

1 Binding letter of intent as advance notification or non-binding letter of intent

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

2 Formal details

- Planned name of the consortium
NFDI for Business, Economic and Related Data
- Acronym of the planned consortium
BERD@NFDI