This document represents the original text of the KonsortSWD application to the DFG on 15.10.2019. Possible funding cuts are not yet taken into account. If funding is reduced not all work packages can be implemented as described.

Bei diesem Dokument handelt es sich um den Originaltext des KonsortSWD-Antrages mit Stand der Abgabe bei der DFG am 15.10.2019. Etwaige finanzielle Kürzungen sind noch nicht berücksichtigt. Ggf. können nicht alle Arbeitspakete wie beschrieben durchgeführt werden.

National Research Data Infrastructure (NFDI)
Consortium Proposal

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

Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften (KonsortSWD)

Applicant institution: GESIS – Leibniz Institut für Sozialwissenschaften (Leibniz Institute for the Social Sciences)

Spokesperson: Christof Wolf

Authors: Maja Adena, Christian Aßmann, Doris Bambey, Katarina Blask, Andreas Blätte, Michael Bosnjak, Daniel Buck, Mathias Bug, Anja Busch, Anna Fräßdorf, Daniel Fuß, Jan Goebel, Betina Hollstein, Monika Jungbauer-Gans, Dagmar Kern, Claus-Peter Klas, Susanne Kretzer, Stefan Liebig, Nicole Mayer-Ahuja, Bernhard Miller, Peter Mutschke, Anja Perry, Regina T. Riphahn, Marc Rittberger, Daniel Schiffner, Silke Schneider, Jürgen Schupp, Pascal Siegers, Ranjit K. Singh, Heike Solga, Klaus Tochtermann, Ralf Toepfer, Christof Wolf, Susanne Zindler

CONTENTS

Table of figures:

Table of Abbreviations and Acronyms

AB	Advisory Board
API	Application Programming Interface
AS	Akademie für Soziologie
AGD	Archiv für Gesprochenes Deutsch (Archive of Spoken German)
BERD@NFDI	NFDI applicant: Business, Economic and Related Data
BMBF	Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research)
BRIDGE4NFDI	NFDI applicant: 'Bridging boundaries among national research data infrastructures'
BZgA	Bundeszentrale für gesundheitliche Aufklärung (Federal Centre for Health Education)
CAMSIS	Cambridge Social Interaction and Stratification
CDA	Council for Data Access
CESSDA	Consortium of European Social Science Data Archives
CLARIN	Common Language Resource and Technology Infrastructure
cf.	confer (= compare)
CSS	Computational Social Science
CTS	Core Trust Seal
da|ra	Registration agency for social science and economic data
DDI	Data Documentation Initiative
DFG	Deutsche Forschungsgemeinschaft (German Research Foundation)
DGfE	Deutsche Gesellschaft für Erziehungswissenschaft (German Educational Research Association)
DGfP	Deutsche Gesellschaft für Politikwissenschaft
DGGÖ	Deutsche Gesellschaft für Gesundheitsökonomie
DGMP	Deutsche Gesellschaft für Medizinische Psychologie
DGPs	Deutsche Gesellschaft für Psychologie (German Psychological Society)
DGPuK	Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (German Communication Association)
DGS	Deutsche Gesellschaft für Soziologie
DGSKA	Deutsche Gesellschaft für Sozial- und Kulturanthropologie (German Anthropological Association)
DGV	Deutsche Gesellschaft für Volkskunde
DIPF	Leibniz-Institut für Bildungsforschung und Bildungsinformation (Leibniz Institute for Research and Information in Education)
DIW	Deutsches Institut für Wirtschaftsforschung (German Institute for Economic Research)
DJI	Deutsche Jugendinstitut (German Youth Institute)
DOI	Digital Object Identifier
DOIP	Digital Object Identifier Protocol
DStatG	Deutsche Statistische Gesellschaft (German Statistical Society)
DVPW	Deutsche Vereinigung für Politische Wissenschaft (German Political Science Association)
DZHW	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (German Centre for Higher Education Research and Science Studies)
EcoSoc-IN	Economic and Social Sciences goINg FAIR Implementation Network
e.g.	For example
EGP-scheme	Erikson-Goldthorpe-Portocarero-scheme

eLabour	Interdisziplinäres Zentrum für IT-basierte qualitative arbeitssoziologische Forschung (Interdisciplinary Centre for Qualitative Research Data from the Sociology of Work)
EOSC	European Open Science Cloud
ERA	European Research Area
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
ESS	European Social Survey
FAIR	Findable, Accessible, Interoperable, and Reusable
FAQ	Frequently asked questions
FDM	Forschungsdatenmanagement
FDZ-AGD	Forschungsdatenzentrum des Archivs für Gesprochenes Deutsch (RDC for the Archive of Spoken German)
FDZ-BO	Forschungsdatenzentrum Betriebs- und Organisationsdaten (RDC for Business and Organizational Data)
fdz.DZHW	Research Data Centre at DZHW
ForumX	NFDI applicant: experimental data across disciplines
FTE	Full-time equivalent
GEBF	Gesellschaft für empirische Bildungsforschung (Society for Empirical Educational Research)
GeRDI	Generic Research Data Infrastructure
GESIS	Leibniz-Institut für Sozialwissenschaften (Leibniz Institute for the Social Sciences)
GDPR	EU General Data Protection Regulation
GND-ID	Gemeinsame Normdatei-Identifikator (Integrated Authority File-Identifier)
GRW	Guest researcher workstation (Gastwissenschaftsarbeitsplatz)
IAB	Institut für Arbeitsmarkt- und Berufsforschung (Institute for Employment Research)
IDAN	International Data Access Network
IDS	Leibniz-Institut für Deutsche Sprache (Leibniz-Institute for the German Language)
ISCO	International Standard Classification of Occupations
ISEI	International Socio-Economic Index of Occupational Status
IQB	Institut für Qualitätsentwicklung im Bildungswesen (Institute for Educational Quality Improvement)
IZA	Institut zur Zukunft der Arbeit (Insitute of Labor Economics)
KA	Konsort Assembly
KLdB	Klassifikation der Berufe (Classification of Occupations)
KonsortSWD	Konsortium für die Sozial-, Bildungs- Verhaltens- und Wirtschaftswissenschaften (Consortium for the Social, Behavioural, Educational, and Economic Sciences)
LIfBi	Leibniz-Institut für Bildungsverläufe (Leibniz Institute for Educational Trajectories)
LOD	Linked Open Data
LoReData	NFDI applicant: Local and Regional Research Data
NEPS	National Educational Panel Study
NFDI	Nationale Forschungsdateninfrastruktur (National Research Data Infrastructure)
NFDI4Agri	NFDI applicant: Agricultural Sciences
NFDI4BioDiversity	NFDI applicant: Biodiversity, Ecology & Environmental Data
NFDI4Earth	NFDI applicant: Consortium Earth System Science

NFDI4Health	NFDI applicant: Data Infrastructure for Personal Health Data
NFDI4Medicine	NFDI applicant: Consortium of the Medical Informatics Initiative (MII) and the German Centers for Health Research (DZG)
NFDI4Objects	NFDI applicant: Archaeology, Material Culture and Objects
NFDI Web	NFDI applicant: Forschungsdateninfrastruktur für das World Wide Web
OCR	Optical Character Recognition
OAI	OpenAPI (Open application programming interface)
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
pairfam	Panel Analysis of Intimate Relationships and Family Dynamics
PID	Persistent Identifier
PR	Public relations
RatSWD	Rat für Sozial- und Wirtschaftsdaten (German Data Forum)
RDA	Research Data Alliance
RDC	Research Data Centre
RDCnet	Research Data Centre Network
RDI	Research Data Infrastructure
RDM	Research Data Management
RfII	Rat für Informationsinfrastrukturen (German Council for Scientific Information Infrastructures)
SEO	Search Engine Optimisation
SERISS	Synergies for Europe's Research Infrastructures in the Social Sciences
SHARE	Survey of Health, Aging, and Retirement in Europe
SIOPS	Standard International Occupational Prestige Scale
SIS	Fachinformationsdienst (Specialised Information Service)
SIS SCA	Specialised Information Service Social and Cultural Anthropology
SOEP	Sozio-oekonomisches Panel (Socio-Economic Panel)
SOFI	Soziologisches Forschungsinstitut Göttingen (Sociological Research Institute Göttingen)
SSH	Social sciences and humanities
SSHOC	Social Sciences and Humanities Open Cloud
StC	Steering Committee
SUF	Scientific Use File
TA	Task Area
Text+	NFDI applicant: Language- and Text-Based Research Data
2LINK4NFDI	NFDI applicant: Linking NFDI to existing Service and Support Structures in German Academia
UCD	User-centred design
VerbundFDB	Verbund Forschungsdaten Bildung (Framework Programme for the Promotion of Empirical Educational Research)
VfS	Verein für Socialpolitik
VHB	Verband der Hochschullehrer für Betriebswirtschaft (German Academic Association for Business Research)
VQualidat	Verbund Qualidaten (federated infrastructure for qualitative data)
WZB	Wissenschaftszentrum Berlin für Sozialforschung (Berlin Social Science Center)
ZBW	Leibniz-Informationszentrum Wirtschaft (Leibniz Information Centre for Economics)
ZPID	Leibniz-Institut für Psychologische Information und Dokumentation (Leibniz Institute for Psychology Information)

1 General Information

- Name of the consortium in English and German

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

Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften (KonsortSWD)

- Summary of the proposal in English

The **social, behavioural, educational, and economic sciences** offer indispensable perspectives and responses to the **grand societal challenges** the world is facing today. In addressing these challenges, the disciplines represented in KonsortSWD work with a range of **different data types**. These data are typically **sensitive** due to legal or ethical constraints and they are frequently generated for non-scientific purposes.

KonsortSWD aims to support the **research data management (RDM) needs of its communities**, which increasingly collaborate in multi- and interdisciplinary research projects. The (co-)applicant institutions bring their long-standing experience in operating community-driven research data infrastructures to the NFDI.

Our mission is to develop – strengthen, widen, deepen – a research data infrastructure for the study of human society. This effort is to be **primarily user-driven** and **addresses the needs of our research communities**. The core of KonsortSWD's RDM strategy is to provide researchers and research data centres (RDCs) with the **tools and services** they need for managing and sharing (new) **sensitive and non-sensitive** data in compliance with the FAIR principles. This includes supporting **sustainable RDM** in all phases of the research data lifecycle and securing **data accessibility** with an awareness of ethical and legal aspects.

KonsortSWD will increase efficiency in RDM processes by coordinating

- data producers with technical expertise to simplify **sustainable RDM throughout the data lifecycle**,
- a growing infrastructure of currently **34 accredited RDCs** with 3,940 datasets, 46,000 data users, and a staff of 285 full time-equivalents (in 2018),
- the **German Data Forum (RatSWD)**, where data users and producers jointly facilitate access to research data,

- a forum for **ethical reflection** to support researchers and their institutions and to discuss and disseminate **standards** for data-driven, high quality research on human subjects up to entire societies.

KonsortSWD combines a multi-stakeholder approach with transparent governance: a **Steering Committee** is in charge of operations, an **Advisory Board** with elected community representatives provides strategic feedback, and a **Consortial Conference** helps to disseminate KonsortSWD's services and to collect feedback. The Task Areas (TAs) focus on: (1) Community Participation. (2) Data Access, (3) Data Production, (4) Research (Data) Ethics, and (5) Technical Solutions.

- Deutsche Zusammenfassung

Die **Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften** eröffnen vielfältige Perspektiven und entwickeln wesentliche Ansätze, um Antworten auf **große aktuelle gesellschaftliche Herausforderungen** zu finden. Die mit KonsortSWD verbundenen Disziplinen arbeiten dabei mit **verschiedenen Datentypen**, die häufig aufgrund rechtlicher oder ethischer Beschränkungen besonders sensibel sind und häufig nicht originär für wissenschaftliche Zwecke erhoben wurden. KonsortSWD bietet den Forschenden, die zunehmend in multi- und interdisziplinären Projekten zusammenarbeiten, **Unterstützung bei ihrem Forschungsdatenmanagement (FDM)**. Die an KonsortSWD beteiligten Einrichtungen bringen dabei ihre Erfahrungen im Betrieb von nutzungsorientierten Forschungsdateninfrastrukturen in die NFDI ein.

Unsere Mission ist es, die Forschungsdateninfrastruktur zur Beforschung der Gesellschaft zu stärken, zu erweitern und zu vertiefen. Sie soll **nutzungsorientiert ausgestaltet** sein und die **Bedürfnisse der Forschungscommunities berücksichtigen**. Die FDM-Strategie von KonsortSWD zielt darauf ab, Forschende und Forschungsdatenzentren technisch und inhaltlich bei Verwaltung und Nachnutzung (neuer) sensibler und nicht sensibler Daten in Übereinstimmung mit den FAIR-Prinzipien zu unterstützen. Dazu gehört neben einem **nachhaltigen FDM über den Forschungsdatenlebenszyklus hinweg** auch die Sicherstellung des Datenzugangs unter **Berücksichtigung ethischer und rechtlicher Aspekte**.

KonsortSWD erhöht die Effizienz in FDM-Prozessen durch Koordination von:

- Datenproduzierenden, denen technischer Sachverstand zur Vereinfachung eines **nachhaltigen FDM** während des gesamten Forschungsdatenlebenszyklus zur Verfügung steht,

- **Qualitätssicherung und Wachstum** einer Infrastruktur von derzeit 34 akkreditierten Forschungsdatenzentren mit 3.940 Datensätzen, 46.000 Datennutzenden und 285 Vollzeitäquivalenten (Stand 2018) werden gewährleistet,
- der **Zusammenarbeit von Datennutzenden und Datenproduzierenden** im Rat für Sozial- und Wirtschaftsdaten (RatSWD) führt zu erweitertem Zugang zu Forschungsdaten,
- einem Forum für **ethische Reflexion**, das Forschende und ihre Einrichtungen unterstützt und flankierende **Standards** für die datenbasierte, qualitätsorientierte Forschung am Menschen bis hin zu ganzen Gesellschaften bietet.

KonsortSWD verbindet die **Einbindung verschiedener Gruppen mit einer transparenten Steuerung:** Ein **Lenkungsausschuss** verantwortet die operativen Abläufe, ein **Beirat** mit gewählten Vertretenden aus der Forschung gibt strategisches Feedback und eine **Konsortial-Konferenz** dient der Verbreitung und Qualitätssicherung der Arbeitsergebnisse von KonsortSWD. Die Task Areas (TAs) entwickeln Dienste für folgende Bereiche: (1) Nutzendeneinbindung, (2) Datenzugang, (3) Datenproduktion, (4) Forschungs(Daten)Ethik und (5) Technische Lösungen.

- Applicant institution

Applicant institution	Location
GESIS – Leibniz Institut für Sozialwissenschaften (Leibniz-Institute for the Social Sciences)	Mannheim

- Name of the consortium's spokesperson

Spokesperson	Institution, location
Christof Wolf	GESIS, Mannheim

- Co-applicant institutions

Co-applicant institutions	Location
DIPF – Leibniz-Institut für Bildungsforschung und Bildungsinformation	Frankfurt am Main
DZHW – Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH	Hannover
LIfBi – Leibniz-Institut für Bildungsverläufe	Bamberg
SOEP am DIW Berlin – Deutsches Institut für Wirtschaftsforschung e.V.	Berlin
SOFI – Soziologisches Forschungsinstitut Göttingen an der Georg-August-Universität	Göttingen
Universität Bremen – FDZ Qualiservice	Bremen
Universität Duisburg-Essen – Professur für Public Policy und Landespolitik	Duisburg, Essen
WZB – Wissenschaftszentrum Berlin für Sozialforschung gGmbH	Berlin
ZBW – Leibniz-Informationszentrum Wirtschaft	Kiel

ZPID – Leibniz-Zentrum für Psychologische Information und Dokumentation	Trier

- Names of co-spokespersons

Co-spokesperson	Institution, location	Task Area(s)
Marc Rittberger	DIPF, Frankfurt am Main	TA 5 Technical Solutions, TA 3 Data Production
Monika Jungbauer-Gans	DZHW, Hannover	TA 3 Data Production
Christian Aßmann	LIfBi, Bamberg	TA 2 Data Access
Jürgen Schupp	SOEP am DIW Berlin	TA 3 Data Production
Nicole Mayer-Ahuja	SOFI, Göttingen	TA 2 Data Access, TA 3 Data Production
Betina Hollstein	Universität Bremen	TA 2 Data Access, TA 3 Data Production
Andreas Blätte	Universität Duisburg-Essen	TA 3 Data Production
Maja Adena	WZB, Berlin	TA 4 Research Ethics
Heike Solga	WZB, Berlin	TA 1 Community Participation
Klaus Tochtermann	ZBW, Kiel	TA 5 Technical Solutions
Michael Bosnjak	ZPID, Trier	TA 2 Data Access

- Participants

Participants	Institution (where applicable), location
Forschungsdatenzentrum am BAMF	Bundesamt für Migration und Flüchtlinge (BAMF), Nürnberg
Forschungsdatenzentrum BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Berlin
Forschungsdatenzentrum im BIBB (BIBB-FDZ)	Bundesinstitut für Berufsbildung (BIBB), Bonn
Forschungsdatenzentrum der BZgA (FDZ BZgA)	Bundeszentrale für gesundheitliche Aufklärung (BZgA), Köln
Forschungsdatenzentrum des Sozio-oekonomischen Panels am DIW Berlin (FDZ SOEP)	Deutsches Institut für Wirtschaftsforschung e.V. (DIW Berlin), Berlin
Forschungsdatenzentrum Betriebs- und Organisationsdaten am DIW Berlin (FDZ-BO)	Deutsches Institut für Wirtschaftsforschung e.V. (DIW Berlin), Berlin
Forschungsdatenzentrum des DJI (FDZ-DJI)	Deutsches Jugendinstitut (DJI), München
Forschungsdatenzentrum des DZA (FDZ-DZA)	Deutsches Zentrum für Altersfragen (DZA), Berlin
Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung (fdz.DZHW)	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW), Hannover
Forschungsdatenzentrum Deutsches Zentrum für Integration und Migration (DeZIM.fdz)	Deutsches Zentrum für Integrations- und Migrationsforschung e.V. (DeZIM), Berlin
Forschungsdatenzentrum Bildung (FDZ Bildung) am DIPF	DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt am Main
Internationales Forschungsdatenzentrum des IZA (FDZ IZA, IDSC)	Forschungsinstitut zur Zukunft der Arbeit (IZA), Bonn
Regina T. Riphahn, Chairperson of the German Data Forum (RatSWD)	Friedrich-Alexander-Universität Erlangen-Nürnberg, Nürnberg
Forschungsdatenzentrum Wahlen bei GESIS (FDZ Wahlen)	GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim
Forschungsdatenzentrum PIAAC bei GESIS (FDZ PIAAC)	GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim
Forschungsdatenzentrum Internationale Umfrageprogramme bei GESIS (FDZ Internationale Umfrageprogramme)	GESIS – Leibniz-Institut für Sozialwissenschaften, Köln
Forschungsdatenzentrum German Microdata Lab bei GESIS (FDZ GML)	GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim
Forschungsdatenzentrum ALLBUS bei GESIS (FDZ ALLBUS)	GESIS – Leibniz-Institut für Sozialwissenschaften, Köln
LOEWE-Zentrum SAFE	Goethe-Universität, Frankfurt am Main
Forschungsdatenzentrum der Hans-Böckler Stiftung am WSI (FDZ WSI)	Wirtschafts- und Sozialwissenschaftliches Institut (WSI) der Hans-Böckler-Stiftung, Düsseldorf
LMU-ifo Economics & Business Data Center (EBDC)	Leibniz-Institut für Wirtschaftsforschung an der Universität München e.V. (ifo Institut), München

Forschungsdatenzentrum der Statistischen Ämter der Länder (FDZ-Länder)	Information und Technik NRW, Düsseldorf
Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB (FDZ BA im IAB)	Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA), Nürnberg
Forschungsdatenzentrum Archiv für Gesprochenes Deutsch am IDS (FDZ AGD)	Institut für Deutsche Sprache (IDS), Mannheim
Forschungsdatenzentrum am IQB (FDZ IQB)	Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Berlin
Forschungsdatenzentrum am IÖR	Leibniz-Institut für ökologische Raumentwicklung (IÖR), Dresden
Forschungsdatenzentrum des IWH (FDZ-IWH)	Leibniz-Institut für Wirtschaftsforschung Halle (IWH), Halle (Saale)
Forschungsdatenzentrum PsychData des ZPID (FDZ PsychData des ZPID)	Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Trier
Forschungsdatenzentrum des LIfBi an der Otto-Friedrich-Universität Bamberg (FDZ-LIfBi)	Leibniz-Institut für Bildungsverläufe e.V. (LIfBi), Bamberg
Forschungsdatenzentrum des Beziehungs- und Familienpanels (FDZ-pairfam)	Ludwig-Maximilians-Universität München, München
Forschungsdatenzentrum des Survey of Health, Ageing and Retirement in Europe (FDZ-SHARE)	Munich Center for the Economics of Aging (MEA), Max-Planck-Institut für Sozialrecht und Sozialpolitik, München
Forschungsdatenzentrum des RKI (FDZ RKI)	Robert Koch-Institut (RKI), Berlin
Forschungsdatenzentrum Ruhr am RWI (FDZ Ruhr am RWI)	Leibniz-Institut für Wirtschaftsforschung (RWI), Essen
Forschungsdatenzentrum eLabour (FDZ eLabour)	Soziologisches Forschungsinstitut Göttingen (SOFI) e.V. an der Georg-August-Universität, Göttingen
Forschungsdatenzentrum des Statistischen Bundesamtes (FDZ-Bund)	Statistisches Bundesamt, Wiesbaden
Forschungsdatenzentrum Wissenschaftsstatistik des Stifterverbandes (FDZ Wissenschaftsstatistik)	Stifterverband für die Deutsche Wissenschaft, Essen
Forschungsdatenzentrum Qualiservice (FDZ Qualiservice)	Universität Bremen, Bremen
Forschungsdatenzentrum der Universitätsbibliothek Mannheim (FDZ UB Mannheim)	Universitätsbibliothek der Universität Mannheim, Mannheim
Forschungsdatenzentrum des ZEW (ZEW-FDZ)	Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim

- Names and numbers of the DFG review boards (DFG Fachkollegien) that reflect the subject orientation of the proposed consortium

106 Social and Cultural Anthropology, Non-European Cultures, Jewish Studies and Religious Studies, 109 Educational Research, 110 Psychology, 111 Social Science, 112 Economics, 113 Jurisprudence.

2 Consortium

2.1 Research domains or research methods addressed by the consortium, objectives

The overall mission of KonsortSWD is to **support sustainable research data management (RDM) for the social, behavioural, educational, and economic** sciences in all phases of the research data lifecycle. KonsortSWD aims to provide researchers and research data centres (RDCs) with the tools and services they need (RfII, 2016) for the management and sharing of sensitive and non-sensitive data in compliance with the FAIR[1] principles (European Commission, 2018).

KonsortSWD is designed to respond to the needs of a **significant share of German academia**: about 37% of students (as of winter semester 2018/2019, Destatis 2019b, p. 21) and about **24% of professors** in German academia (as of 2018, Destatis 2019a, p. 92) study or work in KonsortSWD-related disciplines. These disciplines require specific services and a joint voice to articulate their needs, especially with respect to accessing data produced outside of research and academia.

Research domains and research methods

Research in the social, behavioural, educational, and economic sciences comprises both fundamental research and research addressing the grand societal challenges (as put forward, e.g., in the UN Sustainable Development Goals). These research areas have in common **empirical approaches to the study of human society** and therefore have **similar requirements for a research data infrastructure**. At the same time, they cover a wide range of disciplines as different as anthropology and business administration, sociology and economics, educational science and public health, or psychology and political science. Research topics include ageing, health (Grube et al. 2019, Carstensen and Jungbauer-Gans 2016), labour market dynamics (Gurr et al. 2018, Heising et al. 2019, Protsch and Solga 2015), decision-making (Dohmen et al. 2011), educational methods, ethnographic questions, social inequality (Mayer-Ahuja 2017), and many more. The research questions are addressed both comparatively (e.g., across countries) and over time (Giesselmann et al. 2018).

Researchers from these disciplines depend on comprehensive access to high-quality data, including surveys, administrative data, unstructured or new data types (e.g., web-based or

[1] KonsortSWD approaches FAIR in a broad sense: ‘FAIR is an acronym composed from Findable, Accessible, Interoperable and Reusable and therefore might be expected to be used as an adjective. [...] the FAIR principles do not just apply to data but to other digital objects including outputs of research. Additionally, making digital objects FAIR requires a change in practices and the implementation of technologies and infrastructures.’ (Hodson in European Commission 2018, p. 8).

transaction data, multimode data from computer-based assessments or learning analytics). The increasingly international scope and interconnected nature of research areas motivates researchers to rely on a research data infrastructure that is equipped with a **wide range of interlinked data (sources) and services**.

As a result of the use of different data sources, there is a **broad range of methods** for documenting and analysing these data – ranging from applied statistics (Aßmann et al. 2016), econometrics, experiments, and observations both in the field and in the lab, to interviews as well as analyses of texts, movements, and audio/video data. While survey data are typically subject to statistical analysis (Best and Wolf 2015), text data (e.g., Twitter posts or parliamentary protocols) can be analysed statistically as well as hermeneutically. Data can also contain information about networks, which can be 'extracted' and, in turn, made accessible to a specific set of statistical tools. Researchers working with experimental and (quasi-)experimental data (e.g., different regional policies in a country or variations in survey questions) use a range of methods for causal analysis (Adena and Huck 2019, 2017). Given this multitude of methods, KonsortSWD will both **consolidate RDM to provide high-quality data** for the more established methods and **pilot RDM tools** for more novel methods and data types (e.g., text or sensor data).

Specific challenges to research data management in KonsortSWD

Much of the data used for research in the social, behavioural, educational, and economic sciences are **not originally collected for scientific analysis**. Human interaction can be studied, for example, through messages people exchange on social platforms and through data collected by social insurance systems or central banks. Other data originate in academia – most notably survey data, but also test results, video interviews, and lab or field experiments. Moreover, (epi)genetic data, neuroimaging data, and medical test **data are combined** with social, behavioural, educational, and economic data to better study how societies work (e.g., Kühn et al. 2017). Furthermore, electronic sensors (e.g., in smartphones) and their near-ubiquity both offer new methodological opportunities while also creating ethical challenges (Harari et al. 2019, Bosnjak et al. 2018). Figure 1 gives an overview.

Subject	Data types	Main sharing restrictions
Individuals	Process-generated data from public and private institutions (e.g., from official statistics, social insurances, registers, scoring systems)	Privacy
	Experimental, (longitudinal) survey data	Privacy, ethical
	Behavioural data (e.g., geotracking, sensor, web, social media)	Privacy, ethical
	Qualitative interviews, observations, audio and video records	Privacy, ethical
Groups of individuals	Audio and video records	Copyright, privacy
	Qualitative interviews, observations; experimental, behavioural data (e.g., web, social media)	Privacy, ethical
Firms / organisations	Business data (e.g., earnings, patents), transaction data (e.g., stocks)	Legal, proprietary data, copyright, confidentiality
	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), regional health and climate indicators	Sometimes legal, ethical
Nations	Official statistics, aggregated survey data, texts (e.g., from political decision-making processes, parliamentary protocols), media reports	Copyright, confidentiality

Figure 1: Description of data types

The wide range of data types makes data access and RDM complex and demanding. On a technical level, RDM requires specific expertise because data come in **various formats**, including text, audio, video, qualitative interviews, survey responses, business and administrative records, statistics, data derived from scoring systems (e.g., consumer scores or grades in schools and universities), or internet usage data. Varying degrees of aggregation from the individual level to the nation state (and beyond) result in **different restrictions** on storing, using, and sharing data (cf. Figure 1). For example, data on individuals can affect privacy, while business data or documents from political decision-making processes can raise questions of confidentiality. Apart from legal issues - such as copyright - data also need to be used ethically, especially when dealing with qualitative (e.g., unstructured textual, audio, or video data, often small sample sizes) and new forms of behavioural data. KonsortSWD will build on the established RDC model to facilitate access to such sensitive data for academic researchers.

Research data centres (RDCs) make up **an infrastructure that successfully accommodates the different demands** of researchers and that has continuously adjusted to new requirements: They are considered to be ‘best practice’ by Germany’s Council for Scientific Information Infrastructures (RfII 2016, pp. 30-33, cf. also Wissenschaftsrat 2011, pp. 81-82) and have played a growing role in the German scientific system during the last two decades. Most RDCs are sustainably funded by their host institutions. By accrediting and monitoring such

RDCs, the German Data Forum (RatSWD) has created a highly evolved and **sustainable research data infrastructure** (RfII 2016, pp. 30-33, Wissenschaftsrat 2011, pp. 81-82). This network of currently 34 RDCs is a result of numerous initiatives from within the scientific community and guarantees user-friendly and cost-effective access to a wide range of relevant data for researchers. To maintain and to develop this infrastructure further, KonsortSWD proposes to integrate the German Data Forum (RatSWD) and the RDC network into NFDI because it can no longer be supported on project-based funding by the Federal Ministry of Education and Research (BMBF).

Objectives

Our mission is to develop – strengthen, widen, and deepen – a user-driven research data infrastructure for the study of human society that meets the requirements of the targeted research communities. Therefore, KonsortSWD supports **sustainable RDM** and **data accessibility** with an awareness of ethical and legal aspects**.** The members of KonsortSWD have set the following main objectives for their cooperation within NFDI:

1. **Support sustainable RDM** for the social, behavioural, educational, and economic sciences in all phases of the research data lifecycle and provide researchers and RDCs with the tools and services they need for managing and sharing sensitive and non-sensitive data in compliance with the FAIR principles.
2. **Respond to new challenges** posed by the growing number of data types and sources by
   a. **making data available** that were not initially gathered for research purposes or are not yet available for secondary use (especially qualitative data),
   b. **providing secure access** to additional sensitive data from both public and private sources,
   c. **linking data** (e.g., by combining survey data with spatial, text-based, medical, neuroimaging, or genetic data).
3. Improve **appreciation for the value of data sharing** through training, information, and reliable citation procedures with persistent identifiers. Stimulate a culture of data sharing that takes into account different stages of accessibility, conditions, and requirements given in the various disciplines and methodological schools, data types, and demands concerning the reputation of the researchers.
4. Safeguard and enhance the quality and **FAIRness of (sensitive) data**, including documentation, archiving, privacy, and security, by building on the success of RDC accreditation. Coordinate data and metadata generation as well as standardisation to achieve interoperability across the different data types and academic disciplines represented by KonsortSWD.

5. **Foster skills development** to meet future requirements resulting from the ongoing digital transformation of research; strengthen competencies in data management and data use in research within the RDCs and among data users. Raise awareness of ethical challenges and develop an appropriate infrastructure for meeting them.
6. Gather support for making data, methods, and **research results openly accessible** (Open Science). Where legal or ethical restrictions apply: enhance availability of privacy-sensitive and other restricted data for research purposes.
7. **Act as a national and international voice** and point of contact for RDM in the social, behavioural, educational, and economic sciences.
8. Connect and continuously expand existing activities to **reach all branches of the consortium and related consortia** (e.g., in the field of public health and the humanities).
9. Set up a **flexible and efficient consortium structure**: continually develop the service portfolio and make room for contributions of new partners.

2.2 Composition of the consortium and its embedding in the community of interest

KonsortSWD addresses empirical researchers studying human behaviour. While the disciplines represented in KonsortSWD differ in subject matter and research focus, the methods they use are comparable, thus justifying a broad, yet inclusive subject-oriented approach. KonsortSWD has analysed the existing research and infrastructure landscape using numerous sources and means (detailed below). Three characteristics stand out:

First, KonsortSWD builds on an **existing research infrastructure** of **accredited research data centres (RDCs)** (RfII 2016, pp. 30-33, Wissenschaftsrat 2011, pp. 81-82). RDCs are **user-driven, bottom-up** initiatives concerned with improving access to (sensitive) research data for the disciplines covered by KonsortSWD. In 2018, the RDCs held 3,940 datasets, counted 71,488 downloads of unrestricted datasets (cf. RatSWD 2019a), and attracted over **1,500 researchers** and infrastructure staff to their training courses in 2017 (cf. RatSWD 2018a). The decentral RDC network of currently 34 RDCs provides researchers with user-friendly and cost-efficient access to a wide range of relevant and predominantly sensitive data. **RDCs share metadata tools** for documenting diverse types of data to enhance data comparability and quality. Most utilise proven data **repository and search technologies** to increase findability and interoperability and use da|ra, a common **data registration service** for all RDCs (cf. section 3.1 Metadata Standards). RDC **accreditation** fosters data sharing and high-quality data management based on common (accreditation) guidelines (Bug et al. 2018, Wissenschaftsrat 2011, pp. 81-82). The RDC network has also established standards for **self-monitoring** and **impact measurement**. It is a key asset of KonsortSWD.

Second, KonsortSWD brings to NFDI the **German Data Forum (RatSWD)**, a highly acclaimed coordinating, quality assuring, and advocating body for both researchers and data producers (cf. RfII 2016, pp. 30-33, Wissenschaftsrat 2011, pp. 81-82). Data providers and elected data users each account for half of its 16 members – from July 2020 onwards 20 members. Representatives of the Federal Ministry of Education and Research (BMBF), the *Länder* and the German Research Foundation (DFG) had and will continue to have guest status (RatSWD 2019d, §3). The German Data Forum (RatSWD) pools and develops expertise on research data infrastructures to ensure and continuously improve findability, accessibility, interoperability, and re-usability of research data (cf. Task Area 1, Measure 1 (TA.1-M.1 RatSWD)). Among the data producers are representatives from official statistics agencies, social insurance providers, and research-driven data production who represent relevant data sources for KonsortSWD's disciplines.

Third, KonsortSWD cooperates with a broad **network** of academic associations, specialised information services (SISs, *Fachinformationsdienste*), and university libraries. Furthermore, the German Data Forum (RatSWD) has established a network of decentral **research ethics committees**, which form the basis of a future research ethics infrastructure (cf. TA 4 Research Ethics). In addition KonsortSWD is well integrated into European and international activities around research infrastructures and RDM in the disciplines it represents.

The **needs of the community were identified** in several ways.

To engage the research communities more broadly and to discuss their expectations for KonsortSWD, the German Data Forum (RatSWD) invited 15 **academic associations** to a meeting in June 2019 (RatSWD 2019b). This meeting made important contributions to **Task Area 1 Community Participation**: it clarified that KonsortSWD needs to assure participation of academic associations beyond the opportunities that the German Data Forum (RatSWD) could offer so far. Users will thus be directly represented in several elements of KonsortSWD's governance structure. Additionally, academic associations are actively integrated into the service development and implementation planning of KonsortSWD through Measure TA.1-M.2 (Community AA).

