

Extended Abstract in preparation for the NFDI conference on May 13th and 14th 2019

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

Hintergrundinformation der DFG zum NFDI-Procedere der Initialisierungsphase:

Attending the NFDI conference requires submission of an extended abstract. The extended abstract will assist with preparation for the conference with respect to networking and discussions among participants, the NFDI expert committee and the DFG Head Office. The content is not binding, but should describe the current situation and needs from the perspective of the respective scientific communities.

Please note the following general preliminary remarks:

- The information set out in the extended abstracts is **non-binding** with regard to any later proposal and review, will **not be subject to any review process** and should reflect the **current status of discussions and planning**.
- The extended abstracts **serve the sole purpose of preparation for the conference** with regard to discussions and participant networking as well as organisation on the part of the DFG Head Office.
- The extended abstracts will only be made available to participants in the NFDI conference and the DFG Head Office.
- The description of the “2. Subject-specific and infrastructural focus of the planned consortium” must not exceed a maximum of six pages. The document must be structured according to the template provided (this rtf file, Arial 11).
- Extended abstracts must be written in English.
- Please send your extended abstract/this rtf file to the following email address: nfdi@dfg.de
- **DEADLINE: March 29, 2019**

1 Formal Details

1.1 Planned title of the consortium

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

1.2 Acronym of the planned consortium

- KonsortSWD (ConsortEcoSoc)

1.3 Lead institution or facility

- to be determined

1.4 Name and work address of a contact person (including email address and institutional affiliation)

- Prof. Regina T. Riphahn, Ph.D. (Chair German Data Forum, regina.riphahn@fau.de, office@ratswd.de)

1.5 Members of the planned consortium (including institutional affiliation, without address)

- Dr. Christian Aßmann, LIfBi
- Dr. Doris Bambey, DIPF
- Dr. Mathias Bug, WZB/Geschäftsstelle RatSWD
- Dr. Jan Goebel, SOEP am DIW Berlin
- Prof. Dr. Stefan Liebig, SOEP am DIW Berlin
- Prof. Regina T. Riphahn, Ph.D., RatSWD, FAU Erlangen-Nürnberg
- Prof. Dr. Marc Rittberger, DIPF
- Dr. Pascal Siegers, GESIS
- Prof. Dr. Klaus Tochtermann, ZBW
- Prof. Dr. Christof Wolf, GESIS

1.6 Participants in the NFDI conference (names, institutional affiliation and email address; max. 3 persons)

- Prof. Regina T. Riphahn, Ph.D. (Chair of German Data Forum office@ratswd.de, regina.riphahn@fau.de)
- Prof. Dr. Stefan Liebig (Director of SOEP, sliebig@diw.de)
- Prof. Dr. Christof Wolf (President of GESIS, christof.wolf@gesis.org) – will attend the NFDI conference on May 13th
- Prof. Dr. Marc Rittberger (Deputy Executive Director of DIPF, rittberger@dipf.de) – will attend the NFDI conference on May 14th

1.7 Research area of the planned consortium (research area according to the DFG classification system [not subject areas]:

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

1.8 Participating research and infrastructure institutions as well as potential information service providers (without address)¹

German Data Forum (RatSWD): The consortium was built around the German Data Forum (RatSWD). Established in 2004, the RatSWD is an independent council that links data producers and data-using scientific communities. It advises the German federal government and the federal states (Länder) in matters concerning the research data infrastructure for the empirical social, educational, behavioural, and economic sciences. The RatSWD has 16 members. Membership consists of eight elected representatives of the social, educational, behavioural, and economic sciences and eight appointed representatives of Germany's most important data producers. The RatSWD offers a forum for dialogue between researchers and data producers, who jointly issue recommendations and position papers. The RatSWD furthers the development of a research infrastructure that provides researchers with flexible and secure access to a broad range of data. As of yet, the RatSWD has accredited 32 research data centres (RDCs²) and fosters their interaction and collaboration. The current structures involve the cooperation between the following institutions:

