[Research data repositories]** We promote the model of independent and accredited research data centres (RDCs). RDCs provide science and research with a user-friendly, transparent, trusted and high-quality access to statistically analysable microdata. They play an intermediary role in improving cooperation between the data users and the respective data producers.<sup>4</sup>

**[Accreditation]** We developed minimum standards and clear criteria<sup>7</sup> for accrediting RDCs. These standards and criteria foster coordination and cooperation among RDCs and ensure the quality of the research data infrastructure. As of November 2017, 31 RDCs have been accredited by the German Data Forum. Regular monitoring supports the analysis and communication of improvements of the research data infrastructure.

**[Data Management Plans]** We support researchers in creating research data plans<sup>8</sup> to safeguard sustainable research data use as early as a research project’s planning phase. The German Data Forum provides researchers with transdisciplinary best practice solutions as well as assistance tailored specifically to the needs of social, behavioural and economic research.

---

<sup>1</sup> The **German Data Forum** advises the federal government and the governments of the Länder on matters pertaining to the research data infrastructure in the empirical social, behavioural and economic sciences. The German Data Forum consists of eight representatives from social,

---

behavioural and economic disciplines, who are elected by the scientific community, and eight appointed representatives of important national data producers. It is a forum for dialogue between scientific data users and data producers, who jointly issue recommendations and statements. The two groups work together on developing a research infrastructure that facilitates broad, flexible, and secure access to data for science and research. The German Data Forum has accredited 31 research data centres since 2004 and fosters cooperation between them. <https://www.ratswd.de/en>

<sup>2</sup> [https://ec.europa.eu/research/openscience/pdf/eosc\\_declaration.pdf](https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf)

<sup>3</sup> Commission for the Improvement of the Informational Infrastructure between Research and Statistics (KVI) (ed.): Towards a better informational infrastructure. NOMOS Verlagsgesellschaft, Baden-Baden.

<sup>4</sup> **Research Data Centres (RDCs)**

The German Data Forum has created the research data centre as a model for flexible and comprehensive access to sensitive data for science and research. Since their inception, the research data centres have played a pivotal role in the social, empirical, and economic sciences and have strengthened the international competitiveness of the German research landscape.

The decentralised structure of the research data infrastructure is a tried and tested way to satisfy the demands of data producers, data users in science and research, and data protection. Anonymisation and pseudonymisation procedures are but two of a wide range of data protection measures, which include Public Use Files (PUF), Scientific Use Files (SUF), on-site access at research data centres, remote data access and the generation of synthetic data.

The RatSWD has played a major role in supporting and advancing these model solutions. They have led to a surge in available data for researchers in the past decade.

<https://www.ratswd.de/en/data-infrastructure/info>

<sup>5</sup> RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2017): Tätigkeitsbericht 2016 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.26> (03.11.2017).

<sup>6</sup> Mons et al. (2017): Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud, *Information Services & Use* 37(1), 49-56. <https://doi.org/10.3233/ISU-170824>.

<sup>7</sup> RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2017): Qualitätssicherung der vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdatenzentren (FDZ). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.4> (08.12.2017).

<sup>8</sup> RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2016): Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.7> (07.11.2017).