

Letter of Intent

Consortium for the Social, Behavioural, Educational, and Economic Sciences (KonsortSWD)

1 Binding letter of intent as advance notification or non-binding letter of intent
[Please indicate clearly whether your document is a binding letter of intent as advance notification or a non-binding letter of intent.]

<input checked="" type="checkbox"/>	Binding letter of intent (required as advance notification for proposals in 2019)
<input type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2020)
<input type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2021)

2 Formal details

Planned name of the consortium

- Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften (Consortium for the Social, Behavioural, Educational, and Economic Sciences)

Acronym of the planned consortium

- KonsortSWD (ConsortEcoSoc)

Applicant institution

- GESIS – Leibniz Institut für Sozialwissenschaften (Leibniz-Institute for the Social Sciences)
- President: Christof Wolf

Spokesperson

- Christof Wolf, Christof.Wolf@gesis.org, GESIS - Leibniz Institut für Sozialwissenschaften

3 Objectives, work programme and research environment

3.1 Research area of the proposed consortium (according to the DFG classification system)

- 12 Social and Behavioural Sciences (main subject area)
- 11 Humanities
- 22 Medicine

3.2 Concise summary of the planned consortium's main objectives and task areas

The vision of KonsortSWD is to **support long-term research data management for the social, behavioural, educational, and economic sciences** in all phases of the research data lifecycle. Building on the progress made so far, KonsortSWD will actively involve data users (1) in the establishment of deeper and wider data access, (2) in the support of research-data management for data production, (3) the development of a research ethics infrastructure and (4) in developing and implementing a technical framework for research data management. As an integral part of the KonsortSWD governance the German Data Forum (RatSWD)¹ will continue its activities bringing together data suppliers and data users (cf. [Bug, et al. 2018](#)).

Based on the established network of 34 Research Data Centres (RDC) **KonsortSWD will facilitate research** in the social, behavioural, educational, and economic sciences to analyse fundamental challenges such as the digital transformation, globalisation, demographic aging, migration, and social aspects of climate change. KonsortSWD will simplify **access to innovative data** including linked data from different sources.

Researchers in the disciplines covered by KonsortSWD use **diverse types of data**. In contrast to other disciplines, a large share of research data in the social, behavioural, educational, and economic sciences is generated outside of academia for non-scientific purposes (e.g., administrative data, data from official statistics). In addition, data are often **highly sensitive and** different **legal and ethical restrictions** apply. The **fragmented data-landscape** includes legally restricted data from official statistics and social insurance organisations. It also comprises confidential business data, proprietary data, privacy-sensitive survey-, experimental- and test-data as well as publicly available (aggregate) data on (supra-)national and regional units.

The work programme of KonsortSWD is organised in the following task areas:

- Task Area 1: Community Participation – ensures user participation at various levels.
- Task Area 2: Data Access – coordinates the needs and services of the RDCs.
- Task Area 3: Data Production – supports the needs of data generating units in areas with innovative data types that are not yet well connected as RDCs of KonsortSWD.
- Task Area 4: Research Ethics – aims to set up an infrastructure to support the cultural shift to more research ethics awareness.
- Task Area 5: Technological Solutions – combines activities generating and implementing technological solutions to data-management challenges.
- Task Area 6: Secretariat – offers administrative services to KonsortSWD and supports the spokesperson.

Specifically, the task areas set out to reach the following objectives:

- Refocus existing activities of the German Data Forum (RatSWD) and reflect user interests more specifically

¹ The German Data Forum (RatSWD) comprises two groups: representatives of data providers, such as the statistical offices, social insurance-providers and major scientific research programmes, as well as representatives of the scientific community. The latter are elected by more than 3,000 researchers (turnout in 2017) from at least 15 academic organisations (Akademie für Soziologie, DGfE, DGGÖ, DGMP, DGfP, DGPs, DGPuK, DGS, DGSKA, DGV, DStatG, DVPW, GEBF, VHB, VfS). The duality of data producers and users ensures RatSWD's widespread acceptance and success.