- Forschungsdaten- und Servicezentrum der Bundesbank (FDSZ Bundesbank)
- Forschungsdatenzentrum ALLBUS bei GESIS (FDZ ALLBUS)
- Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen (FDZ IQB)
- Forschungsdatenzentrum Archiv für Gesprochenes Deutsch am Institut für Deutsche Sprache (FDZ AGD)
- Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO)
- Forschungsdatenzentrum Bildung am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation (FDZ Bildung)
- Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ BA im IAB)
- Forschungsdatenzentrum der Bundeszentrale für gesundheitliche Aufklärung (FDZ BZgA)
- Forschungsdatenzentrum der Rentenversicherung (FDZ-RV)
- Forschungsdatenzentrum der Statistischen Ämter der Länder (FDZ-Länder)
- Forschungsdatenzentrum des Beziehungs- und Familienpanels (FDZ-pairfam)
- Forschungsdatenzentrum des Deutschen Jugendinstituts (FDZ-DJI)
- Forschungsdatenzentrum des Deutschen Zentrums für Altersfragen (FDZ-DZA)
- Forschungsdatenzentrum des Kraftfahrtbundesamtes (FDZ im KBA)
- Forschungsdatenzentrum des Leibniz-Instituts für Bildungsverläufe e.V. an der Otto-Friedrich-Universität Bamberg (FDZ-LifBi)
- Forschungsdatenzentrum des Leibniz-Instituts für Wirtschaftsforschung Halle (FDZ-IWH)

¹ Most RDCs/involved institutions conduct research of their own in order to be able to provide data, services, and consulting at an appropriate scientific level. Hence, we do not group the participating institutions.

² The German expression for RDC is "Forschungsdatenzentrum" (FDZ), Research Data and Service Centres (RDSC) are referred to as "Forschungsdaten- und Servicezentrum" (FDSZ).

- Forschungsdatenzentrum des Robert Koch-Instituts (FDZ RKI)
- Forschungsdatenzentrum des Sozio-oekonomischen Panels am DIW Berlin (FDZ SOEP)
- Forschungsdatenzentrum des Statistischen Bundesamtes (FDZ-Bund)
- Forschungsdatenzentrum des Survey of Health, Ageing and Retirement in Europe (FDZ SHARE)
- Forschungsdatenzentrum des Zentrums für Europäische Wirtschaftsforschung (ZEW-FDZ)
- Forschungsdatenzentrum eLabour
- Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung (fdz.DZHW)
- Forschungsdatenzentrum German Microdata Lab bei GESIS (FDZ GML)
- Forschungsdatenzentrum im Bundesinstitut für Berufsbildung (BIBB-FDZ)
- Forschungsdatenzentrum Internationale Umfrageprogramme bei GESIS (FDZ Internationale Umfrageprogramme)
- Forschungsdatenzentrum PIAAC bei GESIS (FDZ PIAAC)
- Forschungsdatenzentrum PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation (FDZ PsychData des ZPID)
- Forschungsdatenzentrum Qualiservice (Qualiservice)
- Forschungsdatenzentrum Ruhr am RWI – Leibniz-Institut für Wirtschaftsforschung (FDZ Ruhr am RWI)
- Forschungsdatenzentrum Wahlen bei GESIS (FDZ Wahlen)
- Forschungsdatenzentrum Wissenschaftsstatistik des Stifterverbandes (FDZ Wissenschaftsstatistik)
- Internationales Forschungsdatenzentrum des Forschungsinstituts zur Zukunft der Arbeit (FDZ IZA, IDSC)
- LMU-ifo Economics & Business Data Center (EBDC)
- Wissenschaftszentrum Berlin für Sozialforschung (WZB)
- WSI-Datenzentrum
- ZBW - Leibniz-Informationszentrum Wirtschaft

1.9 Planned proposal submission date (2019, 2020, 2021)

- 2019

1.10 Overview diagram or organisational chart for the planned consortium

- to be determined

2 Subject-specific and infrastructural focus of the planned consortium

2.1 Key questions/objectives of the consortium

The disciplines represented by KonsortSWD deal with **socially relevant issues**. To address these issues, research communities require **comprehensive access to innovative data**. Researchers in the social, behavioural, educational, and economic sciences use **diverse types of data**, which are often highly sensitive. There is **high potential for making these data FAIR**.