The community of libraries and SISs are increasingly addressing RDM issues. In order to learn more about the specific needs of SISs, the German Data Forum (RatSWD) organised a meeting with representatives of the eight SISs most relevant to the disciplines covered by KonsortSWD

in Cologne in March 2019. The discussions helped shape the Measures in **Task Area 1 Community Participation**, in particular, Measure TA.1-M.3 Community SIS.[2]

Moreover, the accredited RDCs are in **permanent exchange** with their **more than 46,000 data users** (RatSWD 2019a). A large survey amongst the RDCs helped identify areas for deeper cooperation. For the last two years, the RDCs planned this cooperation in four working groups and used their semi-annual forum to decide which services can be advanced to better meet data users' needs. The ideas developed by the RDCs were incorporated in **Task Area 2: Data Access**.

The German Data Forum (RatSWD) organised further networking workshops to discuss the needs and requirements of different communities: a series of preparatory meetings began with representatives of large social science surveys in December 2017 (RatSWD 2017c).[3] In April 2018, an interdisciplinary workshop addressed challenges for archiving and re-using qualitative research data (Hollstein and Strübing 2018). These activities are the base for the Measures in **Task Area 3: Data Production**.

Task Area 4: Research Ethics builds on long-standing activities of the German Data Forum (RatSWD). During its fifth appointment period (2014-2017), it prepared recommendations on the need for research ethics in its disciplines and on setting up an infrastructure for research ethics (RatSWD 2017a). During its sixth appointment period (2017-2020), in November 2018, it brought together interested ethics committees from different institutional backgrounds with other players, such as academic associations. The participants discussed the fundamental ideas underlying the proposed Measures in TA 4.[4] Subsequently, the content of TA 4 was distributed to and commented on by the participants, other experts, and potential users from different institutional backgrounds with responsibilities in research ethics.

Task Area 5: Technical Solutions builds on established research and development activities by several of KonsortSWD's partners in the areas of information systems and research data infrastructures as well as the principles of user-centred design.

---

[2] For documentation, please see https://www.ratswd.de/dl/media/190829_RatSWD-Vernetzungstreffen-Bib_und_FID-Dokumentation.pdf (28 August 2019).

[3] Subsequent discussion with representatives of panel surveys in German language is documented here: https://www.neps-data.de/de-de/datenzentrum/panelworkshop2018.aspx (28 August 2019); https://fdz.dzhw.eu/panelworkshop2019 (28 August 2019).

[4] For documentation, please see https://www.ratswd.de/dl/media/190924_RatSWD-Vernetzungstreffen-Ethik.pdf (28 August 2019).

The Measures presented in this proposal were **prioritised** in continuous exchanges between stakeholders, technology experts, and the German Data Forum (RatSWD). Key evaluation criteria were their added value for researchers and their contribution to making data FAIR.

KonsortSWD pursues a **wide-ranging approach** to ensure fruitful interaction between **users and service providers**: 1) The currently 34 RDCs interact with their users in many ways. Typically, RDCs have user councils and advisory boards, administer user surveys, and are **evaluated** based on the quality of their service, for example, in regular evaluations by the Leibniz Association (cf. RatSWD 2017b, 2018a, 2019a). 2) In order to continuously help users to participate and articulate their demands around RDM, KonsortSWD devotes TA 1 Community Participation to community participation. 3) The German Data Forum (RatSWD) has a long history of **bottom-up, user-oriented activities** with a focus on research data access. The currently eight representatives of data users in the German Data Forum (RatSWD) are elected for three years. In 2017, more than **3,000** established **researchers participated in the election** (individuals with a doctoral degree employed at a university or another independent research institution). User involvement will be expanded further by increasing the size of the German Data Forum (RatSWD) from 16 to 20 members (ten of which will be elected) in 2020. Members of the German Data Forum (RatSWD) will serve in the Advisory Board (AB) to the Steering Committee (StC) (cf. section 2.5). 4) The StC will monitor and prioritise the Measures according to insights gained in TA 1 Community Participation. User perspectives are represented in the StC via the chair of the German Data Forum (RatSWD) and the officers in charge of TA.1.-M.2 Community AA and TA.1-M.3 Community SIS, who participate as guests (cf. section 2.5).

Participating institutions: A strength of KonsortSWD is that it builds on contributions from a large number of experienced partners who have fruitfully cooperated in various infrastructure projects for many years. As such, the group has ample experience in open data documentation and archiving and can guarantee the sustainability of its activities. The cooperation will continue within KonsortSWD to create innovative services using the experience of all partners. The consortium members will contribute the following strengths:

1) The Leibniz Institute for the Social Sciences (**GESIS**), the applicant institution, is the largest infrastructure provider for the social sciences in Europe. GESIS has developed certified long-term repositories, is experienced in establishing new data-related services, and runs a data archive. GESIS also runs datorium, a platform for data and code sharing for replication purposes.[5] **2)** The Leibniz Information Centre for Economics (**ZBW**) hosts the GO FAIR office

[5] https://datorium.gesis.org/xmlui/ (28 August 2019)

for Germany and is a driving force behind establishing the FAIR principles. The Generic Research Data Infrastructure project (GeRDI)[6] aims at the management of research data across disciplines. It has developed software to connect research data repositories and to provide access to services. **3)** The Leibniz Institute for Psychology Information (**ZPID**) provides PsychArchives, a repository for psychological science, which preserves a variety of digital research objects, including 21 different publication types, research data, tests, pre-registrations, multimedia, and code. It provides access according to the FAIR principles. **4)** The Berlin Social Science Center (**WZB**) is a respected research hub with a far-reaching network throughout the social sciences. It supports ethical reflection in research, has a research ethics committee and has been administering the German Data Forum (RatSWD) since 2016. **5)** The Leibniz Institute for Research and Information in Education (**DIPF**) jointly coordinates the German Network for Educational Research Data (VerbundFDB)[7] with GESIS, the Institute for Educational Quality Improvement (IQB), and other relevant partners (cf. section 3). This federated infrastructure allows educational researchers to submit, archive, and re-use data in a trustworthy and quality-assuring way. **6)** The Leibniz Institute for Educational Trajectories (**LIfBi**) makes the National Educational Panel Study (NEPS) available to a large international community. The RDC at LIfBi has extensive expertise in editing and disseminating complex data. RemoteNEPS offers remote access, especially for the analysis of sensitive data. **7)** The RDC **Qualiservice** at the University of Bremen covers the full range of qualitative social science research data, such as interview transcripts, audio, and video data, field notes, and observation protocols. Together, Qualiservice and the University Library of the Humboldt University Berlin manage the Specialised Information Service Social and Cultural Anthropology (SIS SCA) funded by the German Research Foundation (DFG). **8)** The Sociological Research Institute at Georg-August-Universität Göttingen (SOFI) hosts **eLabour**, an interdisciplinary centre for qualitative research data from the sociology of work. The centre will expand its qualitative data pool and make it available to external scientists through its RDC. **9)** The RDC at the German Centre for Higher Education Research and Science Studies (**fdz.DZHW**) provides access to quantitative and qualitative research data from the field of higher education and science studies. Its data pool contains multi-cohort in-house panel surveys as well as external research data. **10)** The RDC of the German Socio-economic Panel Study at the German Institute for Economic Research (DIW Berlin) (**SOEP**) provides rich data to an international research community. Established in 1984, the SOEP household panel has since become an international role model. SOEP's RDC operates at the forefront of innovative data and infrastructure development, for example, in

[6] www.gerdi-project.eu; www.demo.gerdi.org (28 August 2019)
[7] https://www.forschungsdaten-bildung.de/ (28 August 2019)

linking survey-data to administrative data sources and enriching social science data with genetic and epigenetic data. **11)** The **University of Duisburg-Essen** is home to an innovative research group in computational social science (**PolMine**). The group works with text-based data and develops text-mining methods and software architectures for the empirical analysis of corpora, such as parliamentary minutes.

While KonsortSWD has most required expertise within its ranks, it has had a long-standing practice of obtaining additional (e.g., specialised legal) advice, from its expert networks. Furthermore, KonsortSWD will actively seek out new partners when the communities demand new services. The RDCs listed as participants (accredited RDCs as well as applicants in the accreditation process) of KonsortSWD provide ample expertise.

2.3 The consortium within the NFDI

KonsortSWD brings its experience in the design and operation of a research data infrastructure to the NFDI that is deeply embedded in scientific communities and aligned with the needs of scientific data producers and data users. This is based on more than 15 years of experience in setting up and operating research data centres (RDCs), which are considered ‘best practice’ (RfII 2016, pp. 30-33, Wissenschaftsrat 2011, pp. 81-82). Thus, a wide range of FAIR research data services has already been developed. Within NFDI, KonsortSWD can support the development and expansion of this infrastructure to add value, especially in communities without established structures. KonsortSWD makes unique contributions to the NFDI in the following areas:

- Expertise in **sustainable archiving**, technical, and administrative **models for sharing** sensitive or restricted research data, and the development of linked infrastructures located at research institutions and universities.
- A comprehensive **portfolio** of methodological and technical **services** and measures to support researchers in documenting and archiving their self-generated research data.
- Experience in **linking research data** generated both outside and within academia.
- A **network of local ethics commissions** as a basis to systematically build expertise in research data ethics.
- Proven (administrative) models for **building** and operating **research infrastructures** at universities and non-university research institutions.
- Professional **training programmes** on RDM and data-driven research methodology for both researchers and students.

- Proven models for **training scientific and non-scientific personnel** in data-providing infrastructure facilities – including training in Germany's dual apprenticeship system (the so-called 'specialists for market and social research').
- Based on the experience of the German Data Forum (RatSWD) as an independent voice of its research community and an advisor to legislators and science institutions, KonsortSWD can contribute to NFDI's expertise in **advocating** sharing research data.

Together with ten other applicants (NFDI4BioDiversity, NFDI4Earth, NFDI4Chem, NFDI4Health, NFDI4Agri, NFDI4Crime, NFDI4Medicine, DataPLANT, GHGA, Astro-NFDI), KonsortSWD has consolidated a list of **15 cross-cutting topics** that need to be addressed. The Berlin Declaration (Glöckner et al. 2019) has won support from 21 signatories. KonsortSWD has offered to contribute to all of the topics and will actively support the **cross-consortium discussion on ethical aspects** of sharing research data.

Building on KonsortSWD's experience in linking and merging different types of data, we will make more data and new data types available by cooperating with other NFDI consortia. Beyond our commitment to the topics in the Berlin Declaration, we see potential for **cooperation with the following consortia** (in alphabetical order). Measures TA.2-M.6 (RDM Grants) and TA.6-M.3 (Network Refinement) will provide financial means for these efforts:

BERD@NFDI: Cooperation is intended to cover RDM and merging both structured and unstructured (big) data as well as training of social data scientists. ZBW, one of KonsortSWD's co-applicant institutions, is also engaged in BERD@NFDI.

ForumX will join the data access network of KonsortSWD once it has established its RDC services. It will benefit from the repository experience among KonsortSWD's RDCs. Furthermore, collaboration with the RDC PsychData can lead to the harmonisation of metadata from experiments. SOEP opens the validation of experimental results to members of ForumX within the framework of survey experiments. ForumX will make contributions to the repository framework of KonsortSWD regarding RDM of experimental data from different disciplines and profit from access to the broader data infrastructure that KonsortSWD coordinates.

NFDI4BioDiversity and KonsortSWD aim to collaborate in the areas of geodata use and the handling of sensitive data. KonsortSWD will supply social science data relevant to biodiversity research and give support in dealing with sensitive data, in particular, anonymisation services.

NFDI4Agri and KonsortSWD will explore ways to link data on land-use and food production with socio-demographic, economic, and business data. Cooperation promises to be particularly

fruitful with respect to enriching data held by NFDI4Agri with data held by KonsortSWD's RDCs and to exchanging expertise on qualitative interviewing. The latter is frequently used in the agricultural sciences and is also part of KonsortSWD's proposal. Furthermore, the two groups will cooperate on geodata together with NFDI4Earth.

NFDI4Earth: We aim to establish interoperability between geodata and survey data. The collaboration is planned to be three-fold: 1) To create geodata files that can be used easily in the social sciences. 2) NDFI4Earth offers its competence in using geodata. 3) KonsortSWD provides its expertise regarding anonymisation to NFDI4Earth.

NFDI4Health, NFDI4Medicine, and KonsortSWD will coordinate their efforts to make sensitive cohort and survey data both re-usable and interoperable. Participants of NFDI4Health and NFDI4Medicine aim to join the Economic and Social Sciences goINg FAIR Implementation Network (EcoSoc-IN) in order to join forces directly under the auspices of the Go FAIR initiative.[8]

Text+ and KonsortSWD share an interest in text-based unstructured data. A pilot Measure (TA.3-M.3 Textual Data) on text-based research data will collaborate closely with and use metadata and experience from Text+. KonsortSWD will offer its expertise regarding anonymisation. Andreas Blätte, co-applicant within KonsortSWD, is also a participant in the Text+ consortium.

We also envision cooperation with consortia planning to join NFDI in later rounds:

BRIDGE4NFDI and **2LINK4NFDI**: KonsortSWD's experience in collaboration between data centres and consortia may be helpful for BRIDGE4NFDI and 2LINK4NFDI to develop their services. KonsortSWD aims to use the services generated by these two cross-cutting consortia.

LoReData can profit from better integration of geodata and regional data that is accessible via several RDCs 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 RDM needs that are methodologically close to social research (such as surveys) and that share its sensitive nature. KonsortSWD can support NFDI4Objects regarding the legal and ethical re-use of sensitive research data.

[8] https://www.go-fair.org/ (13 September 2019)

NFDI Web: The web archive to be provided by NFDI Web is explicitly intended as a source for social scientists. KonsortSWD will explore cooperation prospects for analysing unstructured text data contained in web sources. KonsortSWD is also interested in linking long-term web data to surveys and economic data.

KonsortSWD expects from its integration into the NFDI above all a substantial expansion of the supply of research data and thus of the possibilities of data-based research in the disciplines it represents. This applies in particular to linking data from KonsortSWD RDCs with existing data from humanities, public health, and geosciences. Through the cooperation in the field of text-related and unstructured research data, we expect impulses for the further development of research data management and the use of innovative analysis methods in the communities of the social, behavioural, educational, and economic sciences, which primarily make use of qualitative or mixed methods. We will contribute to efforts within NFDI to develop common standards and workflows to improve the methodological and technical quality of research data. We see an urgent need for action within the NFDI in the areas of data linkage and data quality.

2.4 International networking

Many RDCs and **KonsortSWD partner institutions are active in international research** and cooperate with other infrastructures. KonsortSWD will base its activities on this strong network to assure that its services are compatible with international developments and contribute to these.

Most notably, **RDC have a significant proportion of users from outside Germany**. For example, 64% of the 15,000 data users of the GESIS data archive in 2018 and about 45% of the SOEP users were not based in Germany. To attract international users, two thirds of the RDCs provide English language contracts, data files, and data documentation (RatSWD 2018a). RDC SHARE and RDC International Surveys (with data from the International Social Survey Programme, the European Values Study, the Comparative Study of Electoral Systems, and the Eurobarometer) primarily serve international user communities.

Given the rising importance of open data infrastructures at the European level, KonsortSWD data and services aim to participate in the emerging pan-European data infrastructure. Aligning the NFDI with the European Open Science Cloud (EOSC) will be an important task for existing service providers. **KonsortSWD will ensure that its data, tools, and services are visible at the international level** and to increase the consortium's impact. The EOSC is particularly relevant for KonsortSWD's international networking strategy because it will set future standards for research data management (RDM) in the European Research Area (ERA). The strategic

implementation plan of the European Commission for the EOSC underscores that it will gradually integrate existing infrastructures into 'a FAIR internet of data, tools, and services' (Budroni et al. 2019) instead of creating new infrastructures.[9] KonsortSWD is working to improve the visibility of its existing infrastructure within the EOSC.

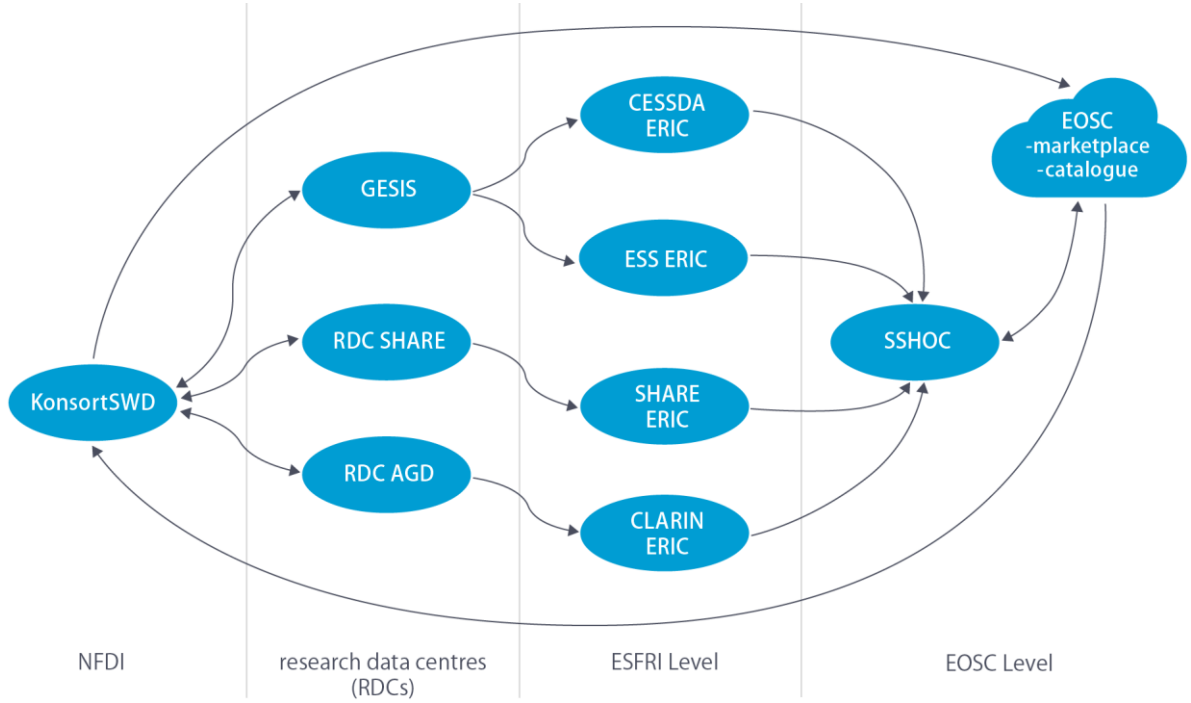

Figure 2: Links of KonsortSWD participants to ESFRI and EOSC

Lifting KonsortSWD to the EOSC: gaining visibility and facilitating transnational access

Members of KonsortSWD already represent Germany in all four social science and humanities European Research Infrastructure Consortia (**ERICs**), as shown in Figure 2. GESIS represents Germany in the Consortium of European Social Science Data Archives (CESSDA). CESSDA ERIC federates 18 data centres from European countries and develops common services for data discovery (e.g., the CESSDA data catalogue[10]) and RDM (e.g., the CESSDA RDM expert guide[11]). GESIS hosts the CESSDA Training support unit (CESSDA Training Working Group 2017-2018) and the CESSDA Metadata Office. The European Social Survey (ESS) ERIC is an international survey programme for the social sciences. The Survey of Health, Aging, and Retirement in Europe (SHARE ERIC) fills an important data gap in research on aging societies. It covers 27 European countries as well as Israel. RDC SHARE is responsible for data

[9] This means that the infrastructures will continue to be funded nationally and the existing national and international infrastructures will align their data, tools, and services with the EOSC (see Jones et al. 2019).

[10] https://datacatalogue.cessda.eu/ (16 September 2019)

[11] https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide (16 September 2019)

documentation and data dissemination for the whole ERIC (including all international users). The Leibniz-Institute for the German Language (IDS) – which hosts the RDC Archive of Spoken German (AGD) (see also appendix 3) – is one of 20 certified Service Providing Centres of the Common Language Resource and Technology Infrastructure (CLARIN) ERIC. CLARIN focuses on tools for automated text annotation and analysis. These tools are important for many social science research topics that deal with unstructured textual data (cf. TA.3-M.3).

The existing ERICs from the social sciences and humanities (SSH) have formed a common consortium, coordinated by CESSDA ERIC, creating the Social Sciences and Humanities Open Cloud (SSHOC) within the EOSC (CESSDA is a member of the EOSC executive board). EOSC has established EOSCportal, which hosts two platforms for services: a catalogue and a marketplace. Services are only listed on the EOSC marketplace after passing a quality review, whereas registering services for the catalogue does not require meeting specific criteria. The SSHOC project currently plans to create a marketplace that will be integrated into the EOSC marketplace. The SSH ERICs will use the SSHOC marketplace with their RDM services and data holdings. Data from KonsortSWD's RDCs will automatically be included. RDM tools and services can be directly registered on the EOSC marketplace without passing through SSHOC or an ERIC. KonsortSWD will therefore:

1. **Register KonsortSWD tools and services** with the **EOSC** and SSHOC marketplaces. The objective is to attract international users to the services and gain visibility as a service provider. Potential candidates for the international promotion of services are: data holdings from RDCs (including transnational access to sensitive data), persistent identifier (PID) services for the disciplines covered by KonsortSWD, and RDM training for data stewards and data managers.
2. **Contribute to forming an international working group with all SSH ERICs to foster transnational access to sensitive data.** Existing initiatives from official statistics, such as the International Data Access Network (IDAN), will be invited to join forces to design a collaborative Horizon Europe Infrastructure (INFRAEOSC) project.
3. Become a partner in INFRAEOSC projects. Currently, the EOSC's implementation is driven by large-scale EU projects (EOSCpilot, EOSChub) and dominated by actors from high-performance computing facilities. To increase visibility within the EOSC, KonsortSWD will take part in international networking events such as the annual EOSC stakeholder meetings, the Digital Infrastructure for Research (DI4R) conferences, the Research Data Alliance (RDA) plenary meetings, and the International Open Science Conference. KonsortSWD will present the scope of its data, its service portfolio, and its technical solutions for RDM. **A primary objective is to become a social science use case partner for European**

infrastructure projects. A second objective is to identify potential partners to initiate EU projects on core issues of KonsortSWD, for example, securing remote transnational access to sensitive data.

Internationally designing FAIR data infrastructures

The German Data Forum (RatSWD) has created the Economics and Social Sciences goINg FAIR Implementation Network (EcoSoc-IN), an open network to discuss solutions for the definition and implementation of FAIR data infrastructures with particular attention to issues arising from sensitive data (see section 2.2). EcoSoc-IN is part of the international networking strategy to **ensure compliance of KonsortSWD's services and solutions with a FAIR data infrastructure** (see section 3.2.2). KonsortSWD will therefore invite European actors to join EcoSoc-IN as an international working group on FAIR data infrastructures in the social sciences to design prototypical FAIR data infrastructures and the interfaces for metadata and data exchange (TA.5-M.3 Interfaces).

Furthermore, DIPF, GESIS, ZBW, and ZPID are members of the GO FAIR Implementation Network GO Inter[12] which aims at fostering semantic interoperability of heterogeneous research data across domains (see section 3.3).

2.5 Organisational structure and viability

Three principles guide KonsortSWD's organisational structure: 1) facilitating communication and the flow of ideas, 2) providing an efficient governance, and 3) keeping all stakeholders informed and involved.

Adhering to these principles, KonsortSWD's organisational structure will have four elements: the **Steering Committee** (StC), the **Advisory Board** (AB), the **Task Areas** (TAs) as well as a the **Konsort Assembly** (KA). Together, these elements ensure that **both data users and providers** are prominently represented.

1. The StC is composed of the consortium's spokesperson and representatives from all TAs, except for TA 6 which is represented by the spokesperson. The **StC is KonsortSWD's executive body and responsible for all operative and strategic decisions**. It implements the work programme and continually adjusts it to developments within the NFDI. All TAs choose a representative and a deputy from their active members (involved in the Measures) of KonsortSWD. Whenever possible, one of the two should be a data user, while

[12] https://www.go-fair.org/implementation-networks/overview/go-inter/ (13 September 2019)

the other should represent the data producers. To ensure a strong voice of KonsortSWD's users in the StC, TA 1 (Community Participation) is additionally represented in the StC by the chair of the AB, the chair of the Committee for Data Access (CDA), and the head of TA1's staff, all of whom may sit in as guests.

2. **The AB advises the StC in matters of strategic importance.** It discusses and approves the consortium's work programme and issues recommendations on its future development. The AB is also involved in the selection of future Measures (cf. TA.6-M.3 Network Refinement). The AB convenes at least once per year. It is autonomous in setting its agenda and bringing in users' perspectives into KonsortSWD's developments. It can invite the spokesperson and other members of the StC as guests. All members of the German Data Forum (RatSWD) (for details see TA.1-M.1 German Data Forum (RatSWD)) who are not members of the StC are ex officio members of the AB. Thus, the elected members represent users of KonsortSWD's services.
3. **TAs coordinate the Measures, which pertain to a defined set of services.** They establish and maintain a dialogue between data users and data producers. Twice a year, the heads of all Measures meet with their TA's heads to exchange results, discuss potential synergies, plan next steps; and assess emerging risks. These meetings will form the basis for brief reports and recommendations to the StC.
4. Members of the KA are all (co-)spokespersons and participants receiving funds from KonsortSWD. **The KA meets annually to discuss matters of general importance to KonsortSWD.** It can make direct recommendations to the StC.

The KonsortSWD ecosystem brings together numerous stakeholders and is part of the research community (shown as the light green circle in Figure 3). The green colours represent channels of immediate user integration whilst the blue colour represents areas where concrete services are developed. It offers a clear structure for the efficient interaction of a variety of actors to guarantee swift operations and reduce complexity. All the structural elements are designed to systematically involve the users.

Communication and quality assurance are safeguarded through mutual guest status settings in the different bodies. Paying special attention to new communities and new user demands, **KA and AB also give feedback on current work and future plans.**

Adding to KonsortSWD's symbolised ecosystem, Figure 4 illustrates how elements of KonsortSWD work together. Since the elements depend on one another, regular communication and coordination will be necessary. As the **democratically legitimised** user representation in KonsortSWD, the AB is set up to shape KonsortSWD activities.

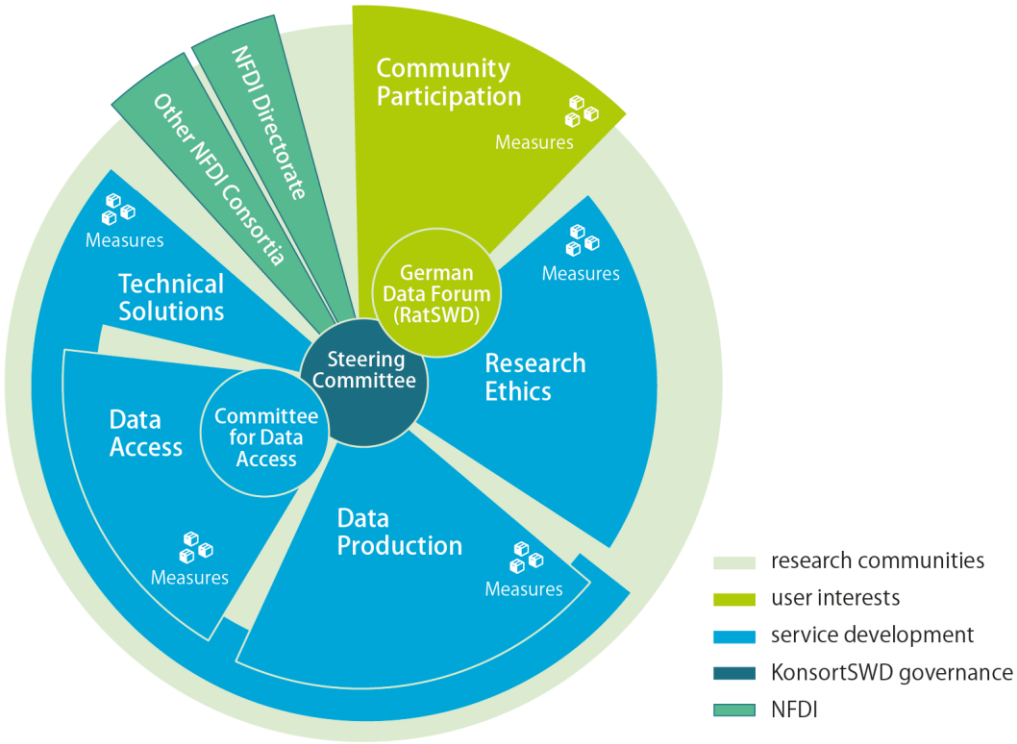

Figure 3: KonsortSWD's ecosystem

Communication within KonsortSWD takes on various forms ranging from informal exchange, to discussions of progress reports and user feedback in meetings, to formal reporting. Communication among the Measures within the TAs is given a formal setting in the biannual TA meetings. Communication between the TAs occurs on a regular basis during the StC meetings. **Cooperation also takes place at different levels:** within the Measures, between the Measures at the TA-level, between TAs in the StC, and based on reports from the StC to the AB and the KA.

In line with the German Research Foundation's (DFG) funding criteria, the responsibility for attaining KonsortSWD's objectives lies with the spokesperson. **Authority for operative and strategic decisions rests with the StC.** It establishes reporting and monitoring procedures with the TAs to be informed about the progress within individual Measures. In this context, the Committee for Data Access (CDA) plays a crucial role, as the RDCs have permanent contact to their users (which turns them into 'specialists of user perspectives') and some service development within KonsortSWD is tailored for their use as well. Further feedback from user representatives is systematically gathered at least annually via the AB and for ongoing decision-making through the different representatives of TA 1 Community Participation in the StC. Figure 4 shows the central role of community and user integration and the multiple user representation within KonsortSWD's ecosystem.

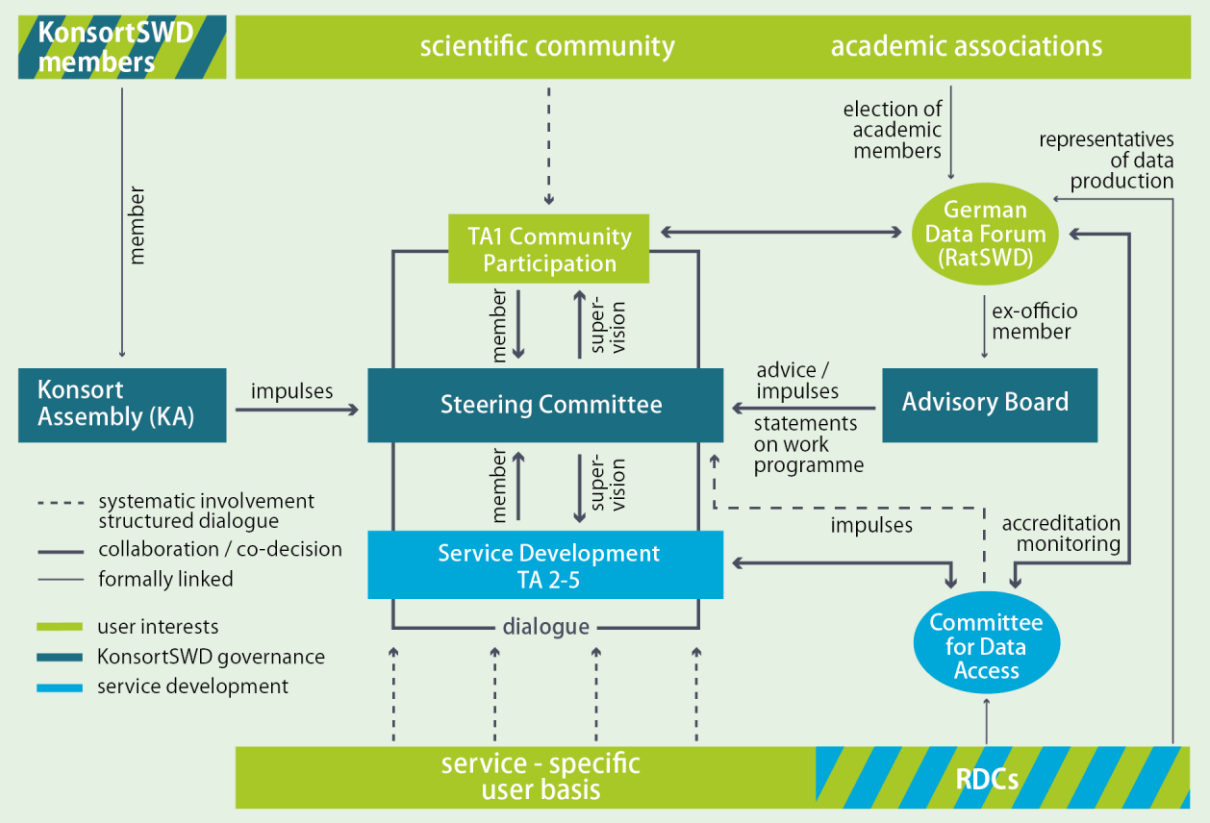

Figure 4: Organisation chart

All (co-)applicant and participant institutions that are funded by KonsortSWD will **contractually agree** on the organisational structure sketched out here.

Internal disbursement

The StC coordinates internal disbursements. Decisions are made by a qualified majority of two thirds of the six voting members (spokesperson plus five TA representatives). **Input for new Measures or changes to planned services can come from all constituent groups of KonsortSWD.** However, in order to ensure user orientation and community-driven service development, the AB must agree to any such proposals. Beyond financial arrangements as implied by contractual obligations, the StC cannot substantially change financial arrangements without a recommendation by the AB.

Viability

Recognising that user demands for research management services will change and that the NFDI's structure is set to adapt, KonsortSWD's governance contains mechanisms to add new Measures for service development and operation. **Both, the addition of new services and their potential termination will be determined on the basis of user feedback** (such as letters of support, usage statistics and direct user feedback on services, or feedback from

academic associations) and recommendations from the AB. As the German Data Forum (RatSWD) will continue issuing recommendations to the communities, these recommendations can trigger the adjustment of services. The proposed Measure Network Refinement (cf. TA.6-M.3) provides the necessary financial resources.

Since the German Data Forum's (RatSWD) has a long history of community participation and most institutions participating in KonsortSWD have a long history of successful cooperation, KonsortSWD considers its operation to be viable. To maintain flexibility under changing circumstances, KonsortSWD incorporates provisions into its legal agreements for adding new co-spokespersons or institutions and to terminate cooperation with or membership in KonsortSWD. The AB will emphasise the requirement for the TAs' services to be reliable and durable. **TAs must set up contingency measures to guarantee the permanent availability of services, and to allow for service adjustments.** These decisions require prior recommendation or ex post approval by the AB to make sure the perspective of the users is incorporated. Decisions in the StC will be made by a qualified majority of two thirds.