- support the cultural development towards **open science** and establish platform-independent open formats for data/metadata exchange. (TA 1, TA 2, TA 5)
- act as an **international and central point of contact** for handling data relevant for the social, behavioural, educational, and economic sciences. (TA 1)
- connect researchers with **international data infrastructures**, the European Open Science Cloud in particular. (TA 1)
- lobby for improved access to non-academic data sources. (TA 1)
- Respond to specific **challenges in research data management posed by diverse data types and sensitivity/disclosiveness of data** by (TA 2, TA 3, TA 5)
 - making available and archiving (sensitive and/or restricted) data from public and private sources that were often not collected for research purposes, (TA 1, TA 2)
 - implementing procedures for making complex data types (audio, video, texts, experimental data, spatial data, unstructured (web-based) data accessible that ensure compatibility of metadata and linkability within and across several data types and academic disciplines, (TA 2, TA 5)
 - warranting uniform, easy access to research data, based on a cross-disciplinary consensus about standards for RDCs, e.g. by expanding accepted, maintained, and well-known repositories for all relevant fields of research. (TA 2, TA 5)
- Secure and enhance the quality and **FAIRness of (sensitive) data**, data documentation, data archiving, data protection, and data security by **building on the success of the accredited RDCs** (cf. [RatSWD, 2017](#)). (TA 2, TA 3, TA 5)
- Increase the **linkability of existing data** through technical means (unique identifiers and linked open data), concepts for data protection, and harmonisation of key concepts. (TA 5)
- Implement transparent and user-friendly **standards on how to adhere to legal restrictions and to implement ethical considerations** regarding data use. (TA 4)
- Foster **skills development** in light of future requirements resulting from the increasing digitisation of research practice; strengthen competencies in data management and data use for research within RDCs, intermediary institutions (e.g. libraries) and among data users to **improve the understanding of the value of data access and data sharing**. (TA 2)

KonsortSWD's services are driven by community needs: community acceptance and implementation of all services will be closely monitored on the basis of established indicators (cf. [RatSWD, 2018](#)), which will also shape future service development. Accordingly the consortium's structure will be flexible to accommodate new service requests and partner institutions. Central decisions on service operation and service development will be tied directly to the needs of the research communities, e.g. via the RDCs' users' councils, the German Data Forum (RatSWD), and academic societies. The consortium will also analyse feedback and demand collected through intermediary institutions with direct contact to users (such as specialised information services (Fachinformationsdienste) and university libraries). Thus, communities will shape the strategic development of the task areas, their measures and developed services measures.

3.3 Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives

KonsortSWD is committed to **improving the availability and accessibility of (sensitive research) data**, while taking **data protection regulations and research ethics** into account. KonsortSWD will build on the network of **accredited RDCs** which provide an evolved and sustainable research infrastructure. RDCs are user-driven bottom-up initiatives from within the scientific community. This decentral network of currently 34 data centres provides researchers with user-friendly and cost-efficient access to a wide range of relevant granular and predominantly sensitive data. The process of accreditation serves as an entry-point for new RDCs to an interdisciplinary context for data sharing. It ensures the quality of service according to guidelines established by the German Data Forum (RatSWD). The services of these RDCs and their cooperation are the backbone of KonsortSWD.

KonsortSWD will build on

- the German Data Forum (RatSWD) in which data users and data providers pool and develop their expertise on research data infrastructures with the long-term aim of **safeguarding and continuously improving findability, accessibility, interoperability, and re-usability of research data** in the social, behavioural, educational, and economic sciences.
- standards and **criteria for self-monitoring** within the RDCs **and new accreditation** of additional RDCs. These quality measures will be continued and further developed in the context of KonsortSWD.
- existing RDC's **metadata tools** for documenting diverse types of data to enhance comparability and quality of these metadata, previous achievements in documentation, open source tools for **best-practice in research data management allowing to build** easy-to-analyse data products.
- decentral **expertise in concept harmonisation** to facilitate comparability of resulting data.
- a **broad network** of academic societies, specialised information services (Fachinformationsdienste), university libraries, and ethics committees.
- the established cooperation with decentral **research ethics committees**.
- proven **data-repository and search technologies** to promote findability and interoperability.
- **da|ra** - general doi-registration service for all RDCs.
- existing **training curricula** at accredited RDCs to expand an affordable training curriculum for (future) researchers and multipliers as well as infrastructure personnel at all stages of their careers.
- the operationalisation of FAIR from within **EcoSoc-IN**, the first **GO FAIR Implementation Network** for the social, behavioural, educational, and economic sciences founded by the German Data Forum (RatSWD) in 2018.
- the integration of new data types drawing on
 - **Qualiservice, eLabour, Research Data Centre Education** at DIPF and further RDCs focussing on sensitive (e.g. video-) data from small-N studies to help establish data-sharing in the respective communities.
 - existing **R libraries** to establish interfaces between text-data from KonsortSWD adjacent disciplines and from other contexts.