The stakeholders of KonsortSWD have set the following **main objectives** for their enhanced cooperation within the framework of NFDI:

- **Facilitate research** in the social, behavioural, educational, and economic sciences to help societies cope with fundamental challenges including digital transformation, globalisation, demographic decline and aging societies, migration, and social aspects of climate change.
- Respond to specific **challenges posed by the multiplicity of data types and data sources** by
 - making (sensitive) data available that were not initially gathered for research purposes and archiving these data,
 - providing and enhancing access to sensitive research data, also from public and private sources,
 - establishing links to and between related data communities (e.g., text-data, spatial data).
- 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 Research Data Centres (RDCs)** ([see also Activities Report 2017](#)).
- Foster **skills development** in view of future requirements resulting from the increasing digitisation of research practice; strengthen competencies in data management and data use for research within the RDCs as well as among data users.
- **Improve the understanding** of the **value of data sharing** through training and lobbying.
- **Enhance availability of restricted data** for research purposes.
- Coordinate data generation and standards to **provide consistent metadata** within and across several data types and academic disciplines.
- Foster an **open** approach to data, methods, and research results (open science).
- Act as an **international and central point of contact** for handling data from the social, behavioural, educational, and economic sciences (research data management).
- Connect and **continuously expand existing activities**: the services of KonsortSWD will be implemented directly in all branches of the consortium.
- Set up a **flexible consortium structure**: “elastic stability” facilitates continuous development of the service portfolio and contribution of additional partners and work packages.

2.2 Known needs/current status of research data management in the relevant discipline/subject-specific relevance of the planned consortium

KonsortSWD is designed to **meet the needs of a significant share of German academia**: in sum, about 37 % of students³ and about 29 % of professors⁴ in German academia (as of 2017) study and work in disciplines relating to law, economics, and the social sciences. Furthermore, in 2016, 19 % of doctoral candidates came from these fields.⁵ These numbers show the relevance KonsortSWD has for academia.

2.2.1 From a research perspective

The social, behavioural, educational, and economic sciences need to draw on a wide range of data in order to address relevant topics. The **fragmented data-landscape** not only consists of generic research data, but also includes legally restricted data from official/administrative statistics and social insurance organisations, confidential business data, proprietary data, privacy-sensitive survey-, experimental- and test-data, and publicly available (aggregate) data on national and regional contexts. KonsortSWD plans to establish a **unified interface**. Furthermore, as the quantity and volume of research-generated data (e.g., data collected in surveys, experiments, and observations) is increasing, systematic efforts and technical solutions across decentral repositories are required to make data findable, accessible, re-usable, and interoperable (FAIR). From a research perspective, a suitable infrastructure for research data management needs to **consider data quality, legal, ethical, and technical aspects throughout the data lifecycle** to establish community-wide standards and provide FAIR data.

Resulting needs are:

- uniform, easy **access to research data**, based on a cross-disciplinary consensus about standards for RDCs,
- **transparent and user-friendly** implementation of legal restrictions regarding **data use**,
- accepted, maintained, and well-known **repositories for all relevant fields** of research,
- increased **linkability of existing data** through technical means (unique identifiers and linked open data), concepts for data protection, and harmonisation of key concepts,
- procedures for **access to complex data types** (audio, video, texts, and experimental data) that ensure compatibility of metadata and linkability,
- enhanced **open science** by establishing and adapting a platform- and software- independent open format for data exchange and metadata,
- **interoperable concepts**, i.e., operationalisations and metadata based on international and interdisciplinary standards, and
- continuous **(political) lobbying for data access** and data sharing for research purposes.