2.6 Operating model

KonsortSWD will fully comply with German Research Foundation (DFG) regulations and non-profit requirements (*Gemeinnützigkeit*). As the lead institution, GESIS will pass on funds to other consortia members based on contracts describing the cooperation framework. The contracts will contain DFG's spending conditions and ensure compliance. They will also define the contributions, responsibilities, and the distribution of overhead along the lines of this proposal. All participant institutions receiving funds are required to be non-profit entities. KonsortSWD will limit other transactions to avoid paying value-added tax through contract work.

Material results (software, documentation, training materials, etc.) will be made available as open-source or open-access products free of charge. Financial transfers will be made only to allow recipients to create new services by scientific means or involving scientific expertise. All (co-)applicants and participants will work on their own account and according to the requirements of their respective legal status. KonsortSWD will not set up a new legal entity for the consortium.

KonsortSWD's partners are academic research institutes, universities, individuals, and entities representing RDCs (e.g., official statistics agencies). **All partners have strong links to data users** through user conferences, user advisory boards, user monitoring, and user surveys. The partners in KonsortSWD are aware that the services developed might need to be transferred to

an NFDI entity, once created, in the coming years. They will support this process within the legal and statutory limits they are subject to.

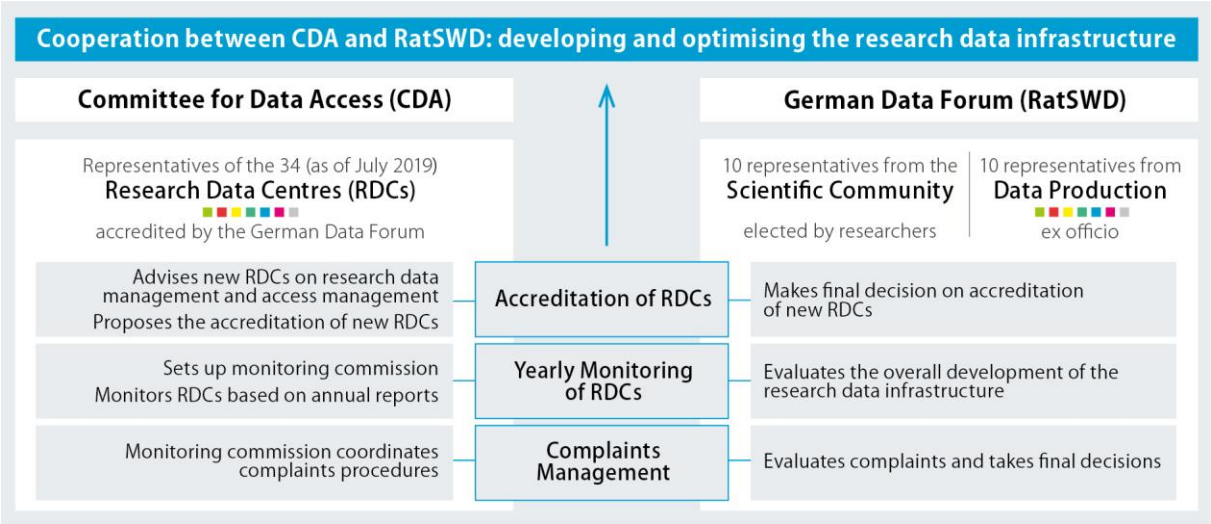

Figure 5: Quality assurance: Cooperation between CDA and German Data Forum (RatSWD)

KonsortSWD's operation will profit from the infrastructures of its member institutions and the RDCs: the German Data Forum (RatSWD) will continue to accredit RDCs and monitor their operations (RfII 2016, pp. 31-32, RatSWD 2018b). **Quality assurance will be continued and further developed** in KonsortSWD, see also Figure 5. On this basis, the data services provided by RDCs are part of KonsortSWD's infrastructure. In addition, KonsortSWD will benefit from existing RDC metadata tools for documenting diverse types of data. These are often open-source tools for best practice in RDM allowing the creation of easy-to-analyse data products. Furthermore, KonsortSWD partners already have curricula for affordable training for (future) researchers as well as infrastructure personnel at all stages of their careers (RfII 2019). Finally, KonsortSWD will benefit from the previously established cooperation with decentral research ethics committees.

Beyond RDCs, **consortium members**, many of which are infrastructure providers within the Leibniz-Association, also **operate infrastructure services available to KonsortSWD**: examples are Data Archives for cross-sectional and longitudinal (survey) data (LIfBi, GESIS, SOEP), a data archive for audio and video data (DIPF), proven data-repository and search technologies to promote findability and interoperability (DIPF, GESIS, ZBW), da|ra, a general DOI registration service for all RDCs (operated by GESIS and ZBW), metadata tools and metadata expertise (GESIS is member of the Data Documentation Initiative DDI), Pretest-Lab for harmonising key survey data (GESIS), existing R libraries to establish interfaces between

text-data from KonsortSWD-related disciplines and other contexts (University Duisburg-Essen), RDM tools for qualitative research data (Qualiservice, Bremen University).

Partners will contribute the following required resources (by Task Area):

- **TA 1 – Community Participation.** Member institutions of KonsortSWD will continue to contribute to the German Data Forum (RatSWD) as data providers by dispatching personnel. Elected members of the German Data Forum (RatSWD) invest their time voluntarily to contribute to TA.1-M.1 and as an Advisory Board to KonsortSWD.
- **TA 2 – Data Access.** Accredited RDCs will continue their cooperation and share their specific expertise, particularly with regard to sensitive data (cf. Annex 3 – Letters of Commitment). Partners will continue to operate existing repositories and search/indexing services for the RDC network. Secure data centres and remote access facilities will continue to be supported without NFDI resources.
- **TA 3 – Data Production:** KonsortSWD will profit from institutional contact to staff working on survey data production, particularly at GESIS, LIfBI, and SOEP. Existing R Code will be contributed as well as data curation resources. Data producers will provide personnel and expertise to standardise data and metadata based on their survey data and other primary data.
- **TA 4 – Research Ethics:** The WZB brings to the table far-reaching networks and experience in the field of research ethics. The efforts of all participants towards creating new infrastructures are made on a voluntary basis.
- **TA 5 – Technical Solutions:** Software architecture for metadata and search infrastructures will serve as a basis to develop more standardised services. Partners will contribute IT infrastructure and personnel for service maintenance.
- **Across all TAs,** partners will contribute rooms and IT infrastructure for personnel working within KonsortSWD.

3 Research Data Management Strategy

Existing information infrastructures, data repositories, or reusable software

Despite the long tradition of data sharing in most disciplines of KonsortSWD (the GESIS Data Archive was founded in 1960), the issue of how to access sensitive microdata from scientific and public sources remained unresolved for a long time. Existing data archives provided access to large amounts of data, especially from survey research (including cross-national survey programmes) but were restricted to anonymised data. Nevertheless, the existence of data archives triggered a culture of data exchange, even though scientists proved to be reluctant to share data in the beginning (Sieber 1991).

RDM in RDCs: KonsortSWD's specific expertise

It is largely due to the German Data Forum (RatSWD) that scientifically valuable data from large-scale data collections from German public authorities – primarily data from the statistical offices (e.g., microcensus data), the Bundesbank, and social insurance providers (especially the German Pension Insurance and the Federal Labour Office) – was made accessible with the creation of research data centres (RDCs) (for details see TA.1-M.1 RatSWD). Since 2004, scientific research institutes also began creating RDCs to improve access to sensitive data, originally from large-scale survey programmes and businesses (cf. RfII 2016, pp. 30-33, Wissenschaftsrat 2011, pp. 81-82).

Sensitive data – like small-scale regional information on survey respondents – became accessible in an organised and transparent fashion after **appropriate access paths were developed**. The solution implemented by almost all RDCs was the creation of guest researcher workstations (*Gastwissenschaftsarbeitsplätze*), which facilitate secure access to sensitive data and check any research output for disclosures. For a map of RDCs including information on guest researcher workstations, see Figure 6. Moreover, several ways of remote data analysis were developed, e.g., JoSuA at the Institute for Labour Economics. In the simplest case, users provide the analytical code, the RDC staff runs it, and then controls the output for sensitive patterns. Latest developments include remote desktop solutions (e.g., LIfBi, DZHW, SOEPremote tool (Goebel 2015)), which allow researchers to analyse data using standard statistical packages without downloading the data. This access path, however, is not legally viable for RDCs that manage data from official statistics (RatSWD 2019a).

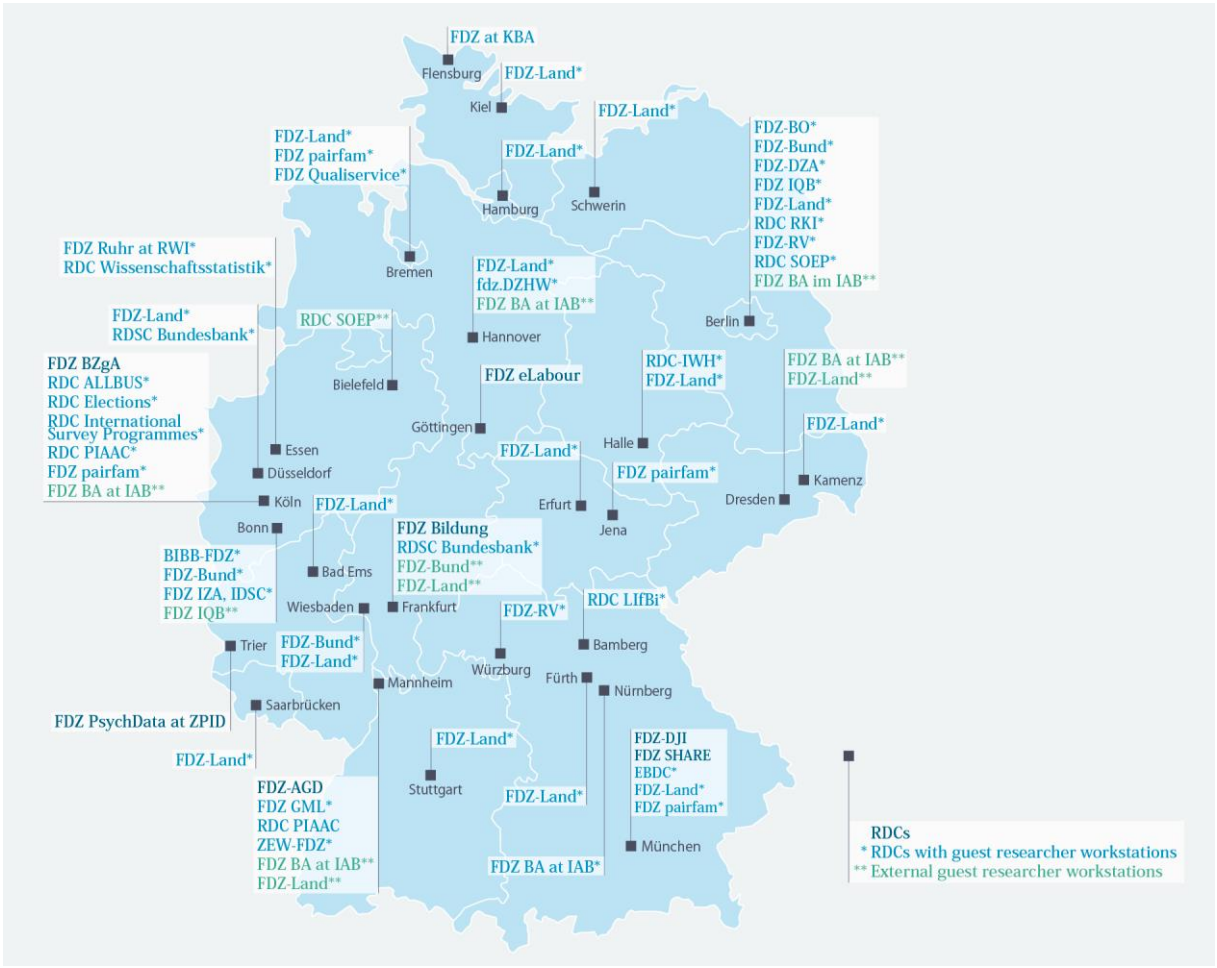

Figure 6: Map of RDCs and locations of guest researcher workstations

To facilitate the use of microdata, RDCs create factually anonymised data (Scientific Use Files, SUF) that can be provided to off-site researchers based on a usage contract. To create SUFs, **RDCs have accumulated extensive expertise in anonymisation of microdata and high-quality data documentation.** Further specific expertise exists to link sensitive data from different sources based on individual consent (e.g., data from SOEP and IAB). Linking data from different databases (e.g., surveys and process-based social insurance data) will be of increasing importance in the future. It is a promising way to enhance the use of existing data, to reduce costs of data collection, and to increase the potential for causal analysis.

KonsortSWD comprises accredited RDCs. To be accredited as RDC by the German Data Forum (RatSWD), applicants must fulfil a broad catalogue of requirements (cf. RatSWD 2018b). **Compliance with the accreditation criteria is monitored through a yearly reporting system.** A monitoring commission elected by the RDCs is operationally in charge, underlining the bottom-up nature of the quality assurance model. As an additional quality measure, the

German Data Forum (RatSWD) has installed, and will bring to NFDI, a complaints office for users to report alleged infringements of accreditation criteria.

In recent years, the RDM of RDCs has increasingly used established metadata standards for documentation (see section 3.1) and DOIs for data citation, linking data to publications, and (cross-)referencing (ZBW, GESIS & RatSWD 2014). The KonsortSWD members GESIS and ZBW established the data registration agency da|ra in 2010 (see section 3.2).

Some partners in KonsortSWD offer RDM services for external data producers. The GESIS data archive disseminates and preserves data from all social sciences and selected public authorities in the long-term. While the focus used to be on survey data, the collection policy was recently broadened to include new data types (e.g., social media data). Moreover, the GESIS data archive offers a portfolio of customised RDM services for individual data producers. Second, Leibniz-Psychology offers archiving services for experimental and survey data from the field of psychology. Third, the RDC Education at the DIPF archives mostly qualitative data from educational research. Fourth, RDC Qualiservice is a generic service provider for data from qualitative research methods in the social sciences, ethnology, and anthropology. KonsortSWD thus has a comprehensive archiving and dissemination infrastructure at its disposal. The four RDCs/archives also have facilities to provide access to sensitive data from external research projects. These generic services to data producers are fully integrated into the KonsortSWD ecosystem. They will continue to provide archiving infrastructures, RDM solutions, and RDM training to researchers at universities and research institutes in Germany and abroad.

There is also a **long history of cooperation in developing research infrastructures among the KonsortSWD members**. Two recent examples are:

- GESIS, ZBW, WZB, and SOEP have developed SowiDataNet[13], a platform supporting data curation and data sharing for research institutes. SowiDataNet combines several core RDM elements (e.g., data curation checklists, a repository run by GESIS, the possibility to run a specific search portal, and DOI registration with da|ra).
- Partners in KonsortSWD have also set up the German Network for Educational Research Data (in German: Verbund Forschungsdaten Bildung – VerbundFDB[14], Meyermann et al. 2017). RDC Bildung at DIPF, RDC IQB, and GESIS - together with other partners - created a federation for archiving and sharing data in empirical educational research. The

[13] http://sowidatanet.de/xmlui/

[14] https://www.forschungsdaten-bildung.de/ (19 September 2019)

cooperation allows for curating data from mixed-methods studies within one federated organisational frame, increasing the transparency and simplicity of data sharing.

Envisaged state of research data management

The vision of KonsortSWD is to support long-term research data management (RDM) for the social, behavioural, educational, and economic sciences in all phases of the research data lifecycle. The core of KonsortSWD's RDM strategy is to provide researchers and RDCs with the tools and services they need for managing and sharing sensitive and non-sensitive data in compliance with the FAIR principles.

Because sensitive data need to be managed in close contact with their producers and curated where expertise for research methods and digital objects is the strongest, KonsortSWD does not intend to create a central data repository for the consortium. This would be neither effective (i.e., provide the best service to research) nor efficient (i.e., create maximum benefit for given investment). Improving and facilitating access to data will remain a core mission of KonsortSWD. To this end, the German Data Forum (RatSWD) will continue to accredit RDCs so that KonsortSWD remains open to new participants, thus further invigorating its federated data infrastructure. KonsortSWD will support the creation of RDCs, especially for new data types (e.g., unstructured data from the internet and experimental data).

The strategy has three pillars. KonsortSWD will

1) **strengthen the FAIRness**[15] **of (meta)data.** RDM must cover all phases of the research process and should not restrict itself to data access and data dissemination. KonsortSWD will therefore consider all RDM aspects by (a) not only covering data archiving and dissemination but extending its RDM strategy to processes of data generation and production (cf. TA 3 Data Production) and (b) by making data and metadata FAIR especially by increasing interoperability across surveys (cf. particularly TA 3 Data Production and TA 5 Technical Solutions).

User benefit: By strengthening the quality of the metadata, KonsortSWD generates FAIR data and new possibilities for data linking. This enhances the analytical potential of the data.

2) **widen the scope of data available through RDCs** for the social, behavioural, educational, and economic sciences. This is accomplished by (a) including (sensitive) data relevant to further communities (e.g., communities using qualitative, i.e., less and unstructured data (Hollstein 2019, 2011), large scale text corpora, data collected using web scraping technologies, public

[15] A very broad overview to the FAIR principles can be found here: European Commission 2018.

health and medical research data, and data from economic and social experiments cf. TA.2-M.3 Federating RDCs, TA.3-M.2 Qualitative RDM and TA.3-M.3 Textual Data), and (b) offering solutions for common challenges, such as the growing importance of research ethics for data collection and data management. In the context of interpretative research, in particular, the re-use of (qualitative) data usually requires broad knowledge about the context of data collection for appropriate interpretation as well as adequate anonymisation (Corti et al. 2005; Medjedović and Witzel 2010). Therefore, re-use of existing data is still the exception in this research tradition, but there are examples underscoring the value of sharing qualitative data (e.g., Behrmann and Hollstein 2012; von Hodenberg 2018).

Text corpus research uses unstructured text-data created for other contexts as a data source, including minutes of parliamentary debates (Blätte and Blessing 2018), party manifestos (Döring & Regel 2019), data scraped from websites (e.g., for Wikipedia see Flöck et al. 2017, for Immoscout see Boelmann and Schaffner 2019), or social media platforms (e.g., Twitter, cf. Quinlan et al. 2018; Sloan et al. 2019). The use of these data is relatively new in the disciplines represented by KonsortSWD, but expertise is highly developed within Text+. Integrating the respective communities into our disciplines will therefore create a productive basis for a cross-consortium infrastructure. While many data sources are in the public domain, others remain restricted by property rights (e.g., newspaper archives) or proprietary terms of services (e.g., Twitter).

User benefit: By widening the scope of the RDCs, new communities will benefit from data access, from RDM expertise available in the KonsortSWD network and from the potential to link data across domains.

3) **deepen the cooperation between RDCs to overcome the current fragmentation of the data landscape** to provide users easier and more transparent access to the services of KonsortSWD. Since 2017, representatives of the RDCs identified strategic fields of action based on their daily experience in user support. Harmonising standards and processes in RDCs (TA.2-M.1 RDC Support) and a network of ‘single points of access’ (TA.2-M.2 Access Points) will facilitate data access and increase transparency. A collaborative training programme for RDC staff and junior scientists (TA.2-M.4 RDM Skills) will deepen RDM skills. To make the exchange between data providers and data users transparent and sustainably replicable, a joint online discussion platform will be implemented (TA.2-M.5 Transparent Interaction).

User benefit: Transparency and simplicity in data access will be increased by deepening cooperation between RDCs.

The RDCs (plus the GESIS data archive) curate more than 8,000 data files from scientific and public data collections. Making these data FAIR will require significant coordinated technological investments and organisational effort. Not all RDCs will be able to (immediately) implement all tools and services for FAIR data developed by the consortium. However, the consortium will develop an overall reference model for a connected, FAIR data infrastructure. It can successively integrate all RDCs and remains open to all researchers and research institutions from academia, public administration, and private business with an interest in FAIR data sharing.

TA 1 Community Participation focuses on the **monitoring of user needs**. All TAs will regularly report to TA 1 Community Participation, in particular, to the Advisory Board (TA.1-M.1 RatSWD). The service portfolio of KonsortSWD will be continually presented at the meetings of the academic associations and specialised information services (SISs). Feedback will be structured and disseminated internally by specialised officers (TA.1-M.2 Community AA and TA.1-M.3 Community SIS). Each TA is responsible for continuous user feedback in developing its services; all web services are developed following the principles of a user-centred design (TA.5-M.4).

Data selection and quality management

KonsortSWD will not have one joint data selection strategy, but will rely on the specific data collection policies of the accredited RDCs and appraisal systems in place. Quality assurance processes are part of the accreditation and monitoring of RDCs (cf. TA 1 Community Participation and TA 2 Data Access). KonsortSWD supports RDCs in developing their selection criteria for data by developing collection policies (TA.3-M.2 Qualitative RDM). RDCs offer a structure to communities that aim at selecting and providing data by creating new RDCs for their purposes. The foundation of new RDCs, and thus the expansion of the KonsortSWD infrastructure, will be based on community needs (cf. Annex 3 – Letters of Commitment by RDCs in the accreditation process).

3.1 Metadata standards

Most of the RDCs in KonsortSWD already use standardised metadata based on the well-established Data Documentation Initiative (DDI).[16] DDI is an international metadata standard for describing data produced by surveys and other observational methods in the social, educational, behavioural, economic, and health sciences. DDI is a free standard that can document and manage all stages of the research data lifecycle, including conceptualisation, collection,

[16] https://www.ddialliance.org (10.09.2019)

processing, distribution, discovery, and archiving. GESIS has been closely involved in the development of DDI since its inception. SOEP, IZA, and IAB are also full members of the DDI alliance. As an internationally established metadata standard, DDI has the potential for laying the foundation of a NFDI-wide standard. While DDI aims to describe data as completely as possible, the Dublin core meta-dataset guarantees basic interoperability - potentially across all disciplines and resource-types (data, publications, web resources, etc.).

A central service for metadata and persistent identifiers (PIDs) in Germany is the registration platform da|ra. Established by the KonsortSWD members GESIS and ZBW in 2010, its services ensure 1) identifying digital resources in a persistent way via Digital Object Identifiers (DOI) and 2) using a metadata schema to describe all registered objects (studies or datasets). The service is widely used by RDCs and research institutions in Germany. To date, da|ra has 108 national and international publication agents. Since 2010, more than 720,000 DOIs were registered since 2010. da|ra is built on an extension of the DataCite Metadata Schema that provides mapping from DataCite metadata to DDI format. Mappings also exist for the most widespread metadata schemas (e.g., Dublin Core, Schema.org) facilitating the harvesting of da|ra's metadata by other search engines (e.g., Google Dataset Search).

The focus of KonsortSWD's metadata strategy

KonsortSWD will invest in making data and metadata FAIR. Currently, the key issues to create a FAIR data infrastructure are to improve interoperability and to provide machine actionable metadata for automated information extraction and data analysis. Metadata quality and the use of PIDs will be enhanced. KonsortSWD will improve the visibility and thus findability of the data of the RDCs and link information across different repositories. KonsortSWD will invest in the advanced harmonisation of the various metadata standards and practices in RDCs (TA.3-M.1 Harmonised Variables) and align them with specific requirements of scientific communities dealing with their subject specific vocabulary (TA.5-M.2 Visibility). KonsortSWD will foster the development and implementation of referencing and identification mechanisms on the level of attributes and questions (TA.5-M.1 PID Services). Automated re-use of the metadata will be facilitated by interfaces to be developed and released (TA.5-M.3 Interfaces). Software solutions for the management of metadata are open. KonsortSWD will distribute them in cooperation with the relevant national and international players (e.g., EOSC, CESSDA ERIC).

It is important to strengthen process-oriented metadata management along the data lifecycle and the re-usability of metadata to reduce the burden of documenting data. Researchers must be encouraged in their roles as data users *and* data producers by supporting the documentation of data as early as possible during the preparation of the data collection using customised

metadata-tools (TA.5-M.3 Interfaces). KonsortSWD will also support users of statistical software packages with integrated (standardised) multilingual metadata (TA.3-M.5 Open Dissemination Format).

3.2 Implementation of the FAIR principles and data quality assurance

For sensitive data, FAIRness improved significantly during the last two decades (e.g., Eder and Jedinger 2018, European Commission 2018, p. 11). 1) Almost all RDCs now use PIDs for their data. The metadata used for data registration with da|ra are hosted by several search indices (e.g., DataCite, GoogleDatasetSearch, GESIS Data Search) making RDC data findable. 2) For sensitive data, appropriate access procedures were developed and systematically implemented. Since the creation of the German Data Forum (RatSWD), many datasets have become accessible to researchers (cf. TA.1-M.1 RatSWD). 3) A first step to strengthen interoperability was to develop and use international standard ontologies for core concepts in public statistics and social research. Their aim was to increase comparability of the measurements in cross-national survey research. Examples are the International Standard Classification of Education (ISCED) and the International Standard Classification of Occupation (ISCO). 4) Re-usability of data was improved by providing extensive data documentation to users, including methods reports, codebooks, questionnaires, and data quality reports. This also includes portals to browse the content of (survey) data at the variable level (e.g., paneldata.org).

Three major challenges remain: first, though the data from the RDCs are findable through specialised search indices, they still lack general visibility because all-purpose search engines (used by the majority of users) do not adequately retrieve information about research data. KonsortSWD plans to improve visibility for true findability (cf. TA.5-M.2 Visibility). Second, data and metadata are often not machine-actionable. Metadata are frequently contained in PDF documents and not easily re-usable or retrievable for automated processing, especially information on the attributes in data files are frequently not accessible in standardised formats. Third, interoperability of the data is limited even across studies using the same data type (e.g., data from different surveys such as SHARE and SOEP), because surveys often do not use the same operationalisation for similar constructs and data documentation normally does not reference the concept level.

The decision about specific technological solutions to address these challenges remains to be made. Given the many existing repositories of the RDCs and the different standards for data documentation, the focus will be on creating interfaces for data and metadata exchange. To define the necessary components to create a FAIR data infrastructure and appropriate

technological solutions, KonsortSWD uses the conceptual frame of the Digital Object Interface Protocol (DOIP). Figure 7 shows the core elements of DOIP. A data infrastructure is FAIR if the four layers enable humans and machines to use the data appropriately.

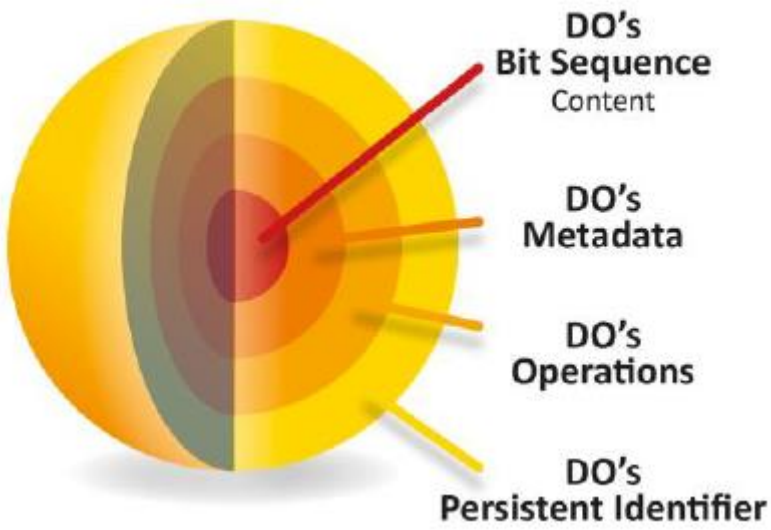

Figure 7: Core elements of Digital Object Interface Protocol (DOIP) (Wittenburg et al. 2019)

The core of the DOIP is the data is the bit sequence containing the information (Digital Object). Most of the data are stored and disseminated in proprietary formats because the majority of the users in the social, behavioural, educational, and economic sciences works with proprietary research software. Regarding the core of the DOIP, KonsortSWD strives at improving FAIRness of data by developing an open data format facilitating the use of non-proprietary software (e.g., R) (TA.3-M.5 Open Dissemination Format).

The second layer of the DOIP is the metadata describing the DO. The FAIR principles are about high-quality data description. Re-use and interoperability require enhanced and more granular metadata available in established standards. To overcome harmonisation problems when pooling data from different sources, we aim at developing more harmonised measurements (TA.3-M.1 Harmonised Variables).

The third layer concerns the operations that can be performed with the DO. With sensitive data, this concerns mostly access restrictions due to privacy and/or confidentiality. Harmonised access conditions (TA.2-M.1 RDC Support) will increase transparency in data access.

The outer layer of the DOIP is the PID for each DO. If the DOs are to be defined on a more granular level of variables/attributes, the existing PID infrastructure must be adapted to this new use case (TA.5-M.1 PID Services). This also includes the use of PIDs for cross referencing metadata, for example, if standardised ontologies are used for the description of the data.

The different measures to design a FAIR data infrastructure need to be integrated into an overall concept of FAIR data in the KonsortSWD disciplines. To avoid creating isolated solutions for KonsortSWD, EcoSoc-IN will serve as an international forum for discussing FAIR solutions with representatives from national and international RDIs (see section 2.4). While EcoSoc-IN is concerned with subject-specific definitions and implementations of FAIR-compliant data infrastructures, members of KonsortSWD (and other players) created GO Inter: Cross-Domain Interoperability of Heterogeneous Research Data implementation network (GO Inter IN). This implementation network focuses on technical challenges of fostering the interoperability of data available in distributed repositories.

Data quality assurance: building (technical) competence for digital long-term preservation

RDCs usually focus on data types that are most relevant for their specialised user communities. RDC staff is responsible for in-depth documentation, anonymisation of personal data, dissemination, and user support. The majority of RDC staff conducts their own research using the data they curate. Therefore, the researchers curating the data have an in-depth understanding of theories and methods underlying the data generating process. This model of combining research and data stewardship is one of the core features of the RDCs and the basis for high quality data provision and user support.

Professional data curation requires specific competencies from library and information science that are not covered by subject-specific university curricula. First, RDC staff needs more training on using and developing metadata standards for improved quality of data documentation. Second, RDC staff is increasingly involved in developing specific data curation and analysis software and managing distributed file systems. This requires competencies in programming and database management. Therefore, KonsortSWD will support advanced RDM training for researchers working at the RDCs and similar institutions (TA.2-M.4 RDM Skills).

Another core component of research data infrastructures is their capacity for digital long-term preservation beyond ten-year bitstream preservation. This requires additional competencies for digital long-term preservation. The best way to prove the trustworthiness of the RDCs is to

acquire a Core Trust Seal[17] or nestor seal[18] certification. Certification ensures that RDCs implement the most significant policies for RDM (e.g., preservation policies, access policies). KonsortSWD will support RDCs going through the certification process thus enhancing the quality of the technical and organisational preservation procedures and increasing trust in the RDCs (TA.2-M.1 RDC Support). Alternatively, RDCs can use certified repositories within the consortium for long-term preservation of data (such as the Archiv für Gesprochenes Deutsch AGD or the GESIS data archive).

3.3 Services provided by the consortium

The new services provided by KonsortSWD will **complement and complete** existing domain-specific repositories, data management services (e.g., data registration), and access infrastructures. The objective behind the provision of new services is to lift the existing infrastructure to a new level of quality by integrating user needs based on an iterative dialogue process. The new services contribute to the three pillars of the KonsortSWD research data management (RDM) strategy: strengthening, widening, deepening. Not all the Measures develop new services, but they may contribute to the KonsortSWD strategy otherwise.

TA 1 Community Participation provides the following services:

1. The German Data Forum (RatSWD) advocates improving data access, cooperates with RDCs, fosters community integration, and serves as an advisory board for KonsortSWD (widening and deepening).
2. A designated community officer links academic associations with the activities of KonsortSWD and translates the relevant information in both directions (widening).
3. A designated officer links specialised information services (SISs, *Fachinformationsdienste*) and libraries with the activities of KonsortSWD and translates the relevant information in both directions (widening).
4. The Outreach Measure provides public relations-related services (widening and deepening).
5. The Mid-term Consortial Conference provides information to other consortia and the wider public and a forum for networking to learn about the needs in the wider communities (widening and deepening).

TA 2 Data Access provides the following services:

[17] https://www.coretrustseal.org/about/history/data-seal-of-approval/ (31 August 2019); The role of Core Trust Seal is underlined by European Commission (2018, p. 44).
[18] https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html (31 August 2019)

6. A support unit for FAIR data management of RDCs (TA.2-M.1 RDC Support) (deepening)
7. A remote access network of interconnected secure guest researcher workstations of participating RDCs (TA.2-M.2 Access Points) (deepening)
8. A federated archiving infrastructure for data from qualitative social research (TA.2-M.3 Federating RDCs) (widening)
9. A training programme to improve data management skills in RDCs and communities (TA.2-M.4 RDM Skills) (strengthening)
10. A moderated online forum for transparent interaction between data providers and data users (TA.2-M.5 Transparent Interaction) (deepening)

TA 3 Data Production provides the following services:

11. Standards for core attributes in surveys to improve interoperability and re-usability of existing and newly generated data (TA.3-M.1 Harmonised Variables) (strengthening)
12. An RDM portfolio for qualitative social research (TA.3-M.2 Qualitative RDM) (widening)
13. An open-source toolset for storing large-scale textual data and linking it to other data sources (TA.3-M.3 Textual Data) (widening)
14. A database-supported tool for the automated coding of textual information (e.g., open responses in surveys) and the derivation of standard variables in the fields of education and occupation with corresponding services. (TA.3-M.4 Open Response Coding) (strengthening)
15. A non-proprietary data format for data dissemination and long-term preservation (TA.3-M.5 Open Dissemination Format) (deepening)

TA 4 Research Ethics provides the following service:

16. It initiates and coordinates an infrastructure for research (data) ethics for the KonsortSWD-related disciplines, including the provision of support material and the organisation of a forum for networking and exchange also with other NFDI consortia on issues of research ethics and data protection (see Berlin Declaration) (TA.4-M.1 Research Ethics)

TA.5 Technical Solutions provides the following services:

17. Extended PID services for variables in data files linked to a metadata schema for cross-referencing (meta)data and improved data discovery (TA.5-M.1 PID Services) (strengthening and deepening)
18. Ensure the visibility of research data (e.g., in standard web searches, TA.5-M.2 Visibility) (widening)

19. Build an open application programming interface specification based on DDI for the whole data exchange process, considering W3C standards and common Web vocabularies as well as methods promoted in the framework of international FAIR initiatives (TA.5-M.3 Interfaces) (deepening)
20. Optimise KonsortSWD online services for user experience and usability (TA.5-M.4 UCD) (deepening)

In order to **foster acceptance** for KonsortSWD among the scientific communities, its services must improve cost-effective access to (sensitive) research data with no or low usage fees. The accreditation of RDCs will remain a core element of KonsortSWD thus increasing trust, acceptance and usage of data and services. The RDC network disseminates new services to their user communities. To **reach out** to other scientific communities represented within the NFDI KonsortSWD will continually improve its services. RDM standards must be strengthened and integrated into the KonsortSWD's Measures. In addition, KonsortSWD will adopt technologies developed and implemented by other consortia and integrate them into the NFDI or the EOSC.