3.4 Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration

Consortia we expect to apply in 2019:

- **BERD@NFDI:** The cooperation is intended to cover research data management and merging of (un)structured (big) data as well as training of social data scientists.
- **NFDI4BioDiversity:** We collaborate in the area of geodata use and handling of sensitive data. KonsortSWD will supply social science data relevant to biodiversity research and will give support in dealing with sensible data, in particular anonymisation services.
- **NFDI4Earth:** The objective is to establish interoperability between geodata and (for example) survey data. The collaboration is planned to be three-fold: 1. Create geodata files that can easily be used in social sciences. 2. NFDI4Earth offers its competencies in using geodata. 3. KonsortSWD provides its expertise regarding anonymisation to NFDI4Earth
- **NFDI4Health:** We coordinate our efforts with NFDI4Health in addressing challenges that are related to making sensitive cohort data and survey data 1. re-usable and 2. interoperable. Participants of NFDI4Health aim to join the EcoSoc Implementation Network in order to join forces directly under the auspices of the Go FAIR initiative.
- **Text+:** Both consortia have an interest in text based unstructured data (for KonsortSWD this is especially the case in political sciences and media studies). A pilot-measure about text-based research data will use metadata and experience from Text+ and will work together strongly with Text+. KonsortSWD will offer its expertise regarding anonymization.

Consortia we expect to apply after 2019:

- **BRIDGE4NFDI** and **2LINK4NFDI:** KonsortSWD's experience in collaborating between data centres and consortia may offer feedback for BRIDGE4NFDI and 2LINK4NFDI regarding their service development. KonsortSWD aims to use the services generated by the two cross-cutting consortia.
- **ForumX:** ForumX will join the data access network of KonsortSWD when it has established its research data centre services. It will benefit from the existing repository experience in the group of RDCs. Furthermore, the synergies with the research data centre PsychData can lead to harmonized metadata for experimental data. ForumX will close a gap within the repository framework of KonsortSWD and profit from access to the broader data infrastructure as coordinated within KonsortSWD.
- **LoReData:** LoReData can profit from better integration of geodata and regional data that is accessible via several research data centres in KonsortSWD.
- **NFDI4Objects:** Within KonsortSWD, anthropologists work with object-based or immaterial non-digital data. KonsortSWD aims to collaborate with NFDI4Objects on specific services for material objects and their contextualisation. KonsortSWD offers services to support the data-management needs that are methodologically close to social research data (such as surveys in the specific research contexts) and that share its sensitive nature. KonsortSWD can support NFDI4Objects regarding the legal and ethical ways of re-use of sensitive research.
- **WEB NFDI:** The web archive to be provided by WEB NFDI is explicitly intended as a source for social scientists. KonsortSWD will explore a cooperation for analysing unstructured text data contained in web sources. KonsortSWD is also interested in exploring to link long-term web-data trends to surveys and economic data.

4 Cross-cutting topics

4.1 Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.

- **Sustainable archiving structures** – how to provide long-term, sustainable data archiving and data access services to users?
- **Standards for metadata/documentation** – how to describe the data and data life cycle comprehensibly and comparably?
- **Research Ethics** – Data (re-) use and exchange bring up ethical aspects that have to be addressed throughout the data lifecycle.
- **Legal standards, particularly for sensitive data** – which legal regulations are relevant for data management and (re-) use?
- **Dealing with sensitive data** – how to balance between anonymization/legal data access and informational degree/analytical potential of data?
- **Linking data** – how to foster the linkage between (non-)sensitive and sensitive data, how to deal with different provenance of data, and how to manage the trade-off between access and data protection regulations?
- **Technology for search and indexing** – improve FAIRness of social science research data through means-tested and open-source tools.
- **Lobbying for data access to non-generic research data** – how to promote access to data that were not initially gathered with the objective of research such as data originating from registers or user generated content on the Internet?

4.2 Please indicate which of these cross-cutting topics your consortium could contribute to and how.

- **Lobbying for data access to non-generic research data** - KonsortSWD has a proven track-record of successful lobbying for opening up non-academic data sources for scientific inquiry based on 15 years of work by German Data Forum (RatSWD).
- **Research Ethics** - The German Data Forum (RatSWD) has started to set up a network with local ethics committees and has gathered broad support for the infrastructure plans. The experiences relevant to (re-)use and management of research data can be shared with other NFDI entities in a systematic fashion.
- **Legal standards, particularly for sensitive data** - Given the very specific restrictions on data used for research in the disciplines covered by KonsortSWD, our RDCs in particular have valuable knowledge and best-practise experience which can be shared with other NFDI entities.
- **Dealing with sensitive data** - Given the varieties of sensitive data used for research in the disciplines covered by KonsortSWD, our RDCs in particular have valuable experience in handling the trade off of satisfying anonymization requirements while maintaining the research potential of data. This experience is likely to be useful for other NFDI entities.