³ https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/StudierendeHochschulenEndg2110410187004.pdf?__blob=publicationFile, p. 33

⁴ https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PersonalHochschulen2110440177004.pdf?__blob=publicationFile, p. 48

⁵ https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/2017/04/PromovierendeDeutschland_042017.pdf?__blob=publicationFile, p. 110

2.2.2 In terms of available information providers and services

So far, 32 RDCs accredited by the German Data Forum (RatSWD)⁶ provide access to granular and predominantly sensitive data. Several additional RDCs have applied for RatSWD accreditation.

At present, most existing services are insular to some extent. KonsortSWD aims to provide a comprehensive structure to **promote the harmonisation of relevant standards for the used data types**. Not all services of information providers in the disciplines covered by KonsortSWD conform to FAIR criteria as of yet.

The RDCs currently lack a **comprehensive search function** across data centres. In addition, there are major obstacles to linking data across surveys and different data types that impede the validation of results as well as new research approaches to address relevant societal issues. The consistent use of high-quality metadata standards based on substantial content will improve community-specific search portals at the RDCs and support users across all research domains employing

- an affordable and universal **training curriculum** for (future) RDC staff, multipliers, and researchers engaging with the data covered by KonsortSWD,
- focused efforts regarding **data that are difficult to anonymise**. This applies, for example, to qualitative social research based on small number of direct interactions with subjects or video data of individuals in particular,
- integration of internationally established **metadata standards** using tools and agreed-upon best practises into major data-generating workflows,
- tools for **standardised documentation**,
- increased **visibility of decentral curated data** by making metadata compliant with established commercial data search engines.

2.3 Summary of the planned research data infrastructure that is specifically intended to address the needs of research users in their respective work processes

The RatSWD-accredited RDCs form a highly evolved and sustainable research infrastructure, which is a result of independent initiatives from within the scientific community. This decentral network of currently 32 data centres provides researchers with user-friendly and cost-efficient access to a wide range of relevant data.

KonsortSWD will connect several community-based coordination committees that focus on specific phases throughout the data lifecycle. Representatives from all RatSWD-accredited RDCs will continue to pool and develop their expertise on the implementation and provision of research data infrastructures with the long-term aim of **safeguarding and continuously improving the findability, accessibility, interoperability, and re-usability of research data** in the social, behavioural, educational, and economic sciences. RDCs accredited by the RatSWD have developed standards and criteria for self-monitoring and new accreditation of additional RDCs. These **quality measures will be continued and further developed** in the context of KonsortSWD.

⁶ The RatSWD comprises **two groups**: representatives of **data providers**, such as the statistical offices, social insurance-providers or major scientific research programmes, and **representatives of the scientific community**, who are **elected** by more than 3,000 researchers (voter turnout in 2017) from at least 14 academic organisations (Akademie für Soziologie, DGfE, DGGÖ, DGMP, DGfP, DGPs, DGPuK, DGS, DGSKA, DStatG, DVPW, GEBF, VHB, VfS). The electorate of the RatSWD also covers additional academic organisations (cf. <https://www.ratswd.de/en/ratswd/election-procedure>). This **duality is the reason for the RatSWD's widespread acceptance** and success. In facilitating cooperation between the scientific community and data-providing institutions, the RatSWD has developed a range of widely acknowledged guidelines (e.g., for data management), which have also advanced research data access in adjacent disciplines.

The RatSWD will monitor this development. It will further develop and guide the self-monitoring and accreditation processes that were established during the last years and continue to run a complaints office based on the monitoring standards.