The proposed services will be adopted by the RDCs within KonsortSWD with a particular focus on **sustainability**. RDCs are typically part of established institutions that provide the infrastructure and compliance regulations necessary for data security. In their Letters of Commitment, the participating institutions declared their support for the development and implementation of new services. For this reason, all Task Areas (TAs) regularly report on service development to the Committee of Data Access, where the RDCs give feedback on the services. The RDC network can secure long-term preservation and access to data even when an RDC is closed.[19] Services that are not used by the RDCs, data collection programmes, or other users will be discontinued.

The proposed platforms for community participation (TA 1 Community Participation) ensure that service development is flexible and **responsive to the dynamic needs** of the research communities. With TA 4 Research Ethics, KonsortSWD responds to a request from the community for coordinated action to raise awareness and to create standards for dealing with (changing) ethical and legal frameworks. Overall, participatory processes are incorporated into each Measure and embedded within the organisational structure of the consortium, especially through TA 1 Community Participation.

[19] In such cases, the data can be transferred to partnering repositories. This has happened in the past with the data of the RDC SFB882.

Resources for Network Refinement (TA.6-M.3) help **adapt the work plan** to demands from the user communities that had not previously been anticipated. The German Data Forum (RatSWD) will propose additional focus areas in the course the funding period. The governance and development process of the TAs allow for the continuous review of developed services and thereby for close monitoring of the acceptance of the measures proposed by KonsortSWD within the scientific communities.

Participatory processes are incorporated at the TA level. The TAs present the developed services to the relevant scientific communities and to the whole NFDI. To ensure the longevity of developed services, all code of new services will be provided in open source format (for details, see TA 5 Technical Solutions). The Milestones and deliverables of the Measures are all formulated to make services long-lasting. The cooperating partners agree on common standards and tools to ensure the re-usability of all services. The partners collaborating within KonsortSWD will incorporate the developed services and tools into their broader portfolios. KonsortSWD promotes the establishment of joint communities of developers. Training material, technical documentation, user support, and best-practice data provision are core deliverables of all Measures.

4 Work Programme

4.1 Overview of Task Areas

KonsortSWD proposes six Task Areas (TAs): TA 1 manages systematic user-involvement and information. TA 2 to TA 5 focus on sustainable research data management (RDM) and ethical aspects along the data lifecycle to deepen, strengthen, and widen a sustainable research data infrastructure for the study of human society. These TAs are of particular relevance to other NFDI Consortia. TA 6 bundles administrative services.

The Measures provide researchers and research data centres (RDCs) with tools and services they need for the management and sharing of sensitive and non-sensitive (new) data types in compliance with the FAIR principles while considering ethical and legal aspects to secure data access.

Task Area	Measures	Responsible co-spokesperson(s)
TA 1: Community Participation	• TA.1-M.1 German Data Forum (RatSWD) • TA.1-M.2 Community involvement: academic associations • TA.1-M.3 Community involvement: specialised information services and libraries • TA.1-M.4 Dissemination and outreach • TA.1-M.5 Mid-term Consortial Conference	Heike Solga (WZB)
TA 2: Data Access	• TA.2-M.1 Supporting research data centres by establishing FAIR principles and certification for sustainable and trustworthy data access • TA.2-M.2 Creating single points of access for sensitive data in the research data centres • TA.2-M.3 Federating the archiving infrastructure of the RDCs: VQualidat • TA.2-M.4 Developing and exchanging research data management skills • TA.2-M.5 Improved use of research data through transparent interaction • TA.2-M.6 Awarding RDM grants to foster cooperation between RDCs and research communities	Christian Aßmann (LIfBi)
TA 3: Data Production	• TA.3-M.1 Facilitating the combination of research data through standardised and harmonised variables • TA.3-M.2 Generation of qualitative data – RDM portfolio for qualitative social research • TA.3-M.3 Piloting the processing and linking of textual data • TA.3-M.4 Improved processes for coding textual information from open response formats and for deriving standard variables • TA.3-M.5 Open, metadata enriched, non-proprietary data format for data dissemination	Jürgen Schupp (SOEP at DIW Berlin)
TA 4: Research Ethics	• TA.4-M.1 Establishing a research ethics infrastructure	Maja Adena (WZB)
TA 5: Technical Solutions	• TA.5-M.1 Enhancing PID services as a base for a FAIR data infrastructure • TA.5-M.2 Enhancing data findability • TA.5-M.3 Interfaces for data exchange processes • TA.5-M.4 User-centred design	Klaus Tochtermann (ZBW)
TA 6: Secretariat	• TA.6-M.1 Accounting, monitoring, and reporting, Annual Consortial Assembly • TA.6-M.2 Contract management • TA.6-M.3 Network Refinement • TA.6-M.4 Consortial governance (Consortial Assembly, Steering Committee)	Christof Wolf (GESIS)

4.2 Task Area 1: Community Participation

Community involvement and user integration are the basis for KonsortSWD's activities. Each of KonsortSWD's Task Areas (TAs) explicitly connects to and involves the scientific community in the development and implementation of its services and activities. KonsortSWD considers the structural, that is, organisational integration of community participation to be inseparable from its mission to improve research data management (RDM). TA 1 comprises a set of Measures that add to and reach beyond the activities in other TAs in KonsortSWD.

The community served by KonsortSWD entails the following partners and stakeholders which are connected to KonsortSWD:

- 3,000 individual researchers, who elected their representatives to the German Data Forum (RatSWD) in 2017,
- 15 academic associations (AS, DGfE, DGGÖ, DGMP, DGfP, DGPs, DGPuK, DGSKA, DGS, DGV, DStatG, DVPW, GEBF, VHB, VfS), which propose candidates for this election,
- user councils and scientific advisory boards of the institutions behind KonsortSWD and of research data centres (RDCs),
- individual and institutional data producers seeking support,
- individual scientists or institutions who are interested in using data in Germany and abroad,
- institutions interested in integrating KonsortSWD's services into their portfolios (e.g., universities, specialised information services (SISs, *Fachinformationsdienste*), libraries, RDCs, repositories, and surveys),
- local research ethics committees,
- other NFDI consortia.

KonsortSWD's services are intended to benefit researchers and students particularly from the social, behavioural, educational, and economic sciences. Already, data and services provided by the existing infrastructure are requested and used by scientists on a global scale. RDCs are an international role model for access to sensitive data (e.g., Wissenschaftsrat 2011, p. 81).

Central challenge: Within the disciplines covered by KonsortSWD, the consortium represents those communities which are characterised by mature and rich cultures of data sharing. However, there are other communities within the same disciplines that follow different research data paradigms. KonsortSWD intends to offer its services to all communities. It aspires to address the needs, expectations, and requirements of researchers in the relevant disciplines.

Objectives: To realise its mission (cf. section 2.1), KonsortSWD identifies researchers' needs which follow from their specific scientific interests and develop along the data life cycle. It integrates them in all its Task Areas (TA) to generate user-focused solutions. TA 1 combines Measures that reach beyond the user involvement that is specific for the development of services for data access (TA 2), data production (TA 3), research ethics (TA 4), or technical solutions (TA 5).

TA 1 combines five Measures to support user involvement: the German Data Forum (RatSWD, TA.1-M.1) systematically involves research communities. It operates as a lobby group advocating data access. It will also take on a new role as the Advisory Board to KonsortSWD.

Academic associations are a vital link to the users in the scientific communities. TA.1-M.2 (Community AA) foresees a community officer to focus on the dialogue with academic associations. The community officer has three tasks: 1) to aggregate and transfer community needs into KonsortSWD, 2) to promote and disseminate information about the activities of KonsortSWD and its services to academic associations, 3) to contribute to the cultural change in terms of data sharing in all disciplines. TA.1-M.3 (Community SIS) is similar in nature but connects to SIS and academic libraries. This helps KonsortSWD to link its development to user needs, contributes to integrate SIS services, and connects them to the KonsortSWD research data infrastructure. TA.1-M.4 (Dissemination) offers a set of general dissemination and outreach services to KonsortSWD and its stakeholders. Finally, TA.1-M.5 (Conference) prepares a networking conference on research data infrastructures with a multi-stakeholder approach that aims at reaching out beyond academia. Overall, the Measures of TA 1 connect KonsortSWD with its user base in a structured, systematic, and sustainable way and provide intermediation services for KonsortSWD and its scientific communities.

TA 1 contributes to the following KonsortSWD objectives (see section 2.1): 1 - by directly addressing researchers and their representatives (TA.1-M.1 RatSWD and TA.1-M.2 Community AA), 3 - by interacting with research communities (TA.1-M.1 RatSWD, TA.1-M.2 Community AA, TA.1-M.5 Conference), 6 - by continuing to advocate for data sharing in research communities and beyond (TA.1-M.1 RatSWD and TA.1-M.2 Community AA), 7 - by connecting with international partners (TA.1-M.1 RatSWD), 8 - by opening the option to be represented and by seeking contact to new fields (TA.1-M.1 RatSWD, TA.1-M.2 Community AA, and TA.1-M.3 Community SIS). Also, TA 1 offers a point of contact for stakeholders from outside KonsortSWD.

Membership and governance: TA 1 monitors its Measures continuously on an operational level and ensures the necessary reporting to the Steering Committee (StC). There will be two annual meetings of the individuals in charge of the separate Measures to foster exchange. The co-spokesperson (or a deputy) in charge of TA 1 moderates the dialogue with other TAs and represents the interests of the research communities behind KonsortSWD in the StC. The head of TA1's staff has guest status in the StC.

Outreach: Past experience regarding the direct exchange with research communities will be helpful in interacting with other consortia. When KonsortSWD-users share needs regarding, for example, medical, geo-coded, text-based or other unstructured data with users in other consortia TA 1 will communicate NFDI-wide developments or services to KonsortSWD and its user-communities. In addition, TA 1 connects to communities covered by other NFDI consortia and disseminates information about KonsortSWD's activities there. Given the increasing

relevance of internationally comparative research in the disciplines covered by KonsortSWD, the international connection of the institutions behind KonsortSWD is an important strength. TA 1 will regularly report to the Committee for Data Access (CDA, TA 2 Data Access) to make sure that services correspond to user needs in the RDCs.

Distribution of responsibility and contribution of applicants: The Measures of TA 1 will be tightly interconnected to create synergies and to establish service-oriented communication within KonsortSWD and with outside stakeholders. Feedback channels between KonsortSWD, national communities, and international stakeholders will be established and systematically monitored in close cooperation with the StC and the consortium's spokesperson. TA 1 and its Measures are jointly managed by the Berlin Social Science Centre (WZB). WZB also contributes by providing office space. Heike Solga, as co-spokesperson of TA 1, will be in charge of the TA.

Resources: 1 postdoctoral researcher, 0.5 doctoral researcher, and 0.25 team assistant to coordinate different channels of community involvement within TA1. They systematically track and communicate community feedback and responses from within KonsortSWD. Additional resources are required for national and international travel, for meetings to address user feedback, and for two annual meetings with individual Measures.

Measure TA.1-M.1 German Data Forum (RatSWD)

Status quo: The German Data Forum (RatSWD) – founded in 2004 and so far financed on a project basis by BMBF. It **consists of appointed data producers** (such as official statistics agencies, social insurance providers, and research-driven data production) **and directly elected academic data users**.[20] The German Data Forum (RatSWD) has been active in five areas.[21]

First, it was the **driving force behind developing the existing research data infrastructure** in the social, behavioural, educational, and economic sciences. To improve access to sensitive data in the relevant disciplines, the German Data Forum (RatSWD) established a model of accredited, decentralised research data centres (RDCs) (RfII 2016, pp. 30-33., Wissenschaftsrat 2009, 2011, pp. 81-82). These RDCs provide flexible and structured access to diverse, research-generated as well as process-produced data (including from official statistics

[20] The representatives of scientific data users are elected in a secret ballot from a pool of candidates that is nominated by (currently) 15 academic associations, guaranteeing representation of different disciplines.
[21] The areas were further developed in response to the evaluation of the Wissenschaftsrat (2009).

agencies, social insurance providers, research projects, and surveys) for independent research. More than 46,000 users used these data in 2018 (RatSWD 2019a). The RDCs cooperate as a network in the Committee for Research Data Infrastructure to improve data management and to institutionalise data access in accordance with the FAIR principles. The German Data Forum (RatSWD) developed standards and criteria for the accreditation of RDCs. For quality assurance purposes, it established a **monitoring commission** to collect and assess **annual RDC reports,** monitor compliance with RDCs' obligations, and maintain a **complaints office for users**.

Second, the German Data Forum (RatSWD) **offers a forum for dialogue between data producers and data users**. Both groups cooperate to develop the data infrastructure and to improve legal and practical procedures for access to sensitive data. The results of the German Data Forum's (RatSWD) consultations are published in openly accessible formats.[22] They support individual researchers and standardisation processes in academia.

Third, as an independent council, the German Data Forum (RatSWD) acts as the voice for its research communities and **advises legislators and science institutions** (e.g., DFG, BMBF). In the 2014-2018 period, it issued almost 40 expert statements dealing with a wide range of topics at the European and national level, including the amendment of the federal law on statistics (BStatG) and the General Data Protection Regulation (GDPR) (WZB 2019, pp. 23-25). The German Data Forum (RatSWD) is a sought-after advisor to the federal government with an established network of contacts and a reputation as a reliable partner for public authorities and the administration (WZB 2019, pp. 23-27). Given the disciplines' focus on sensitive data and on data from administrative sources, it is important to represent these interests in legislative processes.

Fourth, the German Data Forum (RatSWD) **connects research data infrastructures at an international level**. It has been supporting the European Open Science Cloud (EOSC) as a stakeholder since 2017. In 2018, it started the Go FAIR EcoSoc Implementation Network (EcoSoc-IN) to give its disciplines a stronger voice and to apply the FAIR principles with attention to issues arising from sensitive data. In addition, the German Data Forum (RatSWD) has been cooperating with the UK Data Forum and supporting the Research Data Alliance (RDA) as an organisational member since 2015.

[22] In the period 2014-18, the German Data Forum (RatSWD) issued 107 publications, including 12 RatSWD Outputs. For more details, please see https://www.ratswd.de/publikationen/output (25 September 2019).

Fifth, the German Data Forum (RatSWD) **organises communication.** It facilitates the exchange between data users (researchers) and data producers. Since 2003, the Conference for Social and Economic Data (KSWD) has been fostering interdisciplinary exchange on data generation, data management, and data access, and focuses on the exchange with and between various stakeholders from administration, research policy, and the wider public.

Objectives: The German Data Forum (RatSWD) will offer advice to politics and administration that is often requested at very short notice and will continue its activities as an advocacy group for data access (Milestones 2, 3, 5, and 10). In the future, the NFDI directorate might assume some of the latter activities, especially those that have systemic relevance for the NFDI. As data needs in the covered research areas differ from those of many other disciplines, it is important to maintain a strong and competent voice for data access. To do so, the German Data Forum (RatSWD) will initiate up to three working groups[23] that develop position papers which can be the basis for future KonsortSWD services. It will work closely with the other TA 1 Measures to foster community integration and to safeguard a user-driven development of KonsortSWD. The German Data Forum (RatSWD) will continue to accredit RDCs and monitor their network (including running the complaints office). This is achieved through tight cooperation with TA 2 Data Access. The German Data Forum (RatSWD) will rely on its network to support the start-up of the research ethics infrastructure (see TA 4 Research Ethics, Milestone 4). In addition, the German Data Forum (RatSWD) will take on a new role in the KonsortSWD setting and act as a scientific advisory board within KonsortSWD (cf. section 2, Milestones 1 and 6).

Membership and governance: The members of the German Data Forum (RatSWD) represent either data users or data producers. The former typically are nominated by academic associations and elected by the research community. The latter are appointed by the Federal Ministry of Education and Research (BMBF) to establish a balanced representation of data providers (e.g., administrative vs. research-based data, large vs. small producers). The German Data Forum (RatSWD) plans to increase the number of representatives from 8 to 10 for both groups to improve the coverage of disciplines. It offers guest status to the chairs of the CDA and the German Data Forum (RatSWD) chairs enjoy guest status at the CDA meetings.

[23] https://www.ratswd.de/en/activities/working-groups (23 September 2019)

Timetable:

Year	Milestones
1	[M1] Take up the advisory role for KonsortSWD governance, establish procedures [M2] Determine strategic agenda for the next appointment period [M3] Initiate up to three working groups that develop position papers which can be the basis for future KonsortSWD services [M4] Support the setup of a Committee for Research Ethics by activating and involving the long-running network of experts in research ethics
2-4	[M5] Pursue strategic agenda [M6] Act as an Advisory Board and monitor Task Areas and Measures [M7] Contribute to the Mid-term Consortial Conference [M8] Support allocation processes of funds for Network Refinement (TA.6-M.3) [M9] Reach an understanding with the NFDI Directorate regarding advocacy and advisory tasks
5	[M10] Complete working papers and recommendations

Risks of implementation: In light of past experiences, risks are limited. In 2020, new members of the German Data Forum (RatSWD) will be elected who will be introduced to the procedures and activities by members in their second appointment period.

Distribution of responsibility and contribution of applicants: The members of the German Data Forum (RatSWD) work on an honorary basis. The German Data Forum (RatSWD) acts independently. Its chairperson is responsible for the Measures. As in the past, the German Data Forum (RatSWD) is administered by the Berlin Social Science Centre (WZB). The WZB contributes by providing office space.

Resources: The German Data Forum (RatSWD) needs funding under the umbrella of NFDI because the BMBF can no longer provide the German Data Forum (RatSWD) with project-based funding. 1 postdoctoral researcher, 1.5 doctoral researchers, 0.5 team assistant, and 0.26 student assistant organise the meetings of the German Data Forum (RatSWD), prepare statements, represent the German Data Forum (RatSWD) nationally and internationally, and run the cooperation with the Committee for Data Access (accreditation process of new RDCs, annual monitoring of RDCs, complaints office). In addition, funds are required for commissioned work (e.g., legal opinion papers), the election management of the German Data Forum (RatSWD), travel budgets, catering, and rent for event venues.

Measure TA.1-M.2 Community involvement: academic associations

Status quo: Academic associations represent the interests of researchers in their respective disciplines. They are natural partners for infrastructures serving the needs of researchers. Academic associations can reflect the needs of the scientific community and help disseminate

the information on service and infrastructure innovations to their members.[24] In the past, the German Data Forum (RatSWD) frequently involved academic associations in discussions on specific topics such as research ethics, research data management (RDM), and data protection standards. Academic associations differ in their involvement, the extent to which they disseminate information on the German Data Forum's (RatSWD) activities, and in their sophistication regarding RDM. The German Data Forum (RatSWD) intensified its exchange with these communities. A preparatory meeting with 15 academic associations in June 2019 yielded a broad consensus on the necessity to involve academic associations in the development and implementation of KonsortSWD. This will have to be done even more systematically in the future to support distribution of information and a cultural shift towards more data sharing and active RDM (RatSWD 2019b).

Objectives: The measure aims to foster the dialogue with and between users of KonsortSWD's services and to support the user-based development of KonsortSWD's services. The objective is to develop and implement services that meet the needs of KonsortSWD's communities regarding data infrastructures, data linkage, and legal issues.

A community office will be established as a permanent point of contact. The community officer initiates discussions and invites suggestions and feedback from academic associations. Based on this information, KonsortSWD will determine the development of its services.

In addition, the community officer promotes and discusses KonsortSWD's services at the conferences of academic associations, their board meetings (by invitation), and in symposia on methods and RDM. The community officer will organise annual meetings for representatives of all KonsortSWD-related academic associations to build bridges between methodological families and to create a common understanding of community needs and the goals of KonsortSWD. Thus, academic associations, especially those without elected members in the German Data Forum (RatSWD), can participate more actively in the development of KonsortSWD.

The Measure achieves three aims: First, it translates community needs into KonsortSWD services (Milestone 2 and 3). Second, it promotes and disseminates information about the activities of KonsortSWD and its services to academic associations (Milestones 1 and 5). Third, it contributes to cultural change in data sharing in all disciplines (Milestones 4 and 6). Strategically, the community officer will identify and support those individuals and groups within

[24] In 2011, the German Council on Science and the Humanities (Wissenschaftsrat) underscored the importance of academic associations in the development of research infrastructures (cf. Wissenschaftsrat 2011, p. 8).

the communities who further the cause of KonsortSWD and the NFDI, promote the relevant services in their respective communities, and can provide feedback to the consortium. TA.1-M.2 (Community AA) thus goes well beyond the objectives that the German Data Forum (RatSWD) was able to pursue based on its limited resources.

Timetable:

Year	Milestones
1-2	[M1] Establish contacts to academic associations and a process for the timely communication of activities and intended implementations; invite and obtain feedback from communities and academic associations [M2] Establish swift processes on an operative level within KonsortSWD to forward user feedback to the targeted TAs and their Measures and to inform users about the reactions to their feedback [M3] Initiate a structured dialogue to identify goals and pitfalls in specific TAs and Measures of KonsortSWD
3-5	[M4] Raise awareness for a culture of data sharing [M5] Communicate regularly with academic associations and other contacts [M6] Discuss review criteria in journals and to the academic reputation system

Risks of implementation: The successful involvement of the communities relies on the voluntary and complementary collaboration of the stakeholders. The discussion about the development and implementation of KonsortSWD's services might occur at different speeds depending on the research community.

Distribution of responsibility and contribution of applicants: The community officer will cooperate closely within TA 1 Community Participation and with the Steering Committee. While some of the official interaction with the chairs of academic associations remains at the discretion of the co-spokesperson, much of the information exchange and regular presentation and communication activity will be in the hands of the community officer. The Measure is administered by the Berlin Social Science Centre (WZB). WZB contributes by providing office space. The responsibility for the operation rests with the co-spokesperson in charge of TA 1 Community Participation.

Resources: 0.75 doctoral researcher (community officer), 0.25 team assistant, and 0.26 student assistant coordinate the activities under the direction of the co-spokesperson. Additionally, it will require travel funds for trips to board meetings and annual conferences. The officer should also organise at least one central meeting of all involved academic associations per year, which requires funds for venues and catering.

Measure TA.1-M.3 Community involvement: specialised information services and libraries

Status quo: Specialised information services (SISs, *Fachinformationsdienste*) are funded by the German Research Foundation (DFG) to provide 'access to research-specific information quickly and directly, irrespective of work location.'[25] The DFG currently funds several SISs that focus on the disciplines of KonsortSWD.[26] KonsortSWD has touched base with these SISs. The growing interest of SISs in research data management (RDM) suggests an integration and involvement of these stakeholders in KonsortSWD's activities. Their intermediary function between individual researchers, academic associations, and libraries is of great value to KonsortSWD. They may reach users not yet covered by RDCs or academic associations.

Objectives: The measure aims to foster the dialogue between SISs and KonsortSWD to support the user-focused development of KonsortSWD's services and to actively support the RDM developments in this newly established set of research data service providers. A library officer will be established as a permanent point of contact to SIS which may spread the services of KonsortSWD to a wider community. The library officer initiates discussions and invites suggestions and feedback from SISs. The office will widen the SIS network, as there are further SIS that have a significant share of their user community in the disciplines covered by KonsortSWD, for example, the SIS covering the world regions (Seeger 2019).

In addition, the library officer promotes and discusses KonsortSWD's services with individual SISs, at their meetings (by invitation), and in symposia on methods and RDM (Milestones 2, 3, and 5). They will organise annual meetings for representatives of KonsortSWD related SISs and academic libraries to bridge between disciplinary families and to create a common understanding of the SISs, their communities' needs, and the goals of KonsortSWD (Milestones 2 and 4). It is foreseen that they eventually participate more actively in the development of KonsortSWD.

[25] https://www.dfg.de/en/research_funding/programmes/infrastructure/lis/funding_opportunities/informationservice_science/index.html (accessed 11 Oct. 2019)

[26] SIS for Criminology, Education, Media Science, Mobility Studies, Political Science, Social and cultural anthropology, and Sociology. ZBW, co-applicant of KonsortSWD, also offers services comparable to SIS for commerce and economics.

Timetable:

Year	Milestones
1-2	[M1] Establish contacts to SISs; define a process for the timely communication of activities and intended implementations to obtain feedback from libraries [M2] Hold a first meeting with SISs, initiate a structured dialogue to identify goals and pitfalls in the operations of KonsortSWD [M3] Establish swift processes at an operative level within KonsortSWD in order to forward SISs' feedback to the targeted Task Areas and their Measures and to inform SIS about the reactions to their feedback
3-5	[M4] Raise awareness for a culture of data sharing, for example, by promoting data publication and data citation [M5] Communicate regularly with the leadership and RDM operators of SISs [M6] Encourage the proposal of suggestions for Network Refinement by SISs (see TA.6-M.3)

Risks of implementation: The successful involvement of the SISs and academic libraries relies on the voluntary collaboration of the stakeholders, all of whom expressed their support in the attached letters of support

Distribution of responsibility and contribution of applicants: The community officer will cooperate closely within TA 1 Community Participation and with the Steering Committee. While some of the official interaction with the leading personnel of SIS and libraries remains at the discretion of the co-spokesperson, much of the information exchange and regular presentation and communication activity will be in the hands of the community officer. The measure is administered by the Berlin Social Science Centre (WZB). WZB contributes by providing office space. The responsibility for the operation rests with the so-spokesperson in charge of TA 1 Community Participation.

Resources: 0.25 doctoral researchers (community officer) to organise exchange of information. Travel funds for SIS meetings and annual conferences. Funds for setting up a yearly meeting of all involved SISs and academic libraries to foster networking. The budget covers costs for catering and event venues.

Measure TA.1-M.4 Dissemination and outreach

Status quo: KonsortSWD must communicate with various stakeholders such as public administrations, data protection officers, policy makers, and the media, to support individual researchers, (organised) communities, and data producers. For KonsortSWD to reach its growing network of stakeholders, communication requires a mix of formats. Information must be accessible to support cultural change within research communities. As dissemination is of central importance to community involvement, this measure is part of TA 1 Community Participation.

Objectives: A public and community relations office (PR office) will document and promote KonsortSWD services in dynamic online formats and in print. Furthermore presentations in panels and at conferences will be supported and coordinated. A “corporate identity” will support high recognition value of KonsortSWD’s activities. Synergies can be created by coordinating publicity-relevant activities that involve different stakeholders. Past efforts (WZB 2019, pp. 28-43) that were more limited in scope allow for a realistic estimate of effort necessary.

To achieve its objectives in public relations and documentation of KonsortSWD’s activities, the Measure consists of the following tasks:

- to support the preparation of **KonsortSWD papers, statements, recommendations, reports, and press releases** in German and English, including publications throughout KonsortSWD, such as the ‘RatSWD Output Series’ and the annual ‘RDC Activities Report’ (Milestones 1, 5 and 6),
- to **channel public relations of KonsortSWD** (e.g., provide background information to stakeholders, issue press releases and place op-ed articles in newspapers and magazines, prepare and publish auxiliary communications material such as flyers, advertisements, stand-up displays, support Mid-term Consortial Conference etc.), to edit and to layout a newsletter (approximately every two months) and the continuous communication via social networks (relevant for all Milestones/funding period),
- to provide a **central information service** (giving access to KonsortSWD’s services, recommendations, presentation slides, etc.) in cooperation with all organisational elements; trace the **results and media response** of KonsortSWD (relevant for all Milestones),
- to support **KonsortSWD presentations** at conferences and meetings (relevant for all Milestones),
- to develop and maintain **KonsortSWD’s website** in German and English, including hosting content management. The website provides the natural anchor and reference point for numerous services developed in the KonsortSWD Measures (relevant for all Milestones).

Timetable:

Year	Milestones
1-2	[M1] Establish publication standards for KonsortSWD early on and keep them adaptable to new communication formats during the entire funding period [M2] Build an integrated website for all parts of KonsortSWD and link information on all participants (in close cooperation with TA.5-M.4 UCD) [M3] Prepare details of the first Mid-term Consortial Conference
3	[M4] Support the Mid-term Consortial Conference to provide a platform for the discussion of first results within the KonsortSWD community, with other consortia, and the overall NFDI (see also TA.1-M.5 Conference)

4-5	[M5] Communicate adaptations resulting from mid-term conference and other user feedback. [M6] Prepare integrated report of first funding period.

Risks of Implementation: Given that the measure builds on long-running prior experience, there are no implementation risks.

Distribution of responsibility and contribution of applicants: The measure will cooperate closely within TA 1 Community Participation, with the other TAs and with the Steering Committee. It is administered by the Berlin Social Science Centre (WZB) that contributes by providing office space. The responsibility for the operation rests with the co-spokesperson in charge of TA 1 Community Participation.

Resources: Staff of 0.5 doctoral level and 0.9 graphic design expert, 0.26 student assistant is in charge of all press releases, publications, and PR activities of KonsortSWD. Funding for contract awards (e.g., translating, editing, web-programming, and maintenance), printing expenses, production of advertisements, and stand costs is envisioned.

Measure TA.1-M.5 Mid-term Consortial Conference

Status quo: The German Data Forum (RatSWD) has organised seven highly visible Conferences for Social and Economic Data following a multi-stakeholder approach. The conferences successfully brought together research data-driven academia and non-academic data providers, representatives from public administration, and science policy in general.[27] For the German Data Forum (RatSWD), the conferences have been a valuable channel for community building, community-based agenda setting, and communication of outcomes.

Objectives: This multi-stakeholder approach is necessary for the data landscape in the KonsortSWD-disciplines. Hence, the conference design serves as a blueprint for a Mid-term Consortial Conference of KonsortSWD. The conference provides an avenue for 1) networking between academic communities, administration, and political decision-makers (Milestone 1 and 2), 2) assuring the quality of already developed (and developing) KonsortSWD services (Milestone 3), 3) finding out about further needs that could be addressed by KonsortSWD within the communities (Milestone 2 and 3), 4) to support cultural change for data sharing among research communities and representatives of governmental, administrative, or funding institutions (Milestone 2), and 5) to connect to and communicate with other NFDI consortia and NFDI governing bodies.

[27] Documentation can be found here: https://www.ratswd.de/en/info/kswd (19 September 2019)

Timetable:

Year	Milestones
2	[M1] Preparation of the conference: content (general theme and detailed topics, key speakers, panels, etc.) and administration (conference venue, broad invitation, etc.)
3-4	[M2] Organisation of the conference, documentation (collection of feedback, collection of new issues, use of digital and analogue PR channels around the conference) [M3] Documentation of activities for TA 1 Community Participation and StC

Risks of implementation: The conference concept is established. Implementation risks correspond with the general risks of organising and hosting large events.

Distribution of responsibility and contribution of applicants: The measure will cooperate closely within TA 1 Community Participation, with the other TAs, the Steering Committee, Council for Data Access and the further participants and supporters of KonsortSWD. It is administered by the Berlin Social Science Centre (WZB). The responsibility for the operation rests with the co-spokesperson in charge of TA 1 Community Participation.

Resources: Costs for the venue, catering, advertisement, speakers, and documentation.

4.3 Task Area 2: Data Access

Task Area 2 (TA 2) bundles, continues, and develops important activities aiming at an integrated and harmonised data access. With currently 34 research data centres (RDCs) accredited by the German Data Forum (RatSWD) there exists a growing network (cp. Figure 7). It is structured by the Committee for Data Access (CDA, currently: FDI Ausschuss, Committee for Research Data Infrastructure). CDA and TA2 mutually complement each other and shape the coordinated development of the established, highly effective distributed infrastructure (RfII 2016, pp. 30-33., Wissenschaftsrat 2011, pp. 81-82) in the social, behavioural, and economic sciences.

The CDA holds two meetings per year for the community of research data centres (RDCs) and retain its established organisational structure (RatSWD 2017d, see also Figure 5 for cooperation with the German Data Forum (RatSWD)). Membership of the CDA consists of the heads of the RDCs accredited by the German Data Forum (RatSWD). Many represent non-academic data providers that hold highly relevant data for the social, behavioural, educational, and economic sciences such as the Federal Statistical Office (Microcensus), the Federal Employment Agency, the German Statutory Pension Insurance Scheme, the Robert Koch Institute, and Germany's federal motor transport authority (*Kraftfahrt-Bundesamt*). A full overview is published in the RDC Activities Report (RatSWD 2019a).

All efforts of TA 2 and the CDA will aim at safeguarding the continuous and sustainable improvement of the research data infrastructure, which includes advancing the quality and quantity of available data and facilitating data access for researchers.[28]

The network of accredited RDCs provides its users a wide base of research data which are predominantly of sensitive nature. In 2018, almost 4,000 datasets were provided to more than 46,000 users - with a large proportion of international users (cf. RatSWD 2019a). Over the past ten years, the CDA has substantially improved the data infrastructure in three key areas: first, RDCs have generated significant expertise on data protection and successfully facilitated access to sensitive data by setting up guest researcher workstations and procedures for remote execution and remote access. Second, RDCs have improved data documentation and the use of standardised metadata to improve the findability of their data in international catalogues. This involves contributions to the implementation of the international Data Documentation Initiative's (DDI) metadata standard. Most RDCs' datasets are now registered in international catalogues. Third, RDCs have established a reporting and monitoring system. This included setting up a complaints office at the German Data Forum (RatSWD), where users can report alleged infringements of accreditation criteria. To maintain this infrastructure, KonsortSWD proposes to bring the German Data Forum (RatSWD) and the RDC network under the umbrella of NFDI.

Objectives: TA 2 bundles, continues, and develops infrastructures and services that assist researchers in responding to social and economic challenges. The infrastructures and services relate to efforts to ensure and enhance the quality and FAIRness of (sensitive) data (KonsortSWD objective 1), the growing number of data types (objective 2), to coordinate data and metadata generation and standardisation to achieve interoperability (objective 4), and to enhance the availability of privacy-sensitive and other restricted data for research purposes (objective 6). Within NFDI, the RDC's capabilities will be enhanced while preserving their strength as a federated infrastructure. To provide user-friendly services to a broader community KonsortSWD aims at opening up RDCs' expertise and services in RDM to other users (e.g., other disciplines or universities) and to transfer know-how into the scientific communities (cf. TA 1 Community Participation and TA 3 Data Production). TA 2 comprises six Measures:

1. Harmonising RDC processes. This will lead to more transparent access paths. Despite different legal restrictions and institutional setups, data users should find similar structures when contacting different RDCs (see TA.2-M.1 RDC Support).