2.4 Description

2.4.1 of data types

subject	data types	main restrictions
individuals	process-generated administrative data, official statistics (e.g., in context of social insurances and registers)	privacy
	(longitudinal) survey data	privacy, ethical
	Experimental	privacy, ethical
	behavioural data (e.g., geotracking, sensor, web, social media)	privacy
	qualitative interviews, observations, audio, and video records	privacy, ethical
groups of individuals	audio and video records	copyright
firms / organisations	business data (e.g., earnings, patents), transaction data (e.g., stocks), texts (e.g., stemming from decision making processes, parliamentary protocols), web, social media, media reports	legal, proprietary data, copyright, confidentiality
small-scale regional data	crime, regional tax revenue, aggregated survey data, geodata (aggregated in e.g., postal codes, cities, states)	sometimes legal
nations	national statistics, aggregated survey data, texts (e.g., stemming from political decision making processes, parliamentary protocols), media reports	copyright, confidentiality
metadata		sometimes copyright

Different legal restrictions apply and have to be carefully taken into account to make data available as research data.

In contrast to other disciplines, a large share of research data in the social, behavioural, educational, and economic sciences are **generated outside of academia for non-scientific purposes**. Since much of these data originate from publicly funded, administrative sources, the disciplines represented by KonsortSWD need a lobby to articulate the needs of empirical academic researchers. The RatSWD has been a successful lobby [e.g., EU GDPR, BStatG] and will continue to speak up for cognate disciplines in the context of NFDI.

2.4.2 of underlying data processing / data analysis methodologies

An important challenge for data processing in the social, behavioural, educational, and economic sciences is that **confidential or restricted data** cannot be made publicly available for re-use.

The **following methodologies are most common** in the KonsortSWD's disciplines:

- for data processing
 - anonymisation and pseudonymisation,
 - documentation according to established standards,
 - enrichment of data sets,
 - manual coding,
 - transcription of audio / video sources,
 - automated data mining (web-based, text-based sources),
 - data linking (for not-yet anonymised data), and
 - automated or manual merging data from different sources.
- for data analysis
 - statistical, econometric, and psychometric methods,
 - qualitative-hermeneutic methods, and
 - machine learning.

2.5 Planned implementation of the FAIR principles⁷ and information about any existing policies or guidelines in the relevant discipline

KonsortSWD is committed to **improve the availability and accessibility of research data**, while taking data protection regulations and research ethics into account. In 2018, the RatSWD founded **EcoSoc-IN**, the first GO FAIR Implementation Network for the social, behavioural, educational, and economic sciences (see below). In the context of NFDI, key areas of activity will continue to include access to data from public as well as private sources and their interoperability to strengthen research data management.

KonsortSWD will establish and further develop the following services, which secure and advance independent research based on RDC data and safeguard the quality of the research data infrastructure:

- Future research in the fields represented by KonsortSWD depends on continued (political) lobbying for **accessible** data and data sharing. It remains a permanent mission to monitor and respond to proposals for legal adjustments, e.g., with respect to data protection and data access. Empirical research relies on an established and respected voice for science in these processes.
- A joint search infrastructure for RDCs to ensure **findability**.
- A joint metadata framework to generate machine-readable data in order to ensure **interoperability**.
- Processes and standards to harmonise data across major large-scale surveys to ensure **re-use**.
- **Training** for researchers and infrastructure personnel at all stages of their careers (including on-the-job training and apprenticeships), lobbying for research data management as a subject within under- and postgraduate study programmes.
- The implementation of a **central information platform** (website, including a portal with access to the KonsortSWD's services, handouts, presentation slides, etc.) in cooperation with the individual organisational units warranting equal treatment of all data users.
- The documentation of and open source tools for **best-practice in research-data management** creating easy-to-analyse data products that feature quality-assured, standards-compliant metadata, and comprehensive documentation.
- Fostering the **quality of data** by setting standards regarding new methodological approaches and technological solutions.

⁷ <https://www.force11.org/fairprinciples>

2.6 Planned measures for user participation and involvement

The RatSWD will closely monitor all developments initiated by KonsortSWD. All central **service decisions will be tied directly to the needs of the research communities**. The RatSWD's successful activities will be extended both in breadth (number of users) and in depth (coverage of services). **Community-based coordination committees** will shape the strategic development of services / work-packages in their respective areas of expertise. They will be composed of experts from the relevant research fields.

Use and community acceptance of all services provided by KonsortSWD will be closely **monitored** on the basis of indicators, which will also affect future service development.

The consortium's services will focus on community needs.

2.7 Existing and intended degree of networking of the planned consortium

2.7.1 nationally (in particular with other, potential future consortia or existing state-level initiatives)

KonsortSWD will follow a

1) user-centric logic: It will establish close connections to university libraries and other specialised scientific information services (DFG-funded "Fachinformationsdienste").

2) research-driven logic: Existing RDCs accredited and monitored by the RatSWD represent an increasing breadth of data types. They are also used by different disciplines – within the social, behavioural, educational, and economic sciences and beyond. Drawing on this existing network, KonsortSWD will deepen and expand its network to include adjacent disciplines.

Following the demands of the research data landscape, the RatSWD will maintain its networking activities with administration and politics.

3) data type logic: Within NFDI, consortia from the humanities and life sciences as well as from geosciences share similar methodological approaches and research interests. Therefore, these consortia are of special importance to KonsortSWD. Creating synergies through shared services will be of high importance.

2.7.2 internationally

The RatSWD initiated the creation of EcoSoc_IN, a FAIR Implementation Network. EcoSoc_IN is an open network that will provide solutions for the implementation of the **FAIR principles to social science data with particular attention to issues arising from sensitive data**. The implementation network will put forward measures to achieve data interoperability. ZBW, which hosts the German office of the GO FAIR initiative, is an integral part of KonsortSWD.

Participating institutions are engaged in three different **European Research Infrastructure Consortia (ERIC)**. (1) GESIS is the national service provider for the Consortium of Social Science Data Archives (**CESSDA ERIC**). CESSDA connects the service providers for the long-term preservation of research data and innovative research data management. (2) The research data centre of the Survey of Health and Aging in Europe (SHARE) is part of the **SHARE ERIC** - a data infrastructure for the study of social questions related to ageing in Europe. (3) GESIS represents Germany in the European Social Survey (**ESS ERIC**) that provides high-quality survey data for studying social cohesion in Europe.

Throughout the three ERICs, KonsortSWD will participate in establishing and populating the European Open Science Cloud (EOSC). CESSDA ERIC coordinates a large scale H2020 infrastructure project to establish a **Social Science and Humanities Open Cloud (SSHOC)**. SSHOC

gathers all social science and humanities ERICs and other relevant European providers of research infrastructures. The SSHOC will be an EOSC access point for the social sciences. KonsortSWD will bring tools and services developed for the NFDI to the EOSC marketplace. Therefore, the major challenges for infrastructure development will be the **alignment of the NFDI and EOSC services** for optimal benefits of the scientific users.

KonsortSWD has established a close relationship with the **UK Data Forum**, including a regular exchange on access to public data relevant for researchers. Finally, many RDCs are involved in international research networks.

The RatSWD has been an organisational member of **Research Data Alliance (RDA)** since 2015. The cooperation with RDA on a national, European, and international level will be further pursued by KonsortSWD.

2.7.3 between the infrastructure facilities and the research community

Please refer to footnote 6.

2.7.4 with respect to major networking topics

KonsortSWD plans to contribute to NFDI with regard to the following topics

- internationally acclaimed standards and processes for **research access to sensitive data**,
- search and **indexing technology**,
- technology for **persistent identifiers** (DOI),
- **repository technology**,
- **user-centred governance** of consortia.

2.8 Additional information

- KonsortSWD already has a substantial user base, which is sure to further expand once services become more integrated. It comprises empirical researchers from all scientific disciplines, particularly from the social, behavioural, educational, and economic sciences, data producers (individual and institutional), 21,467 active data users (as of 2017) from Germany and abroad, and institutions (e.g., universities, libraries, RDCs, repositories, surveys) that integrate services into their own service portfolio or adapt technical solutions.