[28] https://www.ratswd.de/en/data-infrastructure/fdi (25 September 2019)

2. Facilitating access to sensitive data by creating a nationwide network of access points (e.g., guest researcher workstations at RDCs). This will reduce expenditures (costs and travelling) for data users (see TA.2-M.2 Access Points).
3. Establishing a federated archiving infrastructure. This opens up the RDCs' archiving and curation services to the scientific community and broadens the range of data archived at the RDCs. As a first step, qualitative data, which have been largely excluded from secondary use, will be systematically included in the RDC network (see TA.2-M.3 Federating RDCs).
4. Improving qualifications of RDC staff and data users in the disciplines covered by KonsortSWD. This meets increased demands of RDM (see TA.2-M.4 RDM Skills).
5. Establishing unified means of interaction for supporting users of different RDCs. This will facilitate access to support relating to data (see TA.2-M.5 Transparent Interaction).
6. Making relevant new data FAIR through competitively awarded RDM grants based on their value for research to the disciplines represented by KonsortSWD (TA.2-M.6 RDM Grants).

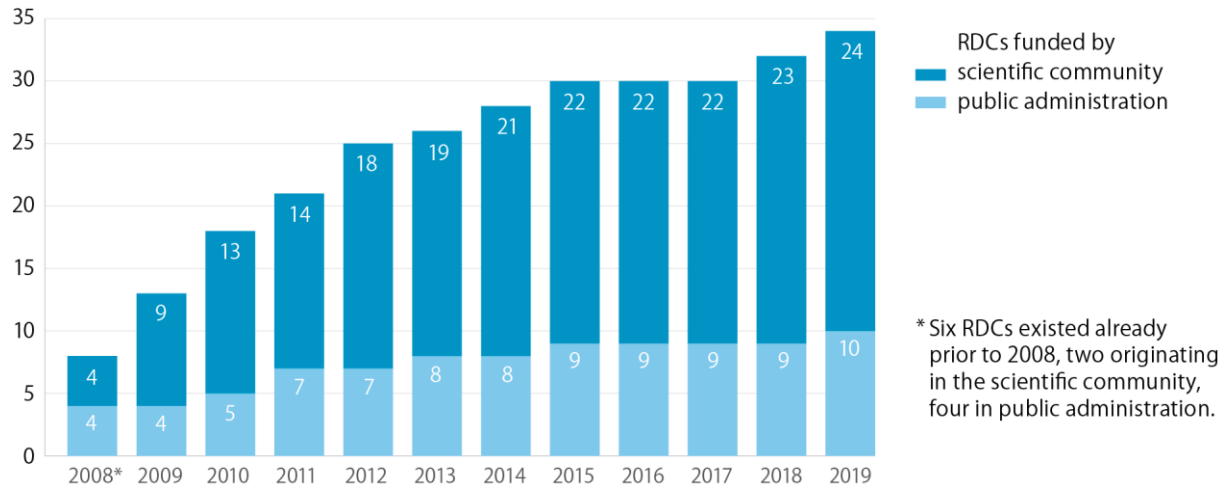

Figure 8: Number of RDCs accredited by the German Data Forum (RatSWD) since 2008

TA 2 relies on the CDA to maintain and invigorate the indispensable dialogue between RDCs funded by the scientific community and public administration (e.g., from statistical offices and social insurance providers). Beyond that, the CDA will have two permanent core tasks: first, continuously developing the research data infrastructure to expand the amount of available data and further facilitating access to data (e.g., by accrediting new RDCs, see the dynamic development of the RDC network in the past as in Figure 8). Every accredited RDC will join the RDC network and thus participate in KonsortSWD. All can benefit from KonsortSWD services and contribute to an improving research data infrastructure. Second, the CDA functions as a forum for knowledge transfer between the RDCs with a focus on legal and technical issues – as well as on the operative implementation of recommendations and development goals in RDM (e.g., GO FAIR, EOSC, RDA).

Membership and governance**:** TA 2 establishes communication between RDCs providing access to research data that arise from research contexts (e.g., survey programmes) and non-research contexts (e.g., administrative data). The services of TA 2 aim at a further improvement of access to these data for the scientific community. To ensure this communication, the working schedule of each of the Measures is accompanied by workshops, which will be organised both on the TA level and on the Measure level. Coordination is provided by the co-spokesperson. This includes monitoring at an operational level and coordination with other TAs through the Steering Committee (StC). There will be two annual meetings of the individuals in charge of the separate Measures to exchange information and coordinate activities. In addition to a regular exchange with TA 1 Community Participation – in the StC, through mutual guest status in regular meetings, as well as through the established cooperation around accreditation processes with the German Data Forum (RatSWD) (TA.1-M.1, see also Figure 5) – TA 2 will have two channels for user participation: first, many RDCs have user boards that help channel specific demands. The CDA will offer a forum to identify user demands common to different RDCs and develop concerted answers. Second, the TA 2 co-spokesperson will be named on KonsortSWD's website as the contact person for user input on services within TA 2. Finally, information on the progress of TA 2 Measures will be discussed in CDA meetings. The demands of RDCs and their users can be communicated via TA 2 to the StC.

Outreach: Challenges relating to access and handling of sensitive data must be addressed across consortia for the entire NFDI. TA 2 provides services for data access that are important for the social, behavioural, educational, and economic sciences. This service portfolio can encourage developments within the NFDI.

Resources: To maintain a close dialogue between the Measures within TA 2 and to organise the work of the Committee for Data Access (CDA), 0.75 postdoctoral researcher and 0.5 administrative support position are budgeted. TA 2 will finance two meetings of the CDA per year.

Measure TA.2-M.1 Supporting research data centres by establishing FAIR Principles and certification for sustainable and trustworthy data access

Status Quo: All research data centres (RDCs) handle sensitive personal data. But they differ on several dimensions, for example, the kind of services provided to data users, processes, or the legal frameworks under which data have been collected or shall be re-used (EU GDPR or the German Social Security Code). These differences must be taken into account when establishing and running RDCs. RDCs currently have heterogeneous workflows and only little

standardised material as well as few process descriptions at their disposal. Moreover, a crucial topic linked to processes is the accreditation of RDCs by the German Data Forum (RatSWD) (national) and the Core Trust Seal certification (CTS, international). Successful accreditation and certification indicates that an RDC is trustworthy in handling data, which is especially important with respect to the long-term archiving of sensitive data. There is a demand for support during the accreditation and certification processes.

Objectives: This Measure supports RDCs with implementing high-quality standards for work processes and services. It stimulates a reciprocal learning process by developing a reference model for FAIR data management and data access across RDCs. The Measure also supports the implementation of the reference model in RDCs by five specific objectives including accreditation and certification processes for new and existing RDCs.

RDCs will be supported by suggesting reference processes to maintain quality standards. The processes include data access, data intake, and data handling within the RDC. The FAIR principles and the Open Archival Information System (OAIS) reference model are crucial for RDC processes. This Measure collects process descriptions from the RDCs and formulates descriptions for standard central processes in line with FAIR and OAIS. This will be achieved in close cooperation with TA.5-M.3 (Interfaces). Documents used by the RDCs for typical processes are collected and will provide the basis for new reference documents, which will be applicable to national and international data users and depositors. These include collection policies, data archiving contracts, data use contracts (for EU GDPR and the German Social Security Code), contract extensions, and data deletion notifications. The contracts must be checked by specialised lawyers. In the first step, processes and related documents for data access conditions in particular are processed (Milestone 1). In the second step, processes and related documents for data input conditions and long-term preservation are processed (Milestone 2).[29] Recommendations will be developed explicating the accreditation process, for which currently no structured preparation process exists. A concept and supplemental material for a workshop for RDCs preparing for the accreditation process will be created. A pilot workshop and regular workshops with updated material will be offered (Milestone 3). The Measure aims at producing material and a workshop format to support RDCs in obtaining Core Trust Seal (CTS) certification. CTS certification verifies a wide range of requirements crucial for trustworthy data repositories (e.g., processes, data integrity, data discovery, data re-use) and thus also provides the foundation for national and international cooperation. The Measure

[29] Qualitative data require specific RDM and fitted documentation with respect to data access and data intake. Respective standards will be developed in TA.2-M.3 (Qualitative RDM).

delivers reference documents for CTS, an instructions manual for the accreditation process, and a workshop format (Milestone 4).[30] Additionally, RDCs will be supported while they undergo the accreditation and certification processes (Milestone 5).[31] The material will be disseminated to the KonsortSWD's RDC network. This Measure strives to support all RDCs (seeking accreditation as well as accredited).

Timetable

Year	Milestones
1	[M1] Create RDC reference work processes with respect to FAIR and OAIS for quality assurance *focusing on data access (data access conditions, processes and tasks within an RDC)* and create related reference documents for RDCs *(data use contracts, contract extensions, data deletion notifications)*
2	[M2] Create RDC reference work processes with respect to FAIR and OAIS for quality assurance *focusing on data intake and long-term preservation (data intake conditions, data handling within the RDC, long-term preservation)* and create related reference documents for RDCs *(collection policies and data archiving contracts)*
3	[M3] Recommendations and workshops to advise on the German Data Forum (RatSWD) accreditation
4	[M4] Reference documents, instructions manual, and workshops to advise on Core Trust Seal accreditation
5	[M5] Support RDCs during German Data Forum (RatSWD) and Core Trust Seal accreditation processes

Risks of implementation: Reluctance of RDCs to achieve CTS certification could challenge achieving the Measure's objectives. However, already at least six RDCs have reported to strive for CTS certification.

Distribution of responsibilities and contribution of applicants: DZHW is responsible for Milestones 1, and 3, DZHW and GESIS jointly for Milestones 2 and 5 (GESIS especially for long-term preservation). GESIS is responsible for Milestone 4. Host institutions will provide rooms for workshops. The administrative fee for CTS accreditation will be covered by the institutions applying for accreditation.

Resources: For five years: 1.25 postdoctoral researchers, student assistants, travel expenses, workshops and meetings, legal counselling, and translations.

[30] Qualitative data require specific RDM and fitted documentation. Respective standards will be developed in TA.2-M.3 (Qualitative RDM).
[31] The CTS certification process for the RDCs must be prepared well in advance. We will work in close cooperation with Measure TA.2-M.3 (Qualitative RDM). Together, we aim to prepare six RDCs for CTS certification. Further RDCs will be supported by Measure TA.2-M.1 (RDC Support).

Measure TA.2-M.2 Creating single points of access for sensitive data in RDCs

Status quo: Ensuring the maximum possible and most convenient data availability for research purposes while at the same time ensuring compliance with high standards of data security and data protection is a central challenge for all providers of datasets at the individual level. A widespread approach to deal with this problem is to set up different access paths to research data in a gradual manner according to data protection requirements. Depending on the extent to which the use of data can be monitored, the different access paths also permit the provision of correspondingly more sensitive information. Factually anonymised microdata can usually be downloaded by users as Scientific Use Files (SUF) after a contract has been signed. Researchers who need more detailed information for their research are forced to use specialised secured guest researcher workstations (GRWs, *Gastwissenschaftsarbeitsplätze*). Currently, more than 28 research data centres (RDC) accredited by the German Data Forum (RatSWD) provide such GRWs (RatSWD 2019 a, p. 24). They are used for data that require a secure access point and cannot be accessed via remote data access. More than 300 datasets are exclusively available to researchers at these GRWs. A total of 732 scientists used a GRW in 2017 (RatSWD, 2018 c, p. 17); most of them had to cover travel expenses.

The use of a GRW by researchers requires extra effort in cost and time due to the necessity of physical presence on the RDC premises. In addition, working at a GRW is not as flexible, as local usage time is limited, longer stays are usually associated with additional costs, and special security arrangements apply (e.g., no internet access, use of smartphone or laptop prohibited at the GRW). This situation is a significant hurdle for potential data users, especially students and researchers from abroad. Consequently, a substantial part of the existing research data potentials remain unused.

Objectives: The objective of the Measure is to improve access to sensitive data by connecting the existing GRW into a network of ‘single points of access.’ Such a research data infrastructure network (RDCnet) utilises the distributed availability of accredited RDCs as an advantage. This allows researchers to access all formal or weakly anonymous data - regardless of the GRW they are working at. The connection between data of a certain institution that is only available on-site and its exclusive accessibility via a GRW at the specific location is no longer required.

At present, only very few RDCs have access to sensitive data from other institutions and only within very specialised and restricted settings. This Measure aims at an ‘upgrade’ of the so-called ‘RDC-in-RDC’ model, such that the data of all participating institutions are available from each GRW (Milestones 1 to 5). In doing so, RDCs could reach a significantly larger number of

users. A nationwide RDCnet combines the possibilities of decentralised data access with the two conditions necessary for this type of data: a) the data providers keep control over when and where (and by whom) their data are accessed (Milestones 4 and 5), and b) data access from a controlled endpoint is assured even with remote access (Milestones 1 to 3).

The infrastructure created by this Measure can be extended further following the expansion of remote access options. Technically, the network architectures created in the first three years are limited to the approved controlled endpoints at on-site GRWs at RDCs, but can also be accessed from the researcher's workstation at a later stage if the data provider allows this (Milestone 11). The structures created in this way can therefore be modified should access criteria change in the future (Milestones 6 to 8). However, each institution should have the possibility to join the network at the time the predefined access conditions are met (‘RDCnet at different speeds‘, Milestone 10).

A website to be integrated in the overall KonsortSWD website will serve as an information platform for the RDCnet's data and regulations as well as a single registration point for GRW work in the RDCnet (Milestone 9). In accordance with the FAIR principles, the RDCnet not only contributes to improve access to sensitive research data (‘accessible’) but also supports the findability of data (‘findable’). In addition, the RDCnet offers perspectives for the enrichment of already available research data by linking it with information from other sources, for example, with appropriately prepared regional indicators (‘interoperable’), and extending the collection with previously inaccessible research data by providing a suitable infrastructure for their use (‘re-usable’).

The RDCnet will always be open for the inclusion of additional GRWs. For data-holding institutions that do not yet have on-site access, the RDCnet can provide an additional incentive to set up GRWs due to its networking infrastructure. With the successful implementation within the framework of KonsortSWD, the established network of distributed GRW is also open to other providers of sensitive data in general, either for use or as a model for setting up specific networks of the research data network. The necessary standardisation of prerequisites and processes is ensured by the development of a transparent catalogue of criteria for the recognition of secure GRW and by an accreditation procedure linked to it (Milestones 1-3).

The internationalisation of data access in RDCnet must be taken into account from the very beginning. Some initiatives in the European context, in particular, have started to optimise international access to microdata. These include the ‘Data without Boundaries’ project (FP7, 2011-2015), the ‘International Data Access Network’ (IDAN), the ESSnet project ‘Decentralised

and Remote Access to Confidential Data in the ESS' (DARA), the OECD's 'Expert Group for International Collaboration on Microdata Access,' and the 'International Network for Exchanging Experience on Statistical Handling of Granular Data' (INEXDA), which was founded in 2017. The expertise gained in these projects with regard to the technical, organisational, and legal possibilities of providing international microdata must be taken into account when designing RDCnet in order to prevent the development of a nationally isolated solution disconnected from international research data infrastructures.

Timetable

Year	Milestones
1	[M1] Development of a catalogue of criteria for the approval of secure GRW [M2] Agreement on an accreditation procedure for GRW certification, definition of the organisational and legal frameworks for RDCnet [M3] Analysis of the requirements for the necessary network structure
2	[M4] Identify and provide secure and fast procedures for data exchange between the RDCs involved in RDCnet [M5] Design and pilot implementation of a suitable network structure for RDCnet
3	[M6] Operation of a pilot network and any necessary adjustments (test user survey) [M7] Announcement of RDCnet within the scientific community (Dissemination) [M8] Develop a business model to ensure long-term operation and a possible balancing of resources between the participating RDCs
4	[M9] Create a website as an information platform for RDCnet's data and regulations as well as a single registration point for GRW work in the RDCnet (Website) [M10] Establish an FDI core network open to external participation among the institutions involved in the work package
5	[M11] Analysis of the requirements for RDCnet to provide an infrastructure that allows linkage of research data across RDCs

Risks of implementation: Future changes in the hard and software equipment could be problematic in a very heterogeneous FDI network.

Distribution of responsibilities and contribution of applicants: Responsibility lies with the Measure Lead (SOEP at DIW Berlin). Hardware equipment for running server and client, the implementation of the local GRW into the RDCnet as well as on-site organisational support of the users are provided by the applicants will be provided (cf. letters of commitment, appendix 3).

Resources: For five years: 1 postdoctoral researcher (coordination and organisational workflow development). For four years: 1 computer scientist (development and implementation of network structure), 0.5 social scientist with expertise in research data management (RDM) (integration into RDC workflow, user RDM support). Additional direct projects cost for additional website service and for annual network meetings, travel costs, and IT security expertise.

Measure TA.2-M.3 Federating the archiving infrastructure of qualitative data

Status quo: Unlike in quantitative empirical research, where data archiving and re-use have been established for several decades, the archiving and re-use of data in qualitative social research is still the exception. So far, only few RDCs started to establish RDM procedures and tools to professionally curate and prepare qualitative, semi-standardised, and unstructured data such as interviews, observational data, field notes, text, visual data, and audio files, which most often comprise sensitive personal data for re-use and secondary analysis. Among these RDCs are the RDC Education at the DIPF, RDC Qualiservice, RDC eLabour, RDC AGD, RDC DZHW, and RDC BO at DIW Berlin. Moreover, the RDM landscape consists of various small, unconnected and highly specialised archives at universities and other institutions. To date, each RDC usually dealt with challenges related to data curation on its own, for example, displaying of metadata, data citation principles, contracts, policies, and licences. At the same time, researchers who are willing to deposit their data often struggle to identify a reliable RDC that offers satisfactory conditions for archiving and disseminating their data. Secondary users often cannot find appropriate data for their research purposes.

One solution to overcome this fragmentation is to create federated data networks. The participating RDCs establish a coordinated approach to data curation defining which RDC is responsible for which types of data. Workflows and quality standards are harmonised and there is a common information portal for users. Expertise for data types and research methods lies with the RDCs. The *Framework Programme for the Promotion of Empirical Educational Research* (*Verbund Forschungsdaten Bildung,* VerbundFDB) was the first successful subject-specific network of accredited RDCs in educational research. It harmonised services, work processes, and regulations, such as user agreements, while, at the same time, ensuring that data are archived and disseminated according to the requirements of each specialised RDC. The achievements and experience of VerbundFDB can serve as a model for the coordination and harmonisation of services provided by other subject or domain-specific networks of RDCs.

Objectives: To promote access and re-use of qualitative data in research and academic teaching and to encourage researchers to deposit their data in research data repositories, KonsortSWD will establish a community-centred federated data network for qualitative data (VQualidat) that replaces the heterogeneous and fragmented landscape with a coordinated, user and service-oriented infrastructure. The network will establish a transparent portfolio of data curation services with harmonised workflows and quality standards (Milestone 1 and 4) and with a central access and information point for researchers covering all types of qualitative data (text, audio, video; Milestone 2). To increase data visibility and address the needs of

researchers, the participants of VQualidat will cooperatively develop an internationally compatible core metadata schema and controlled vocabulary for the various qualitative data (Milestone 3 and 6). The network will enhance data quality, create synergies, and build a service-oriented and trustworthy data infrastructure (Milestone 9) thus fostering data sharing and re-use in qualitative social research.

In accordance with the development of VerbundFDB, two principles guide its implementation: 1) the RDC with most substantive, methodological, and technical expertise for a given data type curates these data and is responsible for their digital long-term preservation and 2) the services for external data providers meet the same quality standards as apply to internal data of the RDCs. The quality standards will be standardised within VQualidat (Milestone 1 and 4).

To promote the integration of RDM into the qualitative research process, standards will be developed in close cooperation with TA.3-M.2 (Qualitative RDM). Both Measures together will offer RDM solutions covering the whole lifecycle of qualitative data and establish a support centre on services (Milestone 2), and funding sources for researchers as well as harmonising workflows for data submission (Milestone 1) and production. The project thus strengthens recent developments towards a culture of data sharing in qualitative research. Moreover, the data pool will be expanded by successively including both small data archives, which are found in institutions and collaborative research centres (Milestone 7), as well as research data management at universities (Milestone 5). All code produced by this Measure will be made available as open source.

Timetable:

Year	Milestones
1	[M1] Development of a functional model for the cooperation within the VQualidat network (coordinated procedures and processes with defined quality standards in accordance with the experiences of VerbundFDB, focusing on displaying of metadata, data citation principles, contracts, policies, and licences) [M2] Implementation of a portal for VQualidat with an interoperable information architecture, including a place for consulting and submission (in cooperation with TA.5-M.4 User-Centred Design)
2	[M3] Creation of a core metadata schema and controlled vocabulary for qualitative data for increased data visibility and a fine data fit, starting with harmonisation of selected procedures of data administration and curation (data citation principles, contracts, policies, and licences [M4] Transition to operations and testing the hitherto achieved workflows for data re-use
3	[M5] Establishment of a broadened network encompassing more archives and institutes and extension of VQualidat's consulting services to university research and education [M6] Extension of the core metadata schema and controlled vocabulary for the description of qualitative data including at the international level, extended harmonisation of selected procedures of data administration, and curation (data citation principles, contracts, policies, and licences)
4	[M7] Inclusion of additional partners and external metadata from smaller data producers and data holders [M8] Measures to promote data sharing and foster data re-use such as secondary analysis in the specific national communities

5	[M9] Coordinated preparation and certification with the Core Trust Seal of the VQualidat-RDCs as quality standard for international compatibility

Risks of implementation: Milestone 9 depends on the successful implementation of Measure TA.3-M.2 (Qualitative RDM).

Distribution of responsibilities and contribution of applicants: Responsibility for implementation lies with each participating RDC. Applicants will contribute Hardware and software packages

Resources: For five years: 1 postdoctoral researcher (coordination), 0.5 doctoral researcher (information transfer with VerbundFDB for 2 years), 0.83 doctoral researchers, 0.3 postdoctoral researcher (information science), 0.25 librarian (metadata), 0.5 student assistant for support, website, travel costs, network meetings, workshops.

Measure TA.2-M.4 Developing and exchanging research data management skills

Status quo: The German Council for Scientific Information Infrastructure (RfII) states that the digital transformation and 'scientification' of infrastructure tasks will shape the future of research data infrastructures. To accommodate these changes, infrastructure staff requires specific skills (RfII, 2019). A survey among research data centres (RDCs), conducted in August 2018, showed that many applicants for positions in RDCs lack specific research data management (RDM) skills as well as RDC-specific competencies, including ethical and legal aspects of RDM as well as data documentation skills (RatSWD, 2018 c, pp. 21-34). These skills are typically not part of classic methodology curricula in most academic disciplines represented by KonsortSWD. At the same time, there are often no incentives for developing these skills (RfII 2019). Moreover, we expect an increasing demand for training and individual knowledge acquisition as only with up-to-date software and IT competencies (e.g., database systems, programming languages, web design) staff will be able to adapt to changing technical and methodical requirements of RDM.

Objectives: To foster acceptance of RDCs and, more generally, data sharing, it is essential that RDCs provide trustworthy, robust, and high-quality services to their users. It is therefore necessary to promote the competencies of research data managers. This Measure's goal is to deepen the cooperation of RDCs (see section 3) by jointly enabling the development of skills and promoting the occupational profile of a research data manager. Strengthening RDM skills is beneficial not only for working at RDCs but also for non-academic employers who value data literacy skills.

The first Milestone of this Measure is to establish a central teaching and information platform, in which teaching materials from RDCs are collected and made available and which will be integrated in the overall KonsortSWD website. The course material will be developed and transformed into online courses and the platform will also provide the opportunity to host webinars as well as information on existing face-to-face training opportunities. The platform will be adequately advertised and maintained in the long term. Target groups are both RDC staff and individuals outside RDCs such as university archiving initiatives or research institutes, people looking for information for their own research data work, and those offering training and looking for teaching material. User and download statistics will be used to evaluate the usage of the material and can be complemented by direct surveys.

A second milestone that we plan to reach over the course of two years is to establish a collaborative training programme based on the competencies provided by the RDCs for RDC staff. The core of this programme will be an annual workshop by RDCs for RDCs covering various aspects of RDM (e.g., anonymisation of survey data, management of text-based and audio-visual data, etc.) to share expertise with each other. Further workshops can be organised when needed in the form of webinars using the online platform. We will also identify training programmes from external providers that can help RDC and university staff to close knowledge gaps. Furthermore, we will research existing study programmes to assess to what extent these contents are currently covered in the social sciences (Milestone 3).

Based on the existing expertise at the RDCs, it is planned to develop a certification course (Milestone 4) within the next five years that will prepare future employees for working at RDCs. The courses will be hosted by this Measure. Since course topics will be based on the RDCs' own expertise, it will be a strong signal for potential employers at the RDCs and will fill the skills gap that candidates currently have when applying to RDCs.

In a further step, we plan to develop teaching material for universities (Milestone 5). This will contribute to integrating RDM into university curricula (see below), including a seminar on RDM, which faculty staff can use for university teaching without making major adjustments. These training opportunities and teaching materials will be developed in cooperation with the Committee for Data Access. The staff will promote the developed material at universities.

To ensure that research data managers are adequately trained to meet the job's practical requirements, the developed training materials and programmes will feed into a set of recommendations for higher education institutions. To this end, we will reach out to universities and higher education institutions to promote these recommendations. We also offer to teach

one RDM seminar session each at about 10 universities and develop and offer online tutorials and webinars for students through the online platform (see Milestone 6).

RDCs also offer opportunities for dual apprenticeships. However, typical activities at RDCs do not always match the profile of traditional apprenticeships, resulting in individual deficits. To compensate for these deficits, we will offer an exchange within the RDC network (Milestone 7) to involve apprentices rotating and working at other RDCs for up to 7 weeks.

This Measure implements the recommendation of RfII (2019) to build training alliances and offer certified courses in order to sustainably develop RDM skills as well as to address objective 5 listed in section 2.1.

Timetable:

Year	Milestones
1-2	[M1] Implement and equip an online information platform to offer courses and make teaching material from RDCs available [M2] Establish a training programme for RDC staff [M3] Assess current study programmes regarding gaps in addressing social science topics and RDM [M5] Develop teaching material for a university seminar on RDM
3-5	[M4] Establish a certification course on RDM for students, young researchers, and professionals [M6] Develop and implement an RDM online tutorial and webinar for students, reach out to universities and higher education institutions, and teach one RDM seminar session each at around 10 universities [M7] Establish an exchange programme for apprentices among the RDCs

Risks of implementation: The Measure relies on the cooperation of RDCs which offer apprenticeship training and of the apprentices. Another risk is that there may be limited interest in the teaching material to be developed.

Distribution of responsibility and contributions of applicants: GESIS and ZPID are jointly responsible for reaching the Milestones. RDCs contribute to the online information platform by providing teaching material ready to be shared. ZPID will provide the web space needed to host the online platform. A free-of-charge RDM workshop for RDC staff will take place once a year. Ten RDCs will contribute to this workshop series (see list above – participating institutions). RDCs will host the courses and provide lecture rooms, required hardware and software, etc. SOEP at DIW Berlin will organise the apprenticeship exchange programme and receiving institutes will devote resources to on-the-job-training for the visiting apprentices. Most RDCs declared their interest to take advantage of these offers (see appendix 3).

Furthermore, GESIS contributes input from the Consortium of European Social Science Data Archives (CESSDA), most notably the CESSDA Strengthening and Widening project which developed a competence framework (CESSDA 2017). ZPID has prior experience in the setting-

up and long-term maintenance of RDM information platforms, such as the DataWiz knowledge base (Kerwer et al., 2017). Besides this technical expertise, ZPID has worked on the research-based conceptualisation of an information architecture designed to implement RDM processes at universities and research institutions within the PODMAN project (Blask and Förster, 2019; Förster et al., 2019).

Resources: 2 postdoctoral researchers for 1) the conceptualisation, establishment, and long-term implementation of training programmes, online tutorials, and webinars for RDC staff and university archiving initiatives and 2) the development and implementation of a certification course for students and graduates. 1 PM per year is devoted to a trainee for the organisation of the apprenticeship exchange. 2 part-time postdoctoral researchers for the equipment of the RDC training platform. After development, 0.5 postdoctoral researcher for maintenance. 2 x 0.5 postdoctoral researchers for 1) the assessment of study programmes and 2) the development of recommendations, university seminars, online tutorials, and webinars for universities. Further direct project costs will include a service for webinars, travel expenses for project meetings and face-to-face workshops, travel expenses and mobility allowances for the apprenticeship exchange programme, and funds for outreach material.

Measure TA.2-M.5 Improved use of research data through transparent interaction

Status quo: The research data infrastructure in the social, behavioural, educational, and economic sciences is characterised by a broad, diverse, and constantly expanding range of available data. (Re)usability of these data, however, not only depends on their availability but also on the extent and quality of support given to researchers dealing with them. Sharing the necessary knowledge is a responsibility of data providers; guaranteeing user services is one of the criteria for RDC accreditation with the German Data Forum (RatSWD). Traditional support tools range from documentation material and tools to training courses, tutorials, hotlines, and FAQ lists. These support tools are currently 1) heterogeneous in scope and user-friendliness, 2) usually tailored to one particular data product, 3) only partially transparent, and 4) in many cases not permanently available. A fundamental problem is that the typically bilateral exchange of additional information ignores the enormous potential of knowledge and experience exchange among the users.

Objectives: This Measure will address these limitations by setting up and permanently operating a joint online discussion platform (cf. Griesbaum et al. 2015) for data providers and data users which is to be integrated in the KonsortSWD website. The main advantage of such a forum is that communication would be completely transparent. This renders 1) the search for

information more efficient, 2) the relevant expertise from data providers *and* data users available in a structured and sustainable way, and 3) data infrastructures and their user communities more visible in general. A platform would create a freely accessible knowledge archive that promotes professional handling of complex research data. At the same time, it enables continuous feedback to data providers. The transparency of interactions may also contribute to standards setting (e.g., for data preparation or citation). Integration of data users into one network and an 'open science' discourse are important motivations for the project.

The Measure builds on existing support services at the RDCs (in accordance with the 'deepening' logic of the RDM strategy – see chapter 3) but extends their reach by providing a platform for interaction, and integrating relevant data providers. These data providers (participating institutions) will define the necessary functionalities in a series of workshops. The overall objective is to create a generic structure in which other data providers from inside and outside of KonsortSWD can participate. The design of the platform will be developed in cooperation with TA.5-M.4 (UCD). The resulting code will be made available as open source.

The project starts with the identification of necessary and desirable functionalities from the perspective of data users and data providers (Milestone 1), the clarification of procedures and data protection issues (Milestone 2), and initial implementation of the forum (Milestone 3). In the second year, the focus is on testing the tool, including feeding in basic content (Milestone 4). Following the introduction of a user administration model and general usage rules (Milestone 5), the forum will be activated for the public (Milestone 6). The next steps are to establish the forum by stimulating user traffic (Milestone 7) and by promoting the tool within the respective data user communities (Milestone 8). The first project phase will conclude with a relaunch of the forum to adapt the software to additional requirements (Milestone 9). The fourth project year is marked by user evaluation (Milestone 10) and an opening of the forum for other data providers (Milestone 11). At the end of the second project phase, a strategy for both the permanent operation of the forum (Milestone 12) and its continued opening (Milestone 13) is to be defined and implemented.

Timetable:

Year	Milestones
1	[M1] Specification of technical requirements [M2] Agreement on joint commitment among the participating institutions [M3] Release of a beta version of the forum
2	[M4] Trial operation of the forum and critical assessment [M5] Introduction of a user management system and a code of conduct [M6] Release of a public version of the forum

3	[M7] Transfer of the forum tool into permanent operation [M8] Activities to advertise the forum, including conference presentations, etc. [M9] Relaunch of the forum with new functionalities and interfaces
4	[M10] Conducting a survey among forum participants for user feedback [M11] Integration of new national and international partners into the forum, including the provision of instructions for new forum managers
5	[M12] Joint operating model for the future organisation and management of the forum [M13] Continuation of the opening and promotion of the forum

Risks of implementation: The success of any online forum depends on its use. Only a forum that is constantly provided with new content serves the purpose of an exchange platform. The main risks are the lack of a response on the part of data users or of activity from data providers. Both factors are manageable: existing experience in RDCs and feedback from several user surveys indicate that researchers are willing to use such a service. It is also in the data providers' own interest to make their data as attractive as possible through reliable and transparent user services. With SOEP, GESIS, DZHW, IAB, IQB, DJI, and pairfam, important data providers are participating in the measure. These partners represent major data collections (particularly longitudinal surveys) in the social, behavioural, educational, and economic sciences. They have an extensive and constantly growing number of data users and are already networked to a high degree. Each data provider must make a commitment to be actively involved on the forum.

Distribution of responsibility and contribution of applicants: The responsibility for developing, implementing, maintaining, and moderating the online forum according to the above-mentioned Milestones rests with the Measure lead institution (LIfBi). It supports the measure with assistance from relevant departments such as IT, data protection, and public relations. LIfBi also provides the necessary facilities, such as a server infrastructure, workstations, and resources for student assistants. The participating institutions contribute their expertise and needs to the concept development. All partners are actively engaged in the operation of the online forum and communicate it to their user communities.

Resources: 1 postdoctoral researcher for content coordination and for development, 0.9 other staff (for technical implementation) for years 1-3, and 0.4 other staff for years 4-5. Funds are also needed to cover software support and travel expenses.

Measure TA.2-M.6 RDM grants to foster cooperation between RDCs and research communities

Status quo: Research data centres (RDCs) disseminate data to research communities based on the highest standards for anonymisation, documentation, and secure data access. Researchers collecting data at institutions without an RDC often cannot benefit from an RDC's domain-specific RDM competencies because funding is unavailable for professional data management (mainly improving accessibility and interoperability as well as re-usability) once the data have been collected. Even if funding extends to cover data collection, research data management plans that include professional RDM are rare (DFG 2019). Opening the RDCs for cooperation with external data producers would therefore generate substantial benefits for researchers in KonsortSWD-related disciplines.

Objectives: This Measure stimulates and facilitates cooperation between researchers and RDCs for sharing complex and sensitive research data for secondary use. Individual researchers from institutions without an RDC specialised in their field of expertise will team up with accredited RDCs. In tandem, they then apply for funding for an RDM grant for one data project. Projects eligible for funding address 1) design of a data management plan, 2) concepts for anonymisation, 3) (transnational) access to sensitive data for secondary users, 4) high-quality metadata documentation, and 5) digital long-term preservation. A condition for funding is that the cooperating RDC agrees to publish the data for secondary use.

The Measure fosters knowledge transfer from RDCs to the research communities. It will also support cross-consortia data integration within NFDI (cf. section 2.3) for datasets linking data from several consortia with data relevant to the KonsortSWD community. Moreover, the funding will contribute to making sensitive data that was collected, for example, at universities, available to the scientific communities. Data gleaned from this RDM grants scheme will help KonsortSWD to anticipate and further support the RDM needs of research communities.

Applications for RDM grants will be collected through a biannual call for projects organised by the Committee for Data Access (CDA, cf. TA 2 Data Access). Outlines (max. 5 pages) will be competitively evaluated at the meetings of the CDA. Evaluation criteria are the 1) the potential scientific contribution of the dataset, 2) the added value of making it FAIR, 3) the scientific quality of the RDM project, 4) the potential for secondary use, and 5) the appropriateness of the funding requested for the project. The potential for scientific contribution will be judged based on the quality of contributions published on a given dataset at the time of application. The StC will decide about funding based on the CDA's evaluation.

Timetable:

Year	Milestones
1-5	[M1] Grants provision
5	[M2] Report by CDA

Risks of implementation: There is a minimal risk that there is insufficient demand for cooperation in research data management with RDCs.

Distribution of responsibility and contributions of applicants: The members of the CDA competitively evaluate the applications for data curation grants at the biannual meetings of the committee. The committee decides on recommendations for funding by simple majority vote.

Resources: As support for RDM projects through grants scales easily, KonsortSWD has planned funding flexibly and predominantly in years where measures with fixed costs (staff) allow for it.

4.4 Task Area 3: Data Production

Research data management (RDM) makes data available to researchers and increases the traceability and quality of empirical research. This works best if RDM systematically focuses on the technical and methodological quality of research data. Both quality aspects are determined by how research data are produced. Therefore, systematic RDM should not solely focus on archiving data and making it accessible, but always take the process of data collection into account. For a number of reasons, the documentation and quality assurance of research data poses particular challenges in the social, behavioural, educational and economic sciences. First, a very broad spectrum of data types is used for empirical research, including both structured data, for example collected in surveys, and unstructured data, such as video, audio, and text data. Second, the aforementioned sciences traditionally use data generated outside science, such as administrative data or, more recently, observational, behavioural, and tracking data. In the latter cases, the methodological quality is less transparent and additional measures are often necessary to make the data usable for science. Task Area 3 (TA 3) therefore aims to increase the quality of data production by establishing and continuously improving RDM in the social, behavioural, educational and economic sciences.

Status Quo: TA 3 addresses three challenges that the disciplines represented by KonsortSWD currently face. First, over the past 20 years, a very diverse and rich data infrastructure has been established, which offers a wide range of, survey-based, or administrative research data. A central problem of this very large stock of data, however, is that many studies and datasets use

different concepts and instruments to measure the same phenomena. As a consequence, researchers cannot fully exploit the potential for analyses. The diversity of measurement concepts and instruments also has an impact on data quality because many instruments are not tested for reliability and validity and the data are often not comparable.

Second, while secondary analysis, data archiving, and re-use in quantitative research have developed over several decades, they have only recently been discussed in qualitative social research. This means that RDM has only been introduced to these data for a short time. Procedures and tools to professionally curate and prepare qualitative, semi-structured, and non-standardised data for secondary analysis, such as interview transcripts or audio and video data, are still under development. As a result, there are still no consistent quality standards for processing unstructured qualitative research data. The amount of documented research data suitable for subsequent use that different research fields using qualitative research provide varies significantly. This lack of standards also applies to unstructured text data, which are increasingly important for the (semi-)automated analysis of text corpora in the social sciences (e.g., from social media, political debates or open ended questions in surveys).

Third, scientific usefulness and quality of research data directly depend on the availability of metadata which describe the raw data as thoroughly as possible. From the point of view of data users, these metadata must not only describe the study design, but should also contain comprehensive information on individual variables and measurement concepts. Currently, data producers generate metadata and make them available ex post and often separate from the actual data. The different technical possibilities of proprietary software used in the social sciences hampers the integration of raw data and the related comprehensive metadata directly into the data production process.

Objectives: TA 3 will establish and optimise interoperability and high-quality data documentation to make research data more suitable for subsequent use in the different scientific communities represented by KonsortSWD. In order to meet the needs of the communities, the Measures are attuned to the methodological peculiarities in data generation and the level to which RDM is already applied in the respective communities. Where standards are well-established, the focus is on increasing the analytical potential of the data and further improving their quality. Where RDM is only just beginning to emerge (e.g., text data and qualitative social science data), methodologically appropriate RDM standards will be developed and, where necessary, a cultural shift towards data sharing will be promoted.

The five Measures of TA 3 initially focus on services to the communities that relate to the three challenges described above. First, the harmonisation of the content of core social surveys facilitates the accumulation and linking of various data and replication studies (TA.3-M.1 Harmonised Variables). This Measure is supplemented by regular roundtables which coordinate the content of major survey programmes and identify content to be harmonised ex post or ex ante. Second, tools and technical standards for data provision and documentation will be developed along established metadata standards for qualitative and unstructured data (TA.3-M.2 Qualitative RDM, TA.3-M.3 Textual Data, and TA.3-M.4 Open Response Coding). Third, an open data format with integrated (standardised) multilingual metadata will be developed as a technical tool that allows data producers to provide detailed metadata as an integral part of data production independently of proprietary software (TA.3-M.5 Open Dissemination Format).

With these Measures, TA 3 responds to the challenges posed by the growing number of data types and sources (cf. section 2.1 objective 2), for example, by linking data from related research communities that use different methods. These Measures coordinate data and metadata generation and standardisation to achieve interoperability among the different data types and academic disciplines represented by KonsortSWD (objective 4). All Measures aim to increase support for making data, methods, and research results openly accessible for everybody (objective 6). As many activities take place in and relate to international contexts, TA 3 also furthers KonsortSWD's international profile (objective 7).

Membership and governance: TA 3 brings together research-based data producers from both quantitative and qualitative social research. The work in the individual Measures is complemented by regular joint workshops in which work results are exchanged and development needs identified. The co-spokesperson in charge of TA 3 coordinates the collaboration and moderates the dialogue with other TAs. The integration of the user perspective takes place at the level of the individual Measures.

Outreach: TA 3 will regularly report to the Committee for Data Access (CDA, TA 2 Data Access) to make sure that services correspond to user needs in the RDCs. TA 3 will develop solutions that allow research data to be linked across disciplines and NFDI consortia, while improving data traceability and documentation. This makes TA 3 a central place for networking with other NFDI consortia.

Resources: TA 3 Measures require coordination with very different stakeholders. TA.3-M.1 (Harmonised Variables) will be coordinated with large survey programmes. Since TA.3-M.2 (Qualitative RDM) and TA.3-M.4 (Open Response Coding) both work on the subject of

unstructured data, close collaboration at the immediate working level is necessary. TA.3-M.5 (Open Dissemination Format), on the other hand, needs to coordinate with the Measures in TA 2 Data Access. This requires administrative support, as does monitoring the activities in the Measures and the outreach to other NFDI consortia. 1 pre-doc is required continually. This person needs additional funding for national business trips and meetings.

Measure TA.3-M.1 Facilitating the combination of research data through standardised and harmonised variables

Status quo: Researchers increasingly rely on data from different surveys. Among other things, this allows drawing more robust conclusions, extending time series, increasing the resolution for georeferencing, examining smaller subpopulations, or linking surveys that cover different populations for international comparisons. However, combining different surveys is challenging due to the use of different measurement instruments for identical concepts. This is most severe for socio-demographic variables, which are widely used as control (or 'background') variables in multivariate models across the disciplines covered by KonsortSWD (and are of substantive interest in social stratification research), but also beyond, for example, in epidemiological research. It is also problematic for other commonly used, non-demographic concepts such as life satisfaction, generalised trust, political interest, or subjective health. The problem of incoherent measurement instruments leads to extra work, information loss, biases, or even spurious findings. We see two avenues for improving the situation: first, defining the standard measurement instruments for socio-demographic variables to be used in new data collections (ex ante harmonisation), and second, developing services to support ex post harmonisation of existing data (including, but not limited to, socio-demographics).

Regarding the first approach (ex ante harmonisation), until 2016, the Federal Statistical Office of Germany published the so-called *Demographische Standards* (German Demographic Standards), a collection of recommended standard measurement instruments for socio-demographic attributes in surveys (Hoffmeyer-Zlotnik et al. 2016). These were updated at irregular intervals in the past. Such a strategy does not appear efficient: Legal and social changes need only to affect one attribute (e.g., the legalisation of same-sex marriage or the use of the term 'housewife' as a potential respondent status), for the standards to become outdated. The future of these standards is currently unclear and KonsortSWD aims at contributing to filling this gap.

In case of existing data, only ex post harmonisation is possible. Even the most widely used concepts in the social, behavioural, educational and economic sciences are measured with

different instruments across national and international surveys. Hence, the 'units' of measurement for the same concepts are not directly comparable. To derive a common, harmonised variable, ex post harmonisation is necessary. However, methodological standards, harmonisation scripts, and educational resources for ex post harmonisation are sorely lacking.

Objectives: This Measure aims at making survey data more interoperable and re-usable, and thereby more FAIR, by standardising ex ante how data are collected in the future and by making existing data more comparable ex post. With respect to ex ante harmonisation of socio-demographic attributes, we aim to build upon prior work in the German 'Demographic Standards' and a data review (Milestone 1) to provide surveys with up-to-date and quality-tested (Milestone 4) best-practice measurement guidelines for selected socio-demographic attributes (Milestone 7), for example, those for which the 2016 version is already outdated, or which surveys and data users deem in need of revision. By means of roundtable discussions (Milestones 2 and 3), we will closely coordinate with survey programmes covered by KonsortSWD, RDCs, and data users. We will also invite official statistics agencies and the German teams from international surveys to make sure we meet their needs. We aim for a modular structure that allows guidelines to be revised independently from one another to ensure timely revisions to capture societal change. Ensuring comparability over time and across countries is another aim of this Measure.

To make ex post harmonisation of variables easier, we aim to provide RDCs and researchers with harmonisation scripts for frequently used socio-demographic and non-socio-demographic variables (Milestones 5 and 7), allowing them to create and publish harmonised datasets. These scripts and the standard occupation classification codes produced in Measure TA.3-M.4 (Open Response Coding) both help make international standard variables available, which will greatly enrich the analytical potential of data. As a first example, LIfBi plans to release a harmonised Scientific Use File (SUF) that facilitates the use of data from the National Educational Panel Study (NEPS) for international comparative analysis (Aßmann et al. 2019) (Milestone 9). Another objective is to provide educational material on ex post harmonisation strategies (Milestone 8). To maintain the necessary rigour of our ex post harmonisation efforts, they will be accompanied by calibration and validation studies using experiments in online panels.

Timetable:

Year	Milestones
1	[M1] Creation of an overview of attributes studied in different surveys (Data review report) [M2] Roundtable to discuss and select variables for ex ante and ex post harmonisation (Roundtable report)
2	[M3] Roundtable to select measurement instruments for revision (Roundtable report)

	[M4] Cognitive pretesting of new proposals (Pretest report) [M5] Development of scripts for ex post harmonisation
3	[M6] Preparation, review and publication of new measurement guidelines [M7] Publication of scripts for ex post harmonisation [M8] Creation of open-access educational material [M9] Provision of a harmonised NEPS SUF

Risks of implementation: There is some risk that the stakeholders may be unable or unwilling to agree on common standards for ex ante harmonisation. In this case, the Measure will focus on producing resources for the ex post harmonisation of socio-demographic variables.

Distribution of responsibility and contributions of applicants: GESIS will steer the work in this Measure. Other institutions will contribute by sending relevant staff and experts to roundtables and by reviewing draft guidelines. LIfBi will implement the harmonisation scripts to produce a SUF for NEPS data. GESIS will contribute 0.4 staff at postdoctoral level, distributed across two current members of staff. LIfBi will provide staff for preparing and releasing the NEPS SUF.

Resources: For three years: 1 postdoctoral researcher for ex ante harmonisation, 1 doctoral researcher for ex post harmonisation, 2 student assistants 25% for routine support tasks, budget for travel to TA/stakeholder meetings, budget for validation studies and cognitive pretesting, budget for two roundtable meetings.

Measure TA.3-M.2 Generation of qualitative data – RDM portfolio for qualitative social research

Status quo: Qualitative research accounts for a significant part of social scientific research, encompassing a great variety of methods (including interviews, observations, and documentary research, often employed in combination) and data collected in a multitude of media and file formats (such as texts, images, and audio-visual material). The heterogeneity and complexity of these data, which often include sensitive personal data, make data preparation and archiving demanding. One specific challenge is that preparing qualitative data for archiving requires researchers and specialised Research Data Centres (RDCs) to cooperate during the entire research process, which often starts even before the primary research project begins (e.g. when applying for funding). Some RDCs have already developed promising solutions: Qualiservice, DIPF, eLabour, DZHW, FDZ-AGD, and FDZ-BO have established close connections with the diverse research community that uses qualitative data. Based on their expertise, this Measure can develop common standards, guidelines, and tools to support all phases of data sharing.

Objectives: The aim of this Measure is to coordinate, harmonise, and link Research Data Management (RDM) services and procedures for data generation among the RDCs of the VQualidat network (TA.2.-M.3) and to develop a common portfolio of professional RDM tailored to qualitative data and research, which ensures high data quality and maximises the potential for data re-use. The portfolio should support researchers to effectively and sustainably handle data throughout the research process, starting with project planning and data collection. The long-term goal is that researchers use instruments developed for data curation already during research and that the subsequent curation at the RDC can begin without delay. The portfolio will be supported by research-focused user trainings that link up users with participating RDCs. This approach targets all phases of the data lifecycle and makes data compatible internationally.

A key task of this Measure is to select which services should be integrated into a comprehensive portfolio and which should only be provided by some specialised RDCs. For this purpose, the existing RDM services and procedures for qualitative data will be examined (Milestone 1). In collaboration with the interdisciplinary community, services and procedures will be developed further and included into a user-friendly service portfolio, which will then be made available to the community.

An additional objective is to establish a culture of data sharing for qualitative data. In a continuous dialogue with researchers, a cultural transformation towards open data will be promoted. Training, professional exchange (Milestone 5 and 7), and good practice-examples together with more available data (Milestone 6) should lead to broader acceptance of re-using qualitative data within different strands of qualitative research.

The envisaged RDM portfolio for data preparation and submission (Milestone 1, 4 and 8) will support researchers with regard to

- Data quality: documentation of the various qualitative research processes; contextualisation of data; external data preparation and curation; development of domain data protocols (coordinated with TA.5-M.3 Interfaces) with common core elements adaptable to specific data types, enabling researchers to transfer data to RDCs without delays
- Legal certainty and preparation of data access: Anonymisation procedures and attribution of safety classifications; consent forms for data archiving and re-use as well as alternatives to written informed consent (e.g., for ethnographic research)
- Cost models/framework: specification of the resources research projects and participating RDCs typically require

- Further development: linking scholarly literature to data; bolstering data documentation for journal articles; provision of metadata forms to researchers (TA.2-M.3 Federating RDCs); further development of tools for anonymising text, image, and audio

The portfolio will be developed in close collaboration with researchers (drawing on use cases (Milestone 2), workshops and analyses of their needs (Milestone 3)) and accompanied by legal consultation for data providers and users. This Measure is closely tied to Measure TA.2-M.3 Federating RDCs: The VQualidat-network provides the structure for coordination and the website for communicating with researchers. Together, the Measures cover the whole data lifecycle and contribute to the promotion of data sharing and re-use of qualitative data.

Timetable:

Year	Milestones
1	[M1] Portfolio/basic catalogue and quality criteria: support for data preparation during research; contextualisation; law/ethics; adaptability assessment of existing technical equipment
2	[M2] Development of a catalogue of domain-specific modules (text-image-audio), testing in use cases [M3] Definition of need for technical support (target specifications); workshop for transfer/connection to the community
3	[M4] Implementation and testing of support equipment, e.g., templates, and tools for anonymisation and RDM [M5] Within extended network (TA.2-M.3 Federating RDCs): workshops and working groups opening up new research data/domains
4	[M6] Final evaluation of use cases; data transfer in RDCs and curation; curated studies available for dissemination [M7] Knowledge transfer through workshops
5	[M8] Finalisation of RDM portfolios for data preparation and submission: transfer into quotable and referenceable domain data protocols encompassing the whole data lifecycle; certification with the Core Trust Seal for quality assurance and knowledge transfer (Milestone in TA.2-M.3 Federating RDCs)

Risks of implementation: Community communication and participation depends at least partly on the successful establishment of the extended network and internet portal in Measure TA.2-M.3 Federating RDCs.

Distribution of responsibility and contributions of applicants: Responsibility for implementation is with the participating RDCs, see also their letters of commitment in appendix 3. Applicants provide legal consulting, hardware and software.

Resources: For five years: 1 postdoc (coordination), 0.5 postdocs information science, 0.5 student assistants for support, 1.6 doctoral researchers (distributed across 5 RDCs), travel expenses, budget for meetings, budget for workshops with scientific community.

Measure TA.3-M.3 Piloting the processing and linking of textual data

Status quo: Textual data are increasingly important in the disciplines covered by KonsortSWD. Made machine-readable corpora, a large variety of materials – such as election programmes, parliamentary debates, newspaper reports, social media communication, or company reports – have a far-reaching potential for research, either by outright algorithmic analysis ('text mining'), or combinations of qualitative and quantitative inquiry. The focus of this measure is on processing quantitatively large text corpora. Current research on the diffusion of populism, or of hate speech in social media using text analysis in combination with computational social science (CSS) approaches demonstrate the relevance of analysing large-scale textual data.

In the digital era, large amounts of textual data are process-generated as a by-product of political, administrative, legal, economic, and other social processes. While new data are digital from the outset, old datasets are increasingly made available in digitised formats, for example, by using optical character recognition techniques. At the same time, algorithmic methods applicable to textual data are developing rapidly and are made available via open-source libraries (predominantly in R and Python). However, the existing libraries (e.g., tm, quanteda, and tidytext) do not fully support processing linguistically annotated data. This is a barrier to using state-of-the-art techniques of computational linguistics. What is more, the tools mentioned process computations *in*-memory, limiting their capacity to process large datasets. Because an infrastructure for sensitive text data (e.g., data with commercial licenses, such as newspaper articles or tweets) would require parallel access that is highly memory-efficient, such infrastructure does not yet exist.

Concerning substantive research, the analysis of textual data is not yet sufficiently connected with research using other data sources in the disciplines covered by KonsortSWD. Linking textual data to other research data is necessary, but has not yet received sufficient funding. Connecting textual data with surveys, administrative data, etc. is still cumbersome at this stage and overburdens projects with extensive manual work.

These issues relate to a core methodological challenge of the digital transformation, namely how to ensure validity and reliability when working with 'big data.' The ability to draw connections and make comparisons between 'new data' (process-generated textual data) and conventional empirical material would improve the soundness of research with new data sources.

Objectives: The first focus of this Measure is to create links between textual and other statistical data. For this purpose, suitable identifiers for matching data will be defined (e.g., for

actors such as GND-ID, Wikidata-ID). Tools will be developed that implement workflows for linking. To arrive at generic solutions, this Measure can use established, internationally visible, and widely used corpora that inspire an enduring interest to be linked with other data sources (such as the corpora of European parliamentarianism GermaParl, AustroParl, and ParisParl). An additional selection criterion will be that the measure creates links to work in the Text+ consortium. Efforts to interlink textual data with other data sources will then aim at 1) other sensitive data from Research Data Centres (RDCs) (e.g., surveys, company data, or regional data as context information for speeches of members of parliament), 2) publicly accessible data (e.g., on persons, or organisations) which are important for the provision of text data in a future RDC, and 3) scientifically well-established data (e.g., from Twitter). The data to be linked will be identified early in the project (Milestone 1). Communication within the scientific community will ensure that the choices match research needs. To facilitate the data linking, a software library will be published (Milestone 2). During the project, updated corpora including increasingly complex annotation layers will be released (Milestones 4, 7, and 9). In addition, the links between textual data and other data will be made available via the open application programming interface specification to be developed in TA.5-M.3 (Interfaces) (Milestone 12). Following up on a set of R packages developed by the Measure Leader (polmineR, RcppCWB, cwbtools), results of the Measure will be published (GitHub and CRAN) as high-quality open source code (Milestones 5 and 13).

Second, this measure will focus on making textual data and linking facilities reproducible according to FAIR principles. To detail the implications, 1) findability *means*: capabilities offered by the Text+ Consortium (CLARIN Virtual Language Observatory) may be used in an evolving cooperation. In addition, protocols will be developed to integrate textual data into existing and widely used social and economic data catalogues (Milestone 3). Concerning 2) accessibility: we will develop procedures for dealing with sensitive aspects of textual data (copyrights, data protection). This measure will make proposals for (remote) access to sensitive text data via RDCs, mainly by creating the infrastructural prerequisites to run calculations on the server side - i.e. without transferring sensitive raw data (Milestone 5 and 13). Concerning 3) interoperability: developing standardisation for textual data from social and economic science contexts (based, e.g., on the Text Encoding Initiative TEI) will facilitate the data exchange with other consortia such as Text+ (and will accordingly be covered by Milestone 3). And concerning 4) re-usability: re-usability presupposes documentation; we will suggest procedures for data preparation that are 100% reproducible. Releases of data will be complemented by full documentation of the procedures how annotations have been generated (Milestones 4, 7, and 9).

A third focus is skills development. An existing set of online slides that integrate video tutorials (Using Corpora in Social Science Research) will serve as a basis for the training component of this Measure and will be updated and adjusted (part of Milestones 2, 6, and 10) and made available on the KonsortSWD website. These online resources will be complemented by a set of workshops (Milestones 8, 11, and 14). The code will be open source.

Timetable:

Year	Milestones
1	[M1] Identification of linkable RDC data (includes community feedback) [M2] Determination of standardised identifiers for data linkage and pilot implementation [M3] Initial arrangements for coordination with Text+
2	[M4] Release of corpus/corpora with first annotations for linking data (Person identifiers) [M5] Implementation of new functionality of open source tools to set up development version of central corpus server for remote access
3	[M6] Publication of workflow for linking data as part of a suite of open source tools [M7] Release of corpus/corpora with extended annotations, for example, organisations [M8] Updates of training material, joint workshop with Text+ (Report)
4	[M9] Release of corpus/corpora with extended annotations, for example, locations [M10] Implementation of services in cooperation with relevant RDCs, including development of monitoring service usage (Documentation) [M11] Training events
5	[M12] Consolidation of data services with relevant RDCs, further development of monitoring service usage (Documentation) [M13] Release of consolidated and robust software libraries (Consolidation of documentation, comprehensive test coverage, issues raised by users are addressed) [M14] Training event (including user feedback to prepare further developments)

Risks of implementation: The accessibility of some relevant data that shall be linked with the corpora may change. For example, Twitter's Terms of Service are steadily evolving and subject to change. However, many restrictions can be overcome by making linkable data available through RDCs for research purposes (not putting them in the public domain).

Distribution of responsibility and contributions of applicants: Responsibility is with the Measure Lead (Andreas Blätte). Contributions are open source libraries: polmineR, bignlp, annolite; support by data steward of the IT of the Department of Social Sciences of the University of Duisburg-Essen (10%); IT resources required for the measure.

Resources: 0.5 doctoral researcher for establishing for contact to RDCs, identification of relevant data, training; 1 doctoral researcher for annotation, service provision. Costs for 3 one-day workshops with 10 persons per year, 7 national, and 1 international trip per year, training materials per year.

Measure TA.3-M.4 Improved processes for coding textual information from open response formats and for deriving standard variables

Status quo: Open and semi-open response formats are part of almost all surveys that measure constructs for which it is impossible to present all relevant categories. The (re-)usability of such data for quantitative analyses largely depends on the – typically ex post – classification of the textual information and the subsequent derivation of standard variables. Manual coding, however, is a time-consuming, error-prone, and costly task in view of hundreds or even thousands of categories. A prominent example is occupational information, the provision of which in the form of classifications (e.g., Dokumentationsziffer DKZ, ISCO, KldB) and derived status, class, or prestige indicators (e.g., ISEI, SIOPS, EGP-scheme, CAMSIS) considerably increases the analytical potential of data collections. Large (panel) studies that capture educational and employment biographies, in particular, are faced with the challenge of coding huge amounts of text entries within a short period of time in a high-quality and consistent manner. In a few cases, data producers have specially developed technical solutions. In other cases, the processes are completely outsourced to commercial survey institutes. For smaller studies, both strategies may not be feasible due to a lack of resources and the information may be completely unavailable in the worst case. A general deficit – in contrast to the relevance of the resulting variables – concerns the insufficient documentation of applied coding and derivation processes.

Objectives: The project aims at establishing an infrastructure for the efficient coding of textual information from surveys with a special focus on the domains of occupation, branches, and sectors, education and training, as well as courses and subjects of study. To achieve this objective, a service centre will be set up to bring together know-how from relevant research contexts (e.g., SurveyCodings/SERISS), administrative bodies (e.g., Federal Employment Agency), coding practices at large-scale studies (e.g., NEPS), and other Measures in this consortium (e.g., TA.3-M.1 Harmonised Variables). In the context of this service centre, a database-supported software tool is to be developed and implemented for structuring and automating coding and derivation processes in such a way that suitable standard variables can be provided at minimum costs, while at the same time ensuring high-quality standards in terms of reliability, consistency, and documentation (in cooperation with TA.5-M.3 Interfaces). The tool will use machine learning methods based on text recognition algorithms. A graphic user interface with numerous functions will ensure flexible operation and supervision of the processes. Its interaction design will be developed in cooperation with TA.5-M.4 (UCD). For data producers and providers, this infrastructure offers new and cost-efficient opportunities for enriching and enhancing their data collections and for offering comprehensive documentation in

the spirit of good scientific practice. For data users, the benefit lies in the increased research potential of these (additionally) standard variables and in their high degree of comparability and interoperability. The code produced by this measure will be made available in an open source format.

Timetable: The Measure is scheduled for three years. It follows up on thematically related projects and builds on corresponding findings (in accordance with the 'strengthen' logic in our RDM strategy, see also chpt. 3). Workshops at regular intervals ensure a close exchange among the participating institutions, invited experts, and representatives from other Measures, accompanied by continuous software development and function optimisation during the entire project period.

The first year serves to compile a catalogue of relevant services (Milestone 1), to specify common coding and derivation conventions (Milestone 2), and to identify necessary software requirements (Milestone 3). This is followed by a first implementation of the coding tool and the central database of consolidated empirical material (Milestone 4). In the second year, the tool is to be extended by a web-based interface (Milestone 5) and supplemented by auxiliary applications (Milestone 6). At the conceptual level, both an internationalisation strategy (Milestone 7) and training material (Milestone 8) will be developed. Activities in the third year will focus on establishing an operating model for the service's continuation (Milestone 9) and the acquisition of further cooperation partners (Milestone 10). This will be accompanied by scientific contributions based on the evaluation studies conducted during the measure (Milestone 11) and by a relaunch of the tool (Milestone 12).

Timetable:

Year	Milestones
1	[M1] Definition of services relating to coding, derivation, and documentation processes including measures to ensure quality standards [M2] Review of coding guidelines and derivation procedures in the respective domains [M3] Inventory of requirements regarding the automation of relevant processes [M4] Programming and testing of a coding tool with functions for process automation and coding/coder administration as well as an interface to a central database
2	[M5] Opening the application to external access and use [M6] Provision of auxiliary utilities (e.g., Stata ado-files) [M7] Development of an internationalisation strategy [M8] Preparation of manuals and tutorials for the use of the tool
3	[M9] Establishment of a sustainable operating model for service and tool maintenance [M10] Announcement of the services to relevant stakeholders [M11] Analysis of data from (experimental) comparative studies for the evaluation of coding quality and algorithm efficiency [M12] Revision of the tool including extension by functions and interfaces

Risks of implementation: Implementation risks relate to the demand for the services offered and to the guarantee of efficiency and quality. Regarding the demand for ex post coding in the domains mentioned above, a considerable and permanent need from already completed, still ongoing, as well as from future surveys can be assumed. This applies to data collections of the participating institutions but also of other KonsortSWD members and other consortia. Current efforts for simultaneous coding during interviews do not reduce this need. Rather, such developments will benefit from the services described (e.g., the database with consolidated coding material) and will be involved in the project accordingly. Regarding the fulfilment of the efficiency and quality requirements, systematic scientific monitoring is planned, including the realisation of (experimental) comparative studies. Finally, close cooperation with the partners and with thematically similar Measures ensures that the development and implementation of the objectives are precisely tailored to the respective needs.

Distribution of responsibility and contributions of applicants: Responsibility for the realisation of the project and for the coordination of the participating institutions is with the Measure lead institution (LIfBi). The lead institution contributes an extensive inventory of already coded material and a long-standing expertise in the required methods and technologies. LIfBi supports the measure with assistance from relevant departments and provides resources for student assistants to be employed in this Measure. It supplies the necessary facilities such as a server infrastructure, workstations, and meeting rooms for the workshops.

Resources: 1 postdoctoral researcher for conceptual work and 0.8 doctoral researcher for software development for the entire project phase of three years. Funds are also needed for travel expenses.

Measure TA.3-M.5 Open, metadata enriched, non-proprietary data format for data dissemination

Status quo: As a common practise in quantitative social sciences, matrix-shaped data products with information about units of observation/cases in rows and their characteristics/variables in columns are offered in formats of statistical software packages (e.g., Stata, SPSS, R, SAS). Data users often need expensive software licences to use the data products or rely on software that translates the data to the statistical software of their choice. This is why data producers provide the same data in various formats, resulting in higher costs. In both spheres, non-trivial conversion problems arise because the information required for the target format is missing in the source format. Different data formats, which are only partially compatible, represent an obstacle for replication studies. Proprietary data formats, in particular, jeopardise the principle of

interoperability as enshrined in the FAIR principles. They are also challenging for long-term archiving. On the other side, documentation and metadata are often only available either as a PDF document or in a web-based information system. To use this additional information, researchers must leave their normal working environment (Statistical package).

Objectives: We aim to specify an open data format with integrated (standardised) multilingual metadata. Import filters for popular statistical software packages will be made available. Depending on the capability of the respective package, at least a subset of the supplied metadata is imported and is then available directly within the statistical package's user interface: in many cases there will be no need for an external documentation or there is a direct link to a metadata portal (cf. SOEPhelp, NEPStools/infoquery). Simplified, the actual data could be produced in a CSV format; the metadata provide information not only on variable labels and value labels, but also on relations between variables, question texts (for variables from surveys), URLs to a metadata portal, the PIDs of variables (cf. TA.5-M.1 PID Services), etc. After the import, which is as simple as to open a file in the package's native format, the statistical software package finally provides a data product that makes optimal use of the capabilities of the package and the provided metadata.

This open data format supports various FAIR data principles: the data products will be described directly by metadata, it will be easier accessible and interoperable, and upstream metadata will be re-used. Therefore, the metadata specification has to be compatible to formats developed within KonsortSWD (cf. TA.5-M.3 Interfaces), but in a first step to other already available standards, including CESSDA's Metadata Model or DDI Codebook.

Other communities, which are also using matrix-shaped data should be approached in order to consider their needs in respect to used software or possible metadata to include and widen the user basis for the new data format. As the package structure gets more stable over time, it will make sense to contact software suppliers (Stata, IBM SPSS, StatTransfer) and package providers (tidyverse/haven, which is an R package importing and exporting data formats used by other statistical packages, based on ReadStat, a widely used C library).

Specification and software, including the source code, will be provided as FLOSS software under license (e.g., CC, MIT, LGPL) which make the products easily usable in different contexts. All code produced by this Measure will be made available in an open source format.

Timetable:

Year	Milestones
1	[M1] Draft: expandable minimum specification for the format (data and metadata) [M2] Discussion within the community (KonsortSWD, CESSDA, SSHOC) [M3] Publish version 1 of (minimum) specification
2	[M4] Produce a data package in the open format with real data that are already published in a proprietary format (Anonymised as public use file) [M5] Develop one import filter for a statistical package (e.g., Stata, SPSS, R, SAS, Excel, LibreOffice Calc) [M6] Develop one conversion filter, which transforms metadata from one established format (e.g., DDI Codebook, CESSDA's metadata model, or statistical software) to the needed metadata [M7] Discussion within the community (KonsortSWD, CESSDA, SSHOC) [M8] Publish version 2 of (minimum) specification, update data package, and import filter
3	[M9] Develop import filters for two additional statistical packages (e.g., Stata, SPSS, R, SAS, Excel, LibreOffice Calc) [M10] Reach out to communities beyond disciplinary boundaries (other NFDI consortia)
4	[M11] Publish version 3 of (mature) specification [M12] Develop import filters for two additional statistical packages used beyond disciplinary boundaries [M13] Develop an additional conversion filter which transforms metadata from another established source to the specified format [M14] Update existing material to new software versions and specification versions
5	[M15] Discussion with the community [M16] Develop a governance for the continuous development of the format

Risks of implementation: Although the standards of the DDI family are already available, the amount of available metadata in these formats is limited. Even if the quantity increases, incompatibilities must be expected. Ongoing exchange with the communities will reduce these risks and help to identify problems at an early stage. Developing a minimal specification early and adapting it continuously to users' needs will help build trust in delivering a working system.

Distribution of responsibility and contributions of applicants: Responsibility is with the Measure Lead. Applicants will provide hardware and software for development and testing.

Resources: For coordination and conceptualisation: 1 postdoctoral researcher; for technical implementation of disciplinary and extra-disciplinary import filters: 2 IT specialists (other staff); for the technical implementation of extra-disciplinary import filters and a user survey: 0.5 survey specialist (other). In addition, travel expenses for meetings and conferences, catering for community workshops, potentially software/licences.

4.5 Task Area 4: Research (Data) Ethics

Several recent developments in the research data management (RDM) in KonsortSWD-related disciplines have created an immediate need for the establishment of a research (data) ethics infrastructure (the urgency was emphasised by eight academic associations in April 2018 already – the need is reconfirmed in the recent LoS by 14 scientific associations, see also

Appendix). First, the use of sensitive data requires protection of respondents. This is even more important when studying vulnerable populations, for example, the poor, people with disabilities, or people experiencing political repression. Since the introduction of the EU GDPR this requirement became salient. Second, innovative methods such as laboratory and field experiments as well as new data collection technologies create new challenges requiring ethical considerations. Increasingly, journals in these fields require ethics reviews from researchers. Also, researchers now use linked (sensitive) data or data that become sensitive through linkage. Here, the challenges of ethical data use are immediate. Third, qualitative data are more and more considered in repositories and archives. With more complex data, the challenges of data ethics – and consequently of research (data) quality – grow. The request for ethics support in relation to the use and management of research data became clear in exchanges with user communities (e.g., Hollstein and Strübing 2018).

While there is an increasing awareness of research ethics, there is a lack of a suitable infrastructure that provides reliable answers to ethical challenges for the disciplines covered by KonsortSWD. The partners behind KonsortSWD and the representatives of the respective research communities both believe that the NFDI is a suitable context to systematically address this transdisciplinary issue (see Letters of Support, appendix 4). Therefore, KonsortSWD proposes to initiate a research (data) ethics infrastructure as outlined below. The infrastructure will take advantage of the long-running experience in the fields of medicine and psychology but will be tailored to the needs of the disciplines covered by KonsortSWD. Together with other NFDI candidates, we have identified research (data) ethics as a salient cross-cutting topic for NFDI. KonsortSWD will initiate a NFDI-wide discussion by setting up a cross-consortium working group on this topic.

The KonsortSWD **Committee for Research Ethics** will offer an ‘anchor point’ for existing local and decentral ethics committees based at research institutions. It will coordinate conceptual work on research ethics issues as they connect to RDM and data protection. The broad spectrum of tasks addressed by existing, often times specialised ethics committees (e.g., on disciplines or institutions) will benefit from centralised and coordinated support, such as position papers and recommendations on founding and operating ethics committees. Also, the Committee for Research Ethics will act as a voice for its users: it will connect the efforts of KonsortSWD with research ethics-related activities in other NFDI consortia and communicate issues such as ethics, RDM, and data protection to central, cross-consortium NFDI bodies. Beyond NFDI, it will connect and exchange knowledge with similar bodies in medicine, psychology, and those addressing dual use issues of security-relevant research in Germany

and abroad. The Committee will establish the Forum Research Ethics and connect the Forum and KonsortSWD with external stakeholders.

The **Forum Research Ethics** systematically involves local ethics committees. The Forum will have two elements: a **general assembly** and a **lean organisation** and leadership team. The general assembly brings together local ethics committees plus representatives from academic associations and organisations on a voluntary basis; the leadership team organises and directs the activities. The assembly offers a platform for the exchange of information and experiences between participants, for example, through bi-annual meetings. Also, it provides support and addresses the needs of local research ethics committees.

Overall, Task Area 4 (TA 4) contributes to the following KonsortSWD objectives (cf. section 2.1): 1) by addressing ethics in the context of RDM, 2) by providing solutions to challenges generated by new types and sources of data and new methods, 5) by supporting the development of new and ethics related skills, 7) by cooperating with international partners in research (data) ethics.

The tasks of the Committee for Research Ethics and of the Forum Research Ethics differ in focus. The Committee is responsible for networking both inside KonsortSWD, for example, with measures touching on ethics challenges (TA.2-M.3 Federating RDCs and TA.3-M.2 Qualitative RDM), and beyond KonsortSWD. In contrast, the Forum focuses on the needs of local ethics committees, academic associations, and scientific institutions. It offers an arena for debate and develops supporting tools.

Tasks of the Committee for Research Ethics

- To monitor developments in research ethics at the national, European, and international level
- To network with other stakeholders and cooperate with ethics-related organisations (e.g., German Ethics Council, German Data Ethics Committee)
- To coordinate the initial activities of the Forum Research Ethics
- To establish and maintain user relations and connect with the users of other consortia
- To communicate the activities and needs of local ethics committees to the outside world (including the scientific community and third-party funding providers)

Tasks of the Forum Research Ethics

- To provide a platform for interaction between local ethics committees, academic associations and organisations, and research funding providers (e.g., German Research Foundation)

- To support local ethics committees by addressing timely topics and developing best practices
- To set the agenda of ethics discussions and develop position papers and recommendations
- To collect, systematise, develop, and provide resources, tools, and proposals towards coordination and harmonisation of the activities of local ethics committees, offer supporting materials (e.g., to address issues encountered when founding new ethics committees such as templates for bylaws, or to support public relations)
- To develop training opportunities and training materials for data producers, data users, and members of local ethics committees with regard to ethical issues developing from social, behavioural, and economic research
- To support the initiation of new ethics committees

Membership and governance structures are needs-based, efficient, and flexible. The members of the Committee for Research Ethics will be appointed by KonsortSWD's Steering Committee (StC) after consultation with and upon proposal by the German Data Forum (RatSWD). The Committee for Research Ethics will consist of a small, effective group of about 3-5 experts in the field of research ethics plus 3 user representatives commissioned on three-year appointment cycles by the academic associations supporting KonsortSWD. The assignment procedure will be organised by the community officer described in TA.1-M.2 (Community AA). The Committee for Research Ethics serves both researchers from the relevant disciplines and the local ethics committees. Membership in the Forum Research Ethics is open to ethics committees, academic associations, and other academic organisations. As the Forum Research Ethics is an open general assembly of interested parties, this structure ensures close contact with researchers' needs, who are the focus of local ethics committees' daily activity. A lean and efficient organisation with frequent informal interaction ensures feedback and constant exchange between the Committee for Research Ethics, Forum Research Ethics, and the users.

The German Data Forum (RatSWD) has been developing the ideas for the Committee for Research Ethics and Forum Research Ethics since 2016 (RatSWD 2017a, Wagner 2017) and in close contact with a number of experts in research ethics from various disciplines. Also, the German Data Forum (RatSWD) interacted with federal agencies (e.g., the Federal Anti-Discrimination Agency, *Antidiskriminierungsstelle*), local ethics committees, and academic associations at various points in this process.[32]

[32] https://www.ratswd.de/en/konsortswd/activities (5 October 2019)

Outreach: The issues of ethics and data protection must be addressed across consortia and for the entire NFDI. The Committee for Research Ethics acts as a voice for the social, behavioural, educational, and economic sciences and can become a driving force to sparking developments in other consortia and the NFDI as a whole. The disciplines behind KonsortSWD can benefit from the experience in other consortia and disciplines (e.g., from medicine and psychology within and beyond KonsortSWD) and at the same time offer support to other consortia (e.g., with respect to handling sensitive data and data protection) as mentioned in the Berlin Declaration.

As more and more scientific journals demand research ethics statements, the issue is attracting more attention in international research organisations. The Committee will act as the eyes and ears for its membership internationally, represent their interests abroad, and learn from experiences in other countries. Further, it will get involved in international bodies, facilitate international cooperation, and follow up on input from international committees and agencies.

Measure TA.4-M.1 Establishing a research (data) ethics infrastructure

Status quo: Currently, there exists no comprehensive, nationwide research ethics infrastructure to address the needs of researchers and to foster ethical reflection in the social, behavioural, educational, and economic sciences. Instead, numerous local ethics committees operate in a decentralised manner. They individually deal with challenges along the data lifecycle, in research data management, data protection, and research ethics.

Objectives: KonsortSWD aims to establish a research ethics infrastructure that meets the needs of research communities and supports local, decentralised ethics committees.

Timetable:

Year	Milestones
1	[M1] Committee for Research Ethics is established and participates in the StC [M2] Academic associations are invited to commission user representatives on three-year appointment cycles
2	[M3] First general assembly of local ethics committees and interested parties (e.g., academic associations) is invited and prepares the foundation of the Forum Research Ethics [M4] Governance of the Forum Research Ethics is agreed on and established [M5] Forum Research Ethics determines its strategic agenda, prioritises a work plan, and sets up working groups
3	[M6] Working groups of Forum Research Ethics are operational, working on position papers, and recommendations [M7] Committee reaches out to other consortia, the NFDI and internationally to identify common interests [M8] Regular assembly meetings for networking, mutual support
4-5	[M9] Working groups complete position papers which are adopted by the Forum Research Ethics

	[M10] Committee for Research Ethics establishes channels of interaction and communication with different partners within the NFDI as well as with other national and international bodies focusing on research ethics

The infrastructure will be evaluated by the end of the funding period. If the organisation is sustainable and funding is available, the research ethics-related activities of KonsortSWD may be separated from the consortium's agenda in future funding periods and operate independently. It may be possible to establish institutional cooperation with ethics infrastructures in other areas or to run an ethics infrastructure at the NFDI level.

Risks of implementation: The planned activities rely on the voluntary activity of individuals in different research ethics-related areas. Such collaboration seems likely based on past interactions of the German Data Forum (RatSWD). However, if the call to action fails, the research ethics infrastructure for the disciplines covered by KonsortSWD cannot be established.

Distribution of responsibility and contribution of applicants: The co-spokesperson in charge of TA 4 works together with and is supported by the current chair of the German Data Forum (RatSWD), Regina T. Riphahn, who has been in charge of developing the research ethics activities of the German Data Forum (RatSWD) so far. Both will contribute 10 % of their working time to support TA 4. The Task Area in general and this measure in particular are supported by 14 academic associations as well as several Research ethics Committees and further actors engaged in research (data) ethics (see appendix 4). Their engagement will guarantee a user oriented establishment of the research (data) infrastructure.

Resources: 1 doctoral researcher in year 1 and 1.5 in years 2-5 to provide scientific support plus 0.75 administrative assistants for years 1-5 to run the administration. Additionally, travel funds for trips to conferences and workshops. Committee and Forum require budgets for meetings of the Committee, the Forum, its leadership team, and working groups as well as for documentation and dissemination of information on activities (professional support and webpage updates) and outreach, i.e., travel expenses for national and international events.

4.6 Task Area 5: Technical Solutions

Status Quo: The research data infrastructure (RDI) landscape is characterised by a high degree of technological heterogeneity. Assuring the connectivity of services in the long run is therefore a challenge. An open application programming interface (OAI) for standardised metadata exchange across different data providers (e.g., RDCs) with different data types (survey data, video, audio and log data, transcripts from qualitative social research, data from digital social networks, text corpora, etc. (see Figure 1)) and across the entire research data

lifecycle would substantially increase connectivity and sustainability. More importantly, open protocols allow for different speeds of adoption, which are critical in a federated data infrastructure.

Challenges: The data from KonsortSWD is typically listed in dedicated data catalogues. New users or users from disciplines outside KonsortSWD, however, will search for data on generic, not discipline-specific channels like DataCite, Google Dataset Search that harvest the metadata of various providers. Because NFDI widens the spectrum of disciplines served by the research infrastructure these users become more relevant. Moreover research data are much harder to find on the web than publications (Halevi et al. 2017). Improving the visibility and findability of KonsortSWD data on the web is therefore of central importance. Assigning Persistent Identifiers (PIDs) to individual survey attributes would make it easy to reconstruct, in a machine-readable way, which element of a dataset was used in a study. Finally, users have only rarely been involved in the development of RDIs or data services from the onset. To increase user orientation and user friendliness in the sense of FAIR, KonsortSWD will systematically involve its users in the development of services.

Objectives: This Task Area (TA) bundles the information technology expertise of the consortium partners to provide cross-cutting services that advance the sustainability and FAIR compliance of the envisioned RDI. It thereby contributes to KonsortSWD's objectives (cf. 2.1): objective 1 (provide tools for sharing data FAIRly), objective 4 (ensure and enhance the FAIRness of (sensitive) data), objective 6 (coordinate data and metadata standardisation to achieve interoperability across different data types and academic disciplines represented by KonsortSWD), and objective 7 (become an international and national voice and point of contact for RDM). All digital services will be developed in accordance with user requirements, which will be ascertained together with TA 1 Community Participation, and respecting data sensitivity and the needs of RDC users as represented in the Committee for Data Access (CDA, see TA 2 Data Access).

This Task Area (TA) will generate a reference architecture for a research data infrastructure that allows data providers to make their data compliant to different maturity levels of FAIR. While FAIR principles are accepted guidelines for data handling, they do not provide a blueprint for infrastructure development because there are various ways to implement them. TA 5 therefore strives for technology leadership to implement the FAIR principles in KonsortSWD's domain. The commitment to develop pioneering FAIR tools is already evident in the founding and coordinating of the GO FAIR Implementation Networks GO Inter and EcoSoc-IN by KonsortSWD partners.

Together with all partners of the consortium TA 5 focuses on strengthening the FAIRness of KonsortSWD data, providing a model for other data providers to advance the FAIRification of their own data and integrating future RDCs more easily. This will be done through the following Measures:

Data should be represented at the level of the individual data objects (e.g., survey attributes) in such a way that they are interoperable and reusable with other data objects according to the FAIR principles. TA.5-M.1 PID Services establishes a PID infrastructure that includes PIDs at that level (variables) and the associated concepts as well as the vocabulary used to describe them. The Measure will be coordinated with TA 2 Data Access.

TA.5-M.2 (Data Findability) improves searchability and findability of KonsortSWD data on general data search engines. It will make (meta)data available using standardised mark-up languages (such as schema.org), general search engines (Google Dataset Search in particular), standard interfaces (such as Open Archives Initiative Protocol for Metadata Harvesting, or OAI-PMH), and by feeding them into the metadata index of the EOSC. Furthermore, a rich and, where possible, lossless high-quality metadata schema will be implemented in standard formats to support both community-specific as well as cross-disciplinary data searches. The Measure will strengthen FAIRness of data by increasing data findability. It will be realised in close cooperation with TA 2 Data Access.

TA.5-M.3 (Interfaces) will develop and establish an OAI Specification for standardised data exchange across different stakeholders and phases of the research data lifecycle. This will stop the fragmentation of services and provide a common protocol for accessing data, fostering the homogeneity and sustainability of technical implementations. It will also improve the connectivity of data services within KonsortSWD as well as with the European Open Science Cloud (EOSC) and other international initiatives (e.g., GO FAIR, RDA, DDI, W3C). This Measure will closely cooperate with TA 3 Data Production.

Finally, to ensure the best possible user experience, TA.5-M.4 (UCD) will develop all digital services applying user-centred design (UCD), allowing users to participate in their development from the beginning. The Measure will be coordinated closely with TA 1 Community Participation.

All code produced by measures in TA 5 will be made available as open source in a git repository accessible through the KonsortSWD website.

Membership: Senior scientists in the field of information technologies (computer science, information science) of the participating consortium members will participate in TA 5.

User basis and integration: TA 5 provides services for both data users and providers. According to the principles of user-centred design, users will be involved in the development of services. User demands for services collected by TA 1 Community Participation will be communicated directly or through the Steering Committee (StC) (cf. section 2.5) and factored into planning. Should more domain-specific demands arise, they could be taken up by way of a consolidating measure (TA.6-M.3 Network Refinement).

Outreach: TA 5 will regularly report to the Committee for Data Access (CDA, TA 2 Data Access) to ensure that services meet RDC user needs. Transfer of knowledge, experience, and technology is planned with other NFDI consortia, particularly potential bridge consortia (e.g., Bridge4NFDI, 2LINKNFDI). Seamless integration into the European Open Science Cloud (EOSC) and the Consortium of European Social Science Data Archives (CESSDA) infrastructures will take place through the partners' involvements (see section 2.4). Participation in relevant standardisation initiatives (such as RDA, W3C) is also planned.

Distribution of responsibility and contributions of applicants: Responsibility for coordination is with ZBW as the TA's lead institution, GESIS will contribute 0.1 postdoctoral researcher to assist with coordination of TA 5.

Resources: 0.3 postdoctoral researcher for coordination of TA 5.

Measure TA.5-M.1 Enhancing PID services as a base for a FAIR data infrastructure

Status quo: A persistent identifier (PID) uniquely identifies a data entity regardless of its storage location. Currently, different PIDs exist (e.g., DOI, ORCID, VIAF), but no single standard has evolved. Of the 34 RDCs accredited by the German Data Forum (RatSWD), 32 register their data with the da|ra data registration agency, developed and operated by two participants of KonsortSWD (GESIS and ZBW). da|ra is a member of DataCite and uses digital object identifiers (DOI). RDCs introduced the PID to improve the findability of data and to facilitate their citation (Reinhold et al. 2016). Correct citation, in turn, facilitates replication and the documentation of the impact of research data. A strength of da|ra is that it uses a metadata schema for registration that was developed especially for the social sciences and economics. There already are data mappings for the most widespread metadata schemas (e.g., Dublin Core, Schema.org) that help other search engines to harvest da|ra metadata. Registration of RDC data with da|ra enabled the creation of a comprehensive search index, using data discovery technology developed by GESIS (gesisDataSearch). RDC holdings registered with da|ra now are discoverable via the search indices of DataCite, Data Citation Index,

gesisDataSearch, the cross-domain search engine developed in the GeRDI project, B2Find of the European Open Science Cloud (EOSC), and Google Dataset Search. Thus, the registration of data and the use of DOIs were significant first steps to improve the findability of data, especially from official statistics agencies and social insurance providers.

To date, da|ra assigns DOIs to data files only. However, each of these files usually contains several hundred attributes (e.g., variables). More granular referencing would have several benefits: It would allow referencing which parts of a dataset were used in a particular publication and support reconstructing a research finding. It would improve documentation since attributes are currently only cited in the textual description, if at all. Furthermore, for many types of research, attributes need to be tracked over versions of a dataset to create time lines. In addition, PIDs for attributes can be used within scripts (do-files, R scripts, etc.) to always link to the original dataset. The next step towards a FAIR data infrastructure therefore is to extend PID services for advanced referencing and cross-referencing of (meta)data.

Objectives: The objective of this Measure is to widen the functionality of da|ra and the use of PIDs to the level of attributes. PIDs are the backbone of a FAIR data infrastructure because they allow for referencing and cross-referencing data and metadata objects. *Widening* da|ra, which currently focusses on the social sciences and economics, means to open up registration of research data for all participating disciplines – in particular, neighbouring disciplines of the social sciences (e.g., the humanities). To this end, metadata schemas for registration must be developed in cooperation with the new user communities. Expanding the use of PIDs implies using them for attributes within datasets and special documents (inline objects). In FAIR terms, identifiers for inline objects are an important step towards *machine-actionable* data, improving the findability and re-usability of data at the level of attributes.

By assigning PIDs to the attributes, individual elements of the data files can be referenced and retrieved with the required metadata. This is particularly important when users access data through an application programming interface (API) or Jupyter Notebooks, or when researchers re-use questions from existing surveys for their own research (which is frequently the case). The starting point for a FAIR data infrastructure is the concept of the Digital Object Interface Protocol (DOIP) (Wittenburg et al. 2019; see Figure 7, section 3.2). This will further *strengthen* the findability of existing data at the sub-dataset (variable) level. The specific tasks of the Measure will be 1) to develop technical solutions for fine-grained referencing and identification at the attribute level (e.g., fragment identifiers), suitable for scaling and for handling sensitive data. These solutions need to be integrated into the existing services (in close cooperation with TA 1 Community Participation). They will not only enable citation, but also the permanent

identification of data objects through APIs for data exchange via new interfaces (e.g., DOIP). 2) To meet an increasing demand for interoperability, data mappings (i.e. translation between machine readable formats) need to use the Data Documentation Initiative (DDI) standard. It structures metadata at the level of attributes in a machine-readable way. In order to simplify and standardise this process, da|ra will be upgraded to process the relevant metadata standards of the communities.

This procedure serves as a blueprint for the inclusion of other subject-specific metadata standards (e.g., from experimental research, text-based research and big data research). The user interface will be developed in consultation with TA.5-M.4 (User-centred design). All code produced by this Measure will be made available as open-source format.

Timetable:

Year	Milestones
1	[M1] Integration of additional data providers from the NFDI into da|ra (GESIS contribution) [M2a] Modelling of an extension of da|ra for PIDs on the level of attributes/questions, including adaptation/extension of the metadata schema (Report)
2	[M2b] Implementation of modelled PIDs for attributes/variables (Repository enhancements)
3-5	[M3] Integration of an efficient and flexible component (DDI-FlatDB) to handle different metadata schemas represented in XML and to enable domain switches (Repository enhancements) [M4] Analysis and integration of metadata schemas from other scientific domains, adaptation of data registration for selected additional scientific domains and extension of da|ra by registration services for metadata objects (Report)

Risks of implementation: Change of the DOI provider structure of da|ra (DataCite) may increase development costs.

Distribution of responsibility and contributions of applicants: GESIS is responsible for the coordination of the Measure, it will contribute 0.5 postdoctoral researchers for coordination by team da|ra.

Resources: 1 doctoral researcher for requirement specification, concept development, lead development and implementation; 1 IT developer (other) for supporting development; cost for travel to workshops and conferences.

Measure TA.5-M.2 Enhancing data findability

Status quo: As the number of data repositories continues to grow, it is increasingly important to make their data searchable and findable on the web not only on special portals or repository websites. In a fragmented data infrastructure landscape, the risk is high that users cannot easily

locate data catalogues from individual data providers. In the envisioned federated data infrastructure of both KonsortSWD and NFDI, this issue is even more pressing. Instead of searching in a specific (e.g., KonsortSWD's) data catalogue, users likely prefer using general search engines (e.g., Google, gwant.com) or data portals that harvest metadata from several providers (e.g., DataCitationIndex, DataCite). Unless data are provided in common web formats and repositories implement search engine optimisation (SEO) instruments well, not all users will find the data they are interested in. Moreover, research data are much harder to find on the web than literature references. This is due to their specific character (more than just text) and the great heterogeneity of both metadata formats and the semantic indexing of data (Halevi et al. 2017).

Objectives: The goal of the Measure is to increase the findability and re-use of KonsortSWD metadata and non-sensitive data, including PIDs on the attribute level (provided by TA.5-M.1 PID Services). The Measure focuses on consistently providing metadata in common web formats (such as schema.org) and systematically implementing appropriate SEO instruments. This is to be based on a high-quality and rich metadata standard, which exists, e.g., for da|ra, and is to be refined in TA 2 Data Access. Only through the consistent usage of a uniform metadata standard during the documentation of research data can KonsortSWD effectively and efficiently help researchers find and re-use data. This Measure will make the integration of RDCs easier in the future and strengthen the "F" of the FAIR principles. The following tasks are planned:

- Conducting user studies and using test cases to document and analyse the findability of data on the web; monitoring the impact of implemented SEO instruments (in close cooperation with TA 1 Community Participation and TA 2 Data Access, esp. CDA Council)
- Developing and implementing a sustainable visibility strategy to promote the searchability and findability of KonsortSWD data on the web, such as applying schema.org, annotating data with common web vocabulary, and creating dedicated landing pages, integrated, for example, in RDC websites (with TA 2 Data Access).
- Developing indicators to evaluate the impact of visibility instruments
- Providing input from the perspective of data visibility to metadata standardisation and enrichment (in close cooperation with TA 2 Data Access, especially TA.2-M.3 Federating RDCs).

Timetable:

Year	Milestones
1	[M1] Study of the searchability and findability of KonsortSWD data on the web (Report) [M2] Development of a strategy to increase the findability of KonsortSWD data on the web (including indicators to measure their impact) based on the requirements derived from the visibility study and using common web standards such as schema.org (Report) [M3] Enrichment of KonsortSWD metadata schemas (e.g., da|ra) with metadata that increase data visibility (Report)
2	[M4] Implementation of the strategy in selected KonsortSWD repositories, in particular by applying schema.org broadly, annotating KonsortSWD webpages with SEO-relevant keywords, implementing dedicated landing pages for digital KonsortSWD objects (Repository enhancements) [M5] Implementation of monitoring instruments to measure the impact of the visibility instruments (Tool)
3	[M6] Evaluation of the impact of the implemented visibility strategies (Report) [M7] Further development of data visibility along new requirements (Report) [M8] Improvement of the visibility strategy (Repository enhancements)

Risks of implementation: Data providers might not implement the visibility strategy consistently.

Distribution of responsibility and contributions of applicants: GESIS will be responsible for coordinating the Measure. ZBW will contribute to the conceptualisation and specification of the visibility strategy and impact measurement. DIPF will contribute 0.1 postdoctoral researchers for concept and specification, GESIS will contribute 0.2 postdoctoral researchers for coordination.

Resources: For years 1-3: 0.75 doctoral researchers for requirement specification, development of concepts, lead development and support of implementation; 0.5 doctoral researchers for concept and specification; costs for travel to workshops and conferences.

Measure TA.5-M.3 Interfaces for data exchange processes

Status quo: Research data are handed from one stakeholder to another throughout their lifecycle, requiring validation each time. All participants, including research institutes, survey field institutes, and RDCs, use specialised software to manage information within their own workflow. This process is inefficient and leads to heterogeneous (meta)data. RDCs address this challenge by providing a repository-specific validation process. This, however, requires specification of information (Latif et al. 2018, 2014). Furthermore, harvesting (meta)data from the RDCs is complicated due to a lack of standardisation. The Data Documentation Initiative (DDI) standard for metadata is a solid foundation for establishing a common specification of metadata and can also be expanded by sub-specifications. However, there is no common interface that addresses researchers, and research data infrastructures (RDIs). This also makes it difficult to trace the provenance of data and to access repositories directly.

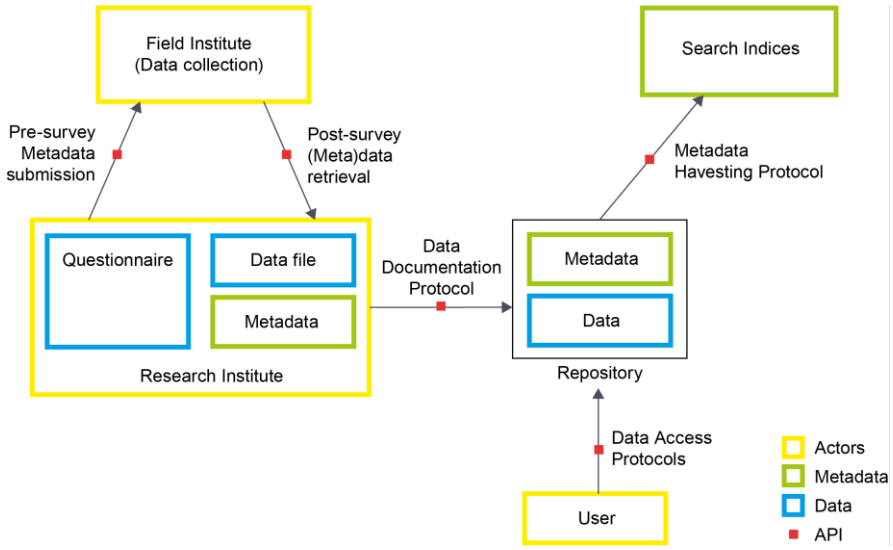

Figure 9: Interfaces of data generation[33]

There exist several tools at KonsortSWD partners' repositories (such as SowiDataNet and RDI Education), including data access tools such as ERDA and submission support tools such as GeRDI. In addition, there are other institutions running further tools such as RADAR or Zenodo, which are multi-disciplinary. All these will be used as a starting point for the development of the interfaces.

Objectives: The main objective is to build an open application programming interface (OAI) specification (Milestone 1) for the entire data exchange process. It will take into account established W3C standards (such as RDF, JSON, and JSON-LD), common web vocabularies, and novel methods developed or promoted by international FAIR initiatives such as GO-Inter[34] (which explicitly addresses interoperability), FAIRsFAIR, [35] or FAIRsharing. [36] The OAI specification allows each participant in KonsortSWD to extend and individualise the data lifecycle while maintaining interoperability and the necessary metadata. Research Data Management (RDM) is d*eepened* by supporting RDCs to exchange data more easily. It is also *strengthened* by the possibility to extend interfaces to meet special requirements, such as more detailed metadata. There are five access points that are being addressed by the Measure (see Figure 9). The specification builds on DDI and will focus both on the interaction among stakeholders and the submission of (meta)data. Software requirements will be defined in close

[33] The OpenAPI specification includes the communication between research and field institutes as well as RDCs. Starting at the initial exchange between research and field institutes, it defines a standardised content exchange between RDCs.

[34] https://www.go-fair.org/implementation-networks/overview/go-inter/ (26 September 2019)

[35] https://www.fairsfair.eu/ (26 September 2019)

[36] https://fairsharing.org/ (26 September 2019)

cooperation with TA 3 Data Production, especially Measures TA.3-M.2 (Qualitative RDM) and TA.3-M.5 (Open dissemination format). The application programming interface (API) will foster automated verification and support access rights management, metadata enrichment as well as the enforcement of quality standards as defined, for example, by KonsortSWD. A prototype for the implementation of selected scenarios will be created (Milestone 3), tested, and evaluated (Milestone 4). The proposed API especially addresses access to raw data by users. For example, researchers may connect Jupyter Notebooks to use data directly and share the results within the community (Milestone 5). It is therefore planned to closely cooperate with TA 3 Data Production, particularly the Committee for Data Access (CDA). The API will homogenise technical applications and thereby foster the development towards the FAIR principles, in particular with regards to the interoperability of data. As the participating partners strive for model implementations, direct validation and incremental development of the standard is possible (Milestone 6). Outreach to external partners, coordinated with TA 1 Community Participation, provides feedback that makes services more durable (Milestone 7). Source code will be published in an open-source format that partners will actively maintain (Milestone 8). This increases the overall outreach and allows external contributions as well as potential implementation within other disciplines.

Timetable:

Year	Milestones
1	[M1] Selection of prototypical data exchange scenarios [M2] Identification of metadata for access points (Workshop, common metadata schema)
2	[M3] Integration of existing technologies for selected access points (Demo workflow) [M4] Iterative evaluation of strengths and weaknesses of the created workflow
3	[M3] Integration of additional access points [M4] Evaluation of access points regarding KonsortSWD requirements (Workshop) [M6] Validation of the interfaces based on feedback and usage logging
4	[M3] Implementation of specialised services and searches [M4] Evaluation of OAI Specification with respect to added services [M5] Experimental extension with different direct data access for analysis
5	[M6] Evaluation of DDI metadata specification [M7] Dissemination of results to NFDI, DDI, and the FAIR community (Workshop) [M8] Support for software products and extension of OAI specification

Risks of implementation

Implementation depends on defined quality standards, heterogeneous DDI specifications with different workflow requirements can limit the scope of API functionality.

Distribution of responsibility and contributions of applicants

DIPF will coordinate the measure, GESIS will contribute to interfaces for repositories and service development; to conceptualisation, and to search infrastructures, ZBW will contribute to harvesting technology and service development; to conceptualisation and to the specification of analysing services, data access, and provenance models. DIPF will contribute 0.2 postdoctoral researchers for concept and specification, GESIS will contribute 0.1 postdoctoral researchers for the communication with research institutes and 0.05 senior researcher with expertise in the DDI standardisation process.

Resources: 2 doctoral researchers for requirement specification, development of concepts, lead development, cooperation with research and field institutes and support of implementation; 0.5 doctoral researchers for concept, specification, and communication with partners; 1.6 IT developers (other staff) for implementation of the API and support of infrastructure; costs for travel to workshops and conferences.

Measure TA.5-M.4 User-centred design

Status quo: To date, the integration of users into infrastructure development most often relies exclusively on user advisory boards. The reception of services by users (output side) – if it is measured at all – is frequently linked to page visits or number of downloads. How users evaluate services is very important for KonsortSWD because it shows whether the services improve the research process.

One way of including users in the development process is the standardised user-centred design (UCD) approach (ICS ISO 9241-210:2019, 2019). UCD is an iterative process in which designers focus on the users and their needs. The process consists of four distinct phases. First, the design team tries to understand the context in which users use a system. Second, user needs are specified based on the results of phase one. In phase three, the design team develops solutions, which in turn are evaluated by users in phase four. A variety of research and design methods are available for each phase. They involve users and ensure the creation of a product with a convincing user experience using several design methods. An example for successful implementation of a UCD process is the integrated search system at GESIS (Hienert et al. 2019), which allows users to efficiently search across different types of information.

Objectives: The main objective of this Measure is to ensure that KonsortSWD's web-based products are easy to use and meet user needs. The broad application of UCD will help achieve this. Users will be involved in the development process from the beginning. Accordingly, the

Measure will *strengthen* the research data infrastructure (RDI) in the sense of the research data management (RDM) strategy and the FAIR principles. The following tasks are planned:

- Interacting with KonsortSWD development projects in Measures TA.1-M.4 Dissemination, TA.2-M.3 Federating RDCs, TA.2-M.5 Transparent interaction, and TA.3-M.4 Open Response Coding
- Identifying user groups for each planned digital service (in cooperation with TA 1 Community Participation)
- Involving end-users from each user group in the development process by applying UCD methods, such as contextual inquiries, questionnaires, design thinking, focus groups, workshops, and usability tests
- Monitoring usage on the base of log data.
- Developing user scenarios, personas, prototypes, and wireframes; adjusting user needs and requirements based on user feedback and analysis of log data
- Providing input for graphic and interaction design of KonsortSWD services based on user feedback as well as recent guidelines and standards, while keeping the applicability to mobile technologies in mind

Timetable:

Year	Milestones
1	[M1] Study of use context by conducting contextual inquiries, interviews, and usability studies of existing portals and digital services (Report) [M2] Specification of user needs and requirements for KonsortSWD services (Report, click prototypes, wireframes)
1-3	[M3] Continuous support of the development process of the KonsortSWD website (TA.1-M.4): application of UCD methods, development of user scenarios, personas, click prototypes/wireframes, recommendations for graphic and interaction design, and usability tests (Report, click prototypes, wireframes) [M4] Continuous support of the development process of the tool for coding textual information (TA.3-M.4 Open response coding): application of UCD methods; development of user scenarios, personas, recommendations for graphic and interaction design, and usability tests (Report, click prototypes, wireframes)
1-5	[M5] Continuous support of the development process of VQualidat (TA.2-M.3 Federating RDCs): application of UCD methods, development of user scenarios, personas, recommendations for graphic and interaction design, and usability tests (Report, click prototypes, wireframes)
1-5	[M6] Continuous support of the development process of the online discussion platform (TA.2-M.5 Transparent interaction): application of UCD methods, development of user scenarios, personas, recommendations for graphic and interaction design, and usability tests (Report, click prototypes, wireframes)

Risks of implementation: The recruitment of participants may encounter unpredictable challenges, which may lead to delays.

Distribution of responsibility and contributions of applicants: GESIS will be responsible for coordinating the Measure and reaching the milestones. GESIS contributes 0.2 senior researchers for coordination and support

Resources: 1 doctoral researcher for managing UCD activities; 0.75 doctoral researchers for User experience design; 0.5 other for frontend development; 0.25 student assistants for user studies; 10 national trips for user studies per year, costs for incentives, and costs for 2 trainings each in year 1 and 2.

4.7 Task Area 6: Secretariat

KonsortSWD's secretariat will manage the consortium's finances, ensure compliance with funding conditions, and provide reporting on the consortium's services. As the consortium's administrative Task Area, TA 6 will ensure that KonsortSWD is equipped for expanding its activities in all branches of the consortium and related consortia (cf. 2.1 KonsortSWD objective 8). In particular, by way of TA.6-M.3 (Network Refinement), we ensure a flexible and efficient consortium structure and the ability to adjust the service portfolio and make room for contributions of new partners (objective 9).

Measure TA.6-M.1 Accounting, monitoring, and reporting

Objectives: Tasks include managing the consortium's finances, ensuring compliance with funding conditions, and providing reporting on the consortium's services. The Measure will process short, standardised bi-annual reports on Measure and TA progress as well as the status on agreed performance indicators (cf. chapter 2.5). Based on this information, standardised reports will be prepared for the Steering Committee, the Advisory Board, and for the German Research Foundation (DFG).

Measure TA.6-M.2 Contract management

Objectives: This Measure will manage contracts between all institutions with (co-)spokespersons, negotiate new contracts (e.g., with new partners for TA.6-M.3 Network Refinement), or terminate contracts with partners whose services are no longer part of KonsortSWD's portfolio.

Measure TA.6-M.3 Network refinement

Status quo: KonsortSWD strives to serve its disciplines with services tailored to their research needs. We anticipate new members of our communities to become aware of KonsortSWD as a new provider of infrastructure. Also, we expect ideas for cross-consortium projects to arise once NFDI is under way.

Objectives: KonsortSWD wants to actively address arising challenges regarding RDM and encourage new partners to contribute services to its portfolio. Network Refinement provides a budget for such additions to the portfolio. Decisions about Network Refinement Measures will be based on **open calls for proposals** among the communities represented by KonsortSWD. Successful proposals will be competitively selected by the Advisory Board to assure they reflect user needs. Network Refinement Measures will start in year 4 of the funding period. We anticipate applications for Network Refinement Measures to be competitive and foresee contributions from at least the following areas:

- Making geographic data easier to use for KonsortSWD communities and developing anonymisation services (in collaboration with NFDI4Earth)
- Making sensitive cohort data and survey data re-usable and interoperable (together with NFDI4HEALTH)
- Adding unstructured (sensor, transaction, tracking, or text) data to KonsortSWD's infrastructure, for example, with NFDI Web
- Supporting cross-consortium projects within NFDI
- Supporting beacon projects for making existing data FAIR
- Supporting the federation of repository infrastructures; interest from the business administration and economics communities has been signalled for a Measure analogous to TA.2-M.2 Access Points
- Fostering cooperation with specialised information services (SIS, *Fachinformationsdienste*) and optimising information exchange (in close cooperation with TA.1-M.3 Community SIS)

As funding guidelines only allow for NFDI funds to be awarded to co-applicant or participating institutions, we count on the German Research Foundation (DFG) to establish a process for approving new members to join the consortia.

Measure TA.6-M.4 Consortial governance (Konsort Assembly, Steering Committee)

Objectives: This Measure organises the annual Konsort Assembly as well as the meetings of the Steering Committee (StC) (cf. section 2.5) to enable regular exchange of information between all partners within KonsortSWD and to discuss further development of KonsortSWD.

(Co-) Applicant contributions

The attached Letters of Commitment show the profound support of the network of Research Data Centres (RDCs) as participants for KonsortSWD. The RDCs participating in KonsortSWD securely store data, metadata, and their documentation on a long-term basis and make them findable, accessible, interoperable, and re-useable. RDCs run quality assurance measures like checks for completeness, consistency, and plausibility of (meta)data. They correct and generate data, for example by coding missing values, correcting implausible values, and harmonising longitudinal survey data (RatSWD 2019a, p. 16). For that purpose, RDCs build on staff of 285 full time-equivalents (FTEs), who provide more than 71,000 datasets to about 46,000 users (as of 2018, cf. Figure 10). RDCs' staff mostly consists of researchers, who have generated more than 500 academic publications in 2018. Hence, RDC staff supports users based on own research experience. RDCs offer qualification measures such as trainings and workshops (RatSWD 2019a, p. 19) and are essential for implementing new services, especially those developed in TA 2 Data Access. Most RDCs have permanent funding, securing the sustainability of KonsortSWD. This powerful network as a whole (represented by key figures in Figure 10) is the crucial contribution to KonsortSWD – the (co-)applicants of KonsortSWD manage a significant share of the network.

Figure 10: Key figures on RDCs as of 2018

Most (co-)applicant institutions contribute their own RDC services to KonsortSWD. Beyond this, the applicant institution, GESIS, will contribute its archiving and search infrastructures (SowiDataNet | datorium and gesisDataSearch) to NFDI. It will commit computational resources for development projects in TA 5, time and advisory activity from its da|ra staff (TA.5-M.1 PID Services) and from staff working for the Data Documentation Initiative (DDI) (TA.5-M3 Interfaces). Furthermore, working time of experts on harmonisation (TA.3-M.1 Harmonised Variables) and of experts on RDM training (TA.2-M.4 Skills), time for coordination and concept

specification (TA.5-M.2 Visibility; TA.5-M.3 Interfaces, TA.5-M.4 UCD) as well as 10% of the working time of Christof Wolf, GESIS' president and KonsortSWD's spokesperson will be contributed.

The co-applicants will contribute office space for KonsortSWD's staff, venues for meetings, and 10% of their co-spokespersons' working time. On top of this, they will make the following contributions:

University of Bremen will provide hardware equipment, software packages, and working time for legal consulting (TA.2-M.4 Federating RDCs, TA.2-M.3 Federating RDCs, and TA.3-M.2 Qualitative RDM). The University of Duisburg-Essen contributes a Data Steward, IT resources, and advisory activity for the development of R packages (TA.3-M.3 Textual Data). DIPF commits working time for concept specification (TA.5-M.2 Visibility, TA.5-M.3 Interfaces), contributes developed software (ERDA, RDC Education), server infrastructure as well as their experience in coordinating a federated research data archiving infrastructure (TA.2-M.3 Federating RDCs). The DZHW staff supports workshop organisation (TA.2-M.1 RDC Support) and provide hardware equipment and software packages (TA.2-M.3 Federating RDCs and TA.3-M.2 Qualitative RDM). LIfBi contributes a server infrastructure, workstations, resources for student assistants and for technologies (TA.2-M.5 Transparent Interaction, TA.3-M.4 Open Response Coding), and working time of NEPS experts (TA.3-M.1 Harmonised Variables). SOEP at DIW Berlin will provide hardware and software (TA.2-M.2 Access Points) and staff for organising the apprenticeship exchange programme (TA.2-M.4 Skills). WZB will support TA 1 Community Participation as well as TA 4 Research Ethics by opening up its broad research (ethics) networks. Elected members of the German Data Forum (RatSWD) – particularly its chair – invest their time free of charge to contribute to TA.1-M.1 (RatSWD) and as an Advisory Board to KonsortSWD. Regina Riphahn – the then former chair of the German Data Forum (RatSWD) – will invest 10% of her working time to support the establishment of TA 4 Research Ethics infrastructure. ZBW commits time and the expertise of FAIR experts for concept specification (TA.5-M.2 Visibility). ZPID supplies time, training expertise, and computing resources (TA.2-M.4 Skills). SOFI (eLabour) and FDZ-BO at DIW Berlin contribute expertise from their networks of qualitative research data holders (TA.2-M.3 Federating RDCs and TA.3-M.2 Qualitative RDM).

Appendix

1 Bibliography and list of references

Adena, M. and S. Huck (2017). Matching Donations Without Crowding Out? Some Theoretical Considerations, a Field, and a Lab Experiment. Journal of Public Economics, 148, 32-42.

Adena, M. and S. Huck (2019). Giving Once, Giving Twice: A Two-Period Field Experiment on Intertemporal Crowding in Charitable Giving. Journal of Public Economics, 172, 127-134.

Aßmann, C.; H. W. Steinhauer, A. Würbach, S. Zinn, A. Hammon, H. Kiesl, G. Rohwer, S. Rässler and H.-P. Blossfeld (2019). Sampling Designs of the National Educational Panel Study: Setup and Panel Development. In: H.-P. Blossfeld and H.-G. Roßbach (eds.). Education as a Lifelong Process (2nd edition). The German National Educational Panel Study (NEPS) (Edition ZfE, vol. 3), 35-55.

Aßmann, C.; J. Boysen-Hogrefe and M. Pape (2016). Bayesian Analysis of Static and Dynamic Factor Models: An Ex-Post Approach towards the Rotation Problem. Journal of Econometrics, 192 (1), 190-206.

B2FIND. http://b2find.eudat.eu/ (11 October 2019).

Behrmann, L. and **B. Hollstein** (2012). Starthilfe oder Hemmschuh? Arbeitsmarkteinstieg und soziale Netzwerke gering qualifizierter junger Erwachsener. In: J. Mansel and K. Speck (eds). Jugend und Arbeit. Empirische Bestandsaufnahmen und Analysen. Weinheim, Juventa, 79-98.

Benoit K.; K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller and A. Matsuo (2018). quanteda: An R package for the quantitative analysis of textual data. Journal of Open Source Software, *3*(30), 774. https://quanteda.io (11 October 2019).

Best, H. and **C. Wolf** (2015, eds.). Sage Handbook of Regression Analysis and Causal Inference. London, Sage.

Blask, K. and A. Förster. (2019). Designing an information architecture for data management technologies. Introducing the DIAMANT model. Journal of Librarianship and Information Science, online first. https://doi.org/10.1177/0961000619841419 (11 October 2019).

Blätte, A. (2019). cwbtools. https://github.com/PolMine/cwbtools (11 October 2019).

Blätte, A. (2018). Interaktion und Dialog mit großen Textdaten: Korpusanalyse mit polmineR. In: K-U Schnapp, J. Behnke, C. Wagemann and A. Blätte (eds.). Computational Social Science - Die Analyse von Big Data. Baden-Baden, Nomos, 119-138.

Blätte, A. (2016). polmineR (v0.6.2). http://www.github.com/PolMine/polmineR (11 October 2019).

Blätte, A. and A. Blessing (2018). The GermaParl Corpus of Parliamentary Protocols. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018). European Languages Resources Association (ELRA). https://www.aclweb.org/anthology/L18-1130 (11 October 2019).

Boelmann, B. and S. Schaffner (2019). FDZ data description: FDZ Data description: Real-Estate Data for Germany (RWI-GEO-RED) - Advertisements on the Internet Platform ImmobilienScout24, RWI Projektberichte, Essen. http://www.rwi-essen.de/media/content/pages/publikationen/rwi-projektberichte/fdz_datenbeschreibung-rwi-geo-red-v1.pdf (11 October 2019).

Bosnjak, M.; R. Bauer and K.W. Weyandt (2018): Mixed devices in online surveys: Prevalence, determinants, and consequences. In: A. Theobald (ed.). Mobile Research: Grundlagen und Zukunftsaussichten für die Mobile Marktforschung. Wiesbaden, Springer Gabler, 53-65.

Budroni, P.; J.-C. Burgelman and M. Schouppe (2019). Architectures of Knowledge: The European Open Science Cloud. ABI Technik, 39 (2), 130-141.

Bug, M.; S. Liebig, C. Oellers and R. T. Riphahn (2018). Operative und strategische Elemente einer leistungsfähigen Forschungsdateninfrastruktur in den Sozial- und Wirtschaftswissenschaften. Jahrbücher für Nationalökonomie und Statistik, 238 (6), 571–590. https://doi.org/10.1515/jbnst-2018-0029 (11 October 2019).

Carstensen, J. and **M. Jungbauer-Gans** (2016). Health returns on education and educational systems. In: A. Hadjar and C. Gross (eds.). Education system and inequalities. International Comparisons. Bristol, UK, Policy Press, 301-321.

CESSDA (2017). Report on the establishment of an international curriculum for professional development in digital data services for the social sciences. Consortium of European Social Sciences Data Archives, http://cessdasaw.eu/content/uploads/2017/11/D5.3_CESSDA_SaW_v1.0.pdf (11 October 2019).

CESSDA Training Working Group (2017-2018). CESSDA Data Management Expert Guide. Consortium of European Social Sciences Data Archives, Bergen. https://www.cessda.eu/DMGuide (11 October 2019).

Corti, L.; A. Witzel and L. Bishop (2005). Sekundäranalyse qualitativer Daten. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 6 (1), Art. 49.

CRAN. https://cran.r-project.org/ (11 October 2019).

da|ra - DOI-Registrierungsservice für Sozial- und Wirtschaftsdaten. https://www.da-ra.de/ (11 October 2019).

DataCite. https://datacite.org/index.html (11 October 2019).

DataCitationIndex. https://clarivate.com/webofsciencegroup/master-data-repository-list/ (11 October 2019).

DataWiz. Research data management tool. https://datawiz.leibniz-psychology.org/DataWiz/ (11 October 2019).

DDI Lifecycle 3.2. https://ddialliance.org/explore-documentation (11 October 2019).

Destatis (2019a). Bildung und Kultur, Personal an Hochschulen, Vorläufige Ergebnisse. Statistisches Bundesamt, Wiesbaden. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/personal-vorbericht-5213402188004.pdf?__blob=publicationFile (11 October 2019).

Destatis (2019b). Bildung und Kultur, Studierende an Hochschulen, Vorbericht. Fachserie 11 Reihe 4.1, WS 2018/2019, Statistisches Bundesamt, Wiesbaden. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/studierende-hochschulen-vorb-2110410198004.pdf?__blob=publicationFile (11 October 2019).

DFG (2019). Umgang mit Forschungsdaten in der DFG-Einzelförderung – Eine Analyse von Anträgen in den Sozial- und Verhaltenswissenschaften. DFG Infobrief 1.19. Deutsche Forschungsgemeinschaft, Bonn. https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/infobriefe/ib01_2019.pdf (11 October 2019).

Dohmen, T.; A. Falk, D. Huffmann, U. Sunde, **J. Schupp** and G. G. Wagner (2011). Individual Risk Attitudes. Journal of the European Economic Association, 9 (3), 522-550.

Döring, H. and S. Regel (2019). Party Facts: A database of political parties worldwide. Party Politics, 25 (2), 97-109.

EcoSoc-IN. https://www.go-fair.org/implementation-networks/overview/ecosoc-in/ (11 October 2019).

Eder, C. and A. Jedinger (2018). FAIR national election studies: How well are we doing? European Political Science, online first, 1-18. https://doi.org/10.1057/s41304-018-0194-3 (11 October 2019).

ERDA. https://gitlab.tba-hosting.de/ForschungsDatenBildung/erda (14 October 2019).

European Commission (2018). Turning FAIR into Reality. https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283 (11 October 2019).

Feinerer, I. and K. Hornik (2018). tm: Text Mining Package. R package version 0.7-6. https://CRAN.R-project.org/package=tm (11 October 2019).

Flöck, F.; K. Erdogan and M. Acosta (2017). TokTrack: A Complete Token Provenance and Change Tracking Dataset for the English Wikipedia. https://arxiv.org/abs/1703.08244 (11 October 2019).

Förster, A.; K. Blask and M. Lemaire (2019). Streamlining research data management - The research perspective of the PODMAN project. BIBLIOTHEK - Forschung und Praxis, 43 (1), 61–67.

GeRDI. https://www.gerdi-project.eu/ (11 October 2019).

GESIS Data Search. https://datasearch.gesis.org/start (11 October 2019).

Giesselmann, M.; M. Stanema, **J. Schupp** and D. Richter (2018). Das Sozio-Oekonomische Panel als Datenbasis für die Arbeits- und Organisationspsychologie: Potentiale, Limitationen und Analysemethoden längsschnittlicher Surveydaten. Zeitschrift für Arbeits- und Organisationspsychologie, 62 (3), 111-125.

GitHub. https://github.com/ (11 October 2019).

Glöckner, F. O.; M. Diepenbroek, J. Felden, J. Overmann, A. Bonn, B. Gemeinholzer, A. Güntsch, B. König-Ries, B. Seeger, A. Pollex-Krüger, J. Fluck, I. Pigeot, T. Kirsten, T. Mühlhaus, **C. Wolf**, U. Heinrich, C. Steinbeck, O. Koepler, O. Stegle, J. Weimann, T. Schörner-Sadenius, C. Gutt, F. Stahl, K. Wagemann, T. Schrade, R. Schmitt, C. Eberl, F. Gauterin, M. Schultz, L. Bernard (2019). Berlin Declaration on NFDI Cross-Cutting Topics. https://doi.org/10.5281/zenodo.3457212 (11 October 2019).

Goebel 2014: Job submission instructions for the SOEPremote System at DIW Berlin – Update 2014, SOEP Survey Papers (195), Series G - General Issues and Teaching Materials. https://www.diw.de/documents/publikationen/73/diw_01.c.570638.de/diw_ssp0195.pdf (11 October 2019).

Griesbaum, J.; N. Mahrholz, K. von Löwe Kiedrowski and **M. Rittberger** (2015). Knowledge generation in online forums: A case study in the German educational domain. Aslib Journal of Information Management, 67 (1), 2-26.

Grube, M.M.; C. Scheidt-Nave, B. Gaertner, D. Lüdtke, **M. Bosnjak**, S. Heinrich, N. Lahmann, G. Meyer, K. Rapp, S. Riedel-Heller, M. Schäufele, K. Wolf-Ostermann, S. Zank and J. Fuchs (2019). Public Health Monitoring 65+ in Deutschland - Auswahl und Definition von Indikatoren. Journal of Health Monitoring, 4 (2), 93-109.

Gurr, T.; S. Unger and **M. Jungbauer-Gans** (2018). Gehen Sanktionen mit einem höheren Stigma-Bewusstsein bei Arbeitslosen einher? Zeitschrift für Sozialreform 64 (2), 217-248.

Halevi, G.; H. Moed and J. Bar-Ilan (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation. Journal of Informetrics, 11 (3), 823-834.

Harari, G. M.; S. R. Müller, C. Stachl, R. Wang, W. Wang, M. Bühner and S. D. Gosling (2019). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. Journal of Personality and Social Psychology, Advance online publication, 20 May 2019. https://doi.org/10.1037/pspp0000245 (11 October 2019).

Heisig, J. P.; M. Gesthuizen and **H. Solga** (2019). Lack of skills or formal qualifications? New evidence on cross-country differences in the labor market disadvantage of less-educated adults. Social Science Research, (online first, 2 July 2019). https://doi.org/10.1016/j.ssresearch.2019.06.005 (11 October 2019).

Hienert, D.; D. Kern, K. Boland B., Zapilko and P. Mutschke (2019). A Digital Library for Research Data and Related Information in the Social Sciences. ACM/IEEE Joint Conference on Digital Libraries (JCDL), 148-157.

Hoffmeyer-Zlotnik, J. H. P (2016). Demographische Standards, Ausgabe 2016: Eine gemeinsame Empfehlung des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und des Statistischen Bundesamtes. Wiesbaden, Statistisches Bundesamt.

Hollstein, B. (2011). Qualitative Approaches. In: J. Scott and P. J. Carrington (eds.). Sage Handbook of Social Network Analysis. London/New Delhi, Sage Publications, 404-417.

Hollstein, B. (2019). What Autobiographical Narratives Tell Us About the Life Course? Contributions of qualitative sequential analytical methods. Advances in Life Course Research, 41, online-first. https://doi.org/10.1016/j.alcr.2018.10.001 (11 October 2019).

Hollstein, B. and J. Strübing (2018). Archivierung und Zugang zu qualitativen Daten. RatSWD Working Paper 267/2018. German Data Forum, Berlin. https://doi.org/10.17620/02671.35 (11 October 2019).

ICS ISO 9241-210:2019 (2019). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. International Organisation for Standardisation. https://www.iso.org/home.html (11 October 2019).

Jones, S. and J.-F. Abramatic (2019). European Open Science Cloud (EOSC), Strategic Implementation Plan, Version 1.0, June 2019. https://doi.org/10.2777/202370 (11 October 2019).

JoSuA. https://fdz.iab.de/en/FDZ_Data_Access/FDZ_Remote_Data_Access.aspx (11 October 2019).

Kerwer, M.; R. Bölter, I. Dehnhard, A. Günther and E. Weichselgartner (2017). Projekt DataWiz: Entwicklung eines Assistenzsystems zum Management psychologischer Forschungsdaten. In: J. Kratzke and V. Heuveline (eds.). E-Science-Tage 2017: Forschungsdaten managen. Heidelberg, heiBOOKS.

Kühn, S.; S. Düzel, P. Eibich, C. Krekel, H. Wüstemann, J. Kolbe, J. Martensson, J. Goebel, J. Gallinat, G. G. Wagner and U. Lindenberger, (2017). In search of features that constitute an “enriched environment” in humans: Associations between geographical properties and brain structure. Scientific Reports 7, Article number: 11920.

Latif A., T. Borst and **K. Tochtermann** (2014). Exposing Data From an Open Access Repository for Economics As Linked Data (2014). D-Lib Magazine, 20 (9/10).

Latif A.; F. Limani and **K. Tochtermann** (2018). Research data management for long tail research data – a generic research data infrastructure approach. CODATA Data Science Journal, Art. 17.

Mayer-Ahuja, N. (2017). Die Globalität unsicherer Arbeit als konzeptionelle Provokation: Zum Zusammenhang zwischen Informalität im Globalen Süden und Prekarität im Globalen Norden. Special Issue Arbeit und Kapitalismus, Geschichte und Gesellschaft. Zeitschrift für Historische Sozialwissenschaft, 2, 264-296.

Medjedovic, I. and A. Witzel (2010). Wiederverwendung qualitativer Daten: Archivierung und Sekundärnutzung qualitativer Interviewtranskripte, Berlin: SpringerVS.

Meyermann, A.; D. Bambey, M. Jansen, R. Mauer, T. Ebel, Eisentraut, M., K. Harzenetter, P. Kuhl, C. Neuendorf, L. Pegelow, M. Porzelt, **M. Rittberger**, T. Schwager, P. Stanat and J. Trixa (2017). Der Verbund Forschungsdaten Bildung–Eine Forschungsdateninfrastruktur für die empirische Bildungsforschung. RatSWD Working Paper Series, No. 266. https://doi.org/10.17620/02671.27 (11 October 2019).

NEPStools; Stata Tools for RDM of NEPS data. https://www.neps-data.de/de-de/datenzentrum/%C3%BCbersichtenundhilfen/statatools.aspx (11 October 2019).

Protsch, P. and **H. Solga** (2015). How employers use signals of cognitive and noncognitive skills at labor market entry. Insights from field experiments. European Sociological Review, 31 (5), 521-532.

PsychArchives. A disciplinary repository preserving a variety of digital research objects (DROs), with 21 different publication types. https://www.psycharchives.org/ (11 October 2019).

Quinlan, S.; T. Gummer, J. Roßmann and **C. Wolf** (2018). “Show me the money and the party!” – Variation in Facebook and Twitter adoption by politicians. Information, Communication, and Society, 21 (8), 1031-1049.

RatSWD (2017a). Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften. RatSWD Output 9 (5). Berlin, German Data Forum. https://doi.org/10.17620/02671.1 (11 October 2019).

RatSWD (2017b). Tätigkeitsbericht der akkreditierten Forschungsdatenzentren des Rates für Sozial- und Wirtschaftsdaten (RatSWD) für das Jahr 2015. RatSWD Output 7 (5). Berlin, German Data Forum. https://doi.org/10.17620/02671.3 (11 October 2019).

RatSWD (2017c). Initiative des RatSWD zur strategischen Vernetzung deutscher Panelsurveys. Berlin, German Data Forum. https://www.ratswd.de/dl/media/180215_RatSWD_Paneltreffen2017_Protokoll.pdf (11 October 2019).

RatSWD (2017d). Geschäftsordnung des Ständigen Ausschusses Forschungsdateninfrastruktur (FDI Ausschuss) des Rates für Sozial- und Wirtschaftsdaten (RatSWD) in der am 07.12.2017 verabschiedeten Fassung. Berlin, German Data Forum. https://www.ratswd.de/dl/downloads/GO_Ausschuss_FDI.pdf (11 October 2019).

RatSWD (2018a): Tätigkeitsbericht 2017 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ). Berlin, German Data Forum. https://doi.org/10.17620/02671.33 (11 October 2019).

RatSWD (2018b). The German Data Forum (RatSWD) and Research Data Infrastructure: Status Quo and Quality Management. RatSWD Output 1 (6). Berlin, German Data Forum. https://doi.org/10.17620/02671.30 (11 October 2019).

RatSWD (2018 c): FDI Clusterbefragung. Internal report for 19th meeting of FDI Ausschuss.

RatSWD (2019a). Tätigkeitsbericht 2018 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ). Berlin, German Data Forum. doi: https://10.17620/02671.40 (15 October 2019).

RatSWD (2019b). RatSWD Vernetzungstreffen der Fachgesellschaften. Berlin, German Data Forum. https://www.ratswd.de/dl/media/190710_RatSWD-Vernetzungstreffen-FG-Dokumentation.pdf (11 October 2019).

RatSWD (2019c) https://www.ratswd.de/dl/media/190924_RatSWD-Vernetzungstreffen-Ethik.pdf (11 October 2019).

RatSWD (2019 d) Geschäftsordnung des Rates für Sozial- und Wirtschaftsdaten (RatSWD) in der am 01.07.2020 gültigen Fassung. Berlin, German Data Forum. https://www.ratswd.de/dl/downloads/RatSWD_GO_Sep19.pdf (14 October 2019).

ReadStat. https://github.com/WizardMac/ReadStat (11 October 2019).

Reinhold, A.; **M. Rittberger** and N. Mahrholz (2016). Data Citation Policies of Data Providers within the scope of Longitudinal Studies in Life Course Research. In: I. Ràfols, J. Molas-Gallart, E. Castro-Martínez and R. Woolley (eds.). 21st International Conference on Science and Technology Indicators - STI 2016. Book of Proceedings, 115-121.

RfII (2016). Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen, Rat für Informationsinfrastrukturen, 30. http://www.rfii.de/?p=1998 (11 October 2019).

RfII (2019). Digitale Kompetenzen - dringend gesucht! Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft. Göttingen, Rat für Informations Infrastrukturen, 56. http://www.rfii.de/?p=3883 (11 October 2019).

Seeger, M. (2019). Ergebnisse der Umfrage zu Forschungsdaten in den asienbezogenen Wissenschaften. CrossAsia Blog. https://blog.crossasia.org/ergebnisse-der-umfrage-zu-forschungsdaten-in-den-asienbezogenen-wissenschaften/ (11 October 2019).

Sieber, J. E. (1991). Sharing social science data: Advantages and Challenges. Sage focus editions, 128. Thousand Oaks, Sage.

Silge, J. and D. Robinson (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. JOSS,1(3). https://doi.org/10.21105/joss.00037 (11 October 2019).

Sloan, L.; C. Jessop, T. Al Baghal and M. Wiliams (2019). Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving. Journal of Empirical Research on Human Research Ethics, Special Issue on Ethical Issues in Social Media Research, 1-14.

Soephelp. A stata-ado, to present metadata documentation in a user-friendly way. http://companion.soep.de/Working%20with%20SOEP%20Documentation/Working%20with%20SOEPhelp.html (11 October 2019).

SOEPremote. https://www.diw.de/sixcms/detail.php?id=diw_01.c.32035.en (11 October 2019).

VerbundFDB. https://www.forschungsdaten-bildung.de/ (11 October 2019).

von Hodenberg, C. (2018). Das andere Achtundsechzig: Gesellschaftsgeschichte einer Revolte. München, CH Beck.

Wagner, G. G. (2017). Anmerkungen zu den vielfältigen Dimensionen einer Forschungsethik in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. RatSWD Working Paper Series 265. https://doi.org/10.17620/02671.10 (11 October 2019).

Wickham, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse (11 October 2019).

Wickham, H. and E. Miller (2019). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.1.0. https://CRAN.R-project.org/package=haven (11 October 2019).

Wissenschaftsrat (2009): Wissenschaftspolitische Stellungnahme zum Status und der zukünftigen Entwicklung des Rates für Sozial- und Wirtschaftsdaten (RatSWD). Drs. 9504-09. Berlin. https://www.ratswd.de/dl/downloads/9504-09.pdf (11 October 2019).

Wissenschaftsrat (2011). Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Drs. 10465-11. Berlin. https://www.wissenschaftsrat.de/download/archiv/10465-11.pdf;jsessionid=947594311989891E2366B5CB396E9580.delivery2-master?__blob=publicationFile&v=3 (11 October 2019).

Wittenburg, P.; G. Strawn, B. Mons, L. Boninho and E. Schultes (2019). Digital Objects as Drivers towards Convergence in Data Infrastructures, EUDAT. https://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11 (11 October 2019).

WZB (2019). Schlussbericht des Projekts "Geschäftstelle des Rates für Sozial- und Wirtschaftsdaten (RatSWD) sowie wissenschaftliche, administrative, kommunikative und technische Unterstützung des RatSWD" nach Nr. 3.2 BNBest-BMBF 98: Berichtszeitraum: 01.06.2014-31.10.2018. https://doi.org/10.2314/KXP:1671910109 (11 October 2019).

ZBW, GESIS & RatSWD (2014). Auffinden, Zitieren, Dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften, Hamburg. https://doi.org/10.4232/10.fisuzida2015.2 (11 October 2019).

2 Curricula vitae and lists of publications

- Dr. Maja Adena; WZB, Berlin
- Prof. Dr. Christian Aßmann; LIfBi, Bamberg
- Prof. Dr. Andreas Blätte; Universität Duisburg-Essen
- Prof. Dr. Michael Bosnjak; ZPID, Trier
- Prof. Dr. Betina Hollstein; Universität Bremen
- Prof. Dr. Monika Jungbauer-Gans; DZHW, Hannover
- Prof. Dr. Nicole Mayer-Ahuja; SOFI, Göttingen
- Prof. Dr. Marc Rittberger; DIPF, Frankfurt am Main
- Prof. Dr. Jürgen Schupp; SOEP am DIW Berlin
- Prof. Dr. Heike Solga; WZB, Berlin
- Prof. Dr. Klaus Tochtermann; ZBW, Kiel
- Prof. Dr. Christof Wolf; GESIS, Mannheim

3 Letters of commitment by the participants

- Forschungsdatenzentrum am BAMF
- Forschungsdatenzentrum BAuA
- Forschungsdatenzentrum im BIBB (BIBB-FDZ)
- Forschungsdatenzentrum der BZgA (FDZ BZgA)
- Forschungsdatenzentrum des Sozio-oekonomischen Panels am DIW Berlin (FDZ SOEP)
- Forschungsdatenzentrum Betriebs- und Organisationsdaten am DIW Berlin (FDZ-BO)
- Forschungsdatenzentrum des DJI (FDZ-DJI)
- Forschungsdatenzentrum des DZA (FDZ-DZA)
- Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung (fdz.DZHW)
- Forschungsdatenzentrum Deutsches Zentrum für Integration und Migration (DeZIM.fdz)
- Forschungsdatenzentrum Bildung (FDZ Bildung) am DIPF
- Internationales Forschungsdatenzentrum des IZA (FDZ IZA, IDSC)
- Regina T. Riphahn, Chairperson of the German Data Forum (RatSWD)
- Forschungsdatenzentrum Wahlen bei GESIS (FDZ Wahlen)
- Forschungsdatenzentrum PIAAC bei GESIS (FDZ PIAAC)
- Forschungsdatenzentrum Internationale Umfrageprogramme bei GESIS (FDZ Internationale Umfrageprogramme)
- Forschungsdatenzentrum German Microdata Lab bei GESIS (FDZ GML)
- Forschungsdatenzentrum ALLBUS bei GESIS (FDZ ALLBUS)
- LOEWE-Zentrum SAFE
- Forschungsdatenzentrum der Hans-Böckler Stiftung am WSI (FDZ WSI)
- LMU-ifo Economics & Business Data Center (EBDC)
- Forschungsdatenzentrum der Statistischen Ämter der Länder (FDZ-Länder)
- Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB (FDZ BA im IAB)
- Forschungsdatenzentrum Archiv für Gesprochenes Deutsch am IDS (FDZ AGD)
- Forschungsdatenzentrum am IQB (FDZ IQB)
- Forschungsdatenzentrum am IÖR
- Forschungsdatenzentrum des IWH (FDZ-IWH)
- Forschungsdatenzentrum PsychData des ZPID (FDZ PsychData des ZPID)
- Forschungsdatenzentrum des LIfBi an der Otto-Friedrich-Universität Bamberg (FDZ-LIfBi)
- Forschungsdatenzentrum des Beziehungs- und Familienpanels (FDZ-pairfam)

- Forschungsdatenzentrum des Survey of Health, Ageing and Retirement in Europe (FDZ-SHARE)
- Forschungsdatenzentrum des RKI (FDZ RKI)
- Forschungsdatenzentrum Ruhr am RWI (FDZ Ruhr am RWI)
- Forschungsdatenzentrum eLabour (FDZ eLabour)
- Forschungsdatenzentrum des Statistischen Bundesamtes (FDZ-Bund)
- Forschungsdatenzentrum Wissenschaftsstatistik des Stifterverbandes (FDZ Wissenschaftsstatistik)
- Forschungsdatenzentrum Qualiservice (FDZ Qualiservice)
- Forschungsdatenzentrum der Universitätsbibliothek Mannheim (FDZ UB Mannheim)
- Forschungsdatenzentrum des ZEW (ZEW-FDZ)

4 Letters of support

15 academic associations:

- Akademie für Soziologie (AS)
- Deutsche Gesellschaft für Erziehungswissenschaft (DGfE)
- Deutsche Gesellschaft für Gesundheitsökonomie (DGGÖ)
- Deutsche Gesellschaft für Medizinische Psychologie (DGMP)
- Deutsche Gesellschaft für Politikwissenschaft (DGfP)
- Deutsche Gesellschaft für Psychologie (DGPs)
- Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPuK)
- Deutsche Gesellschaft für Soziologie (DGS)
- Deutsche Gesellschaft für Sozial- und Kulturanthropologie (DGSKA)
- Deutsche Gesellschaft für Volkskunde (DGV)
- Deutsche Statistische Gesellschaft (DStatG)
- Deutsche Vereinigung für Politische Wissenschaft (DVPW)
- Gesellschaft für empirische Bildungsforschung (GEBF)
- Verband der Hochschullehrer für Betriebswirtschaft (VHB)
- Verein für Socialpolitik (VfS)

7 specialised information services (*Fachinformationsdienste*):

- Fachinformationsdienst Erziehungswissenschaft und Bildungsforschung
- Fachinformationsdienst Sozial- und Kulturanthropologie
- Fachinformationsdienst Mobilitäts- und Verkehrforschung
- Fachinformationsdienst Politikwissenschaft
- Fachinformationsdienst Soziologie (SocioHub)
- Fachinformationsdienst für Medien-, Kommunikations- und Filmwissenschaft (adlr.link)
- Fachinformationsdienst Kriminologie

8 stakeholders in the field of research ethics:

- Antidiskriminierungsstelle des Bundes
- Ethikausschuss der Deutschen Gesellschaft für Publikzistik- und Kommunikationswissenschaft
- Deutsche Gesellschaft für Soziologie

- Ethikkommission des Fachbereichs Wirtschafts- und Sozialwissenschaften der Friedrich-Alexander Universität Erlangen-Nürnberg
- Ethikkommission GESIS
- Ethikkommission der Volkswirtschafteilchen Fakultät der Ludwigs-Maximilians-Universität München
- Ethikkommission der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Universität Heidelberg
- Ethikkommission des Wissenschaftszentrums für Sozialforschung Berlin

Forschungsdatenzentrum im Kraftfahrtbundesamt (FDZ im KBA)

Leibniz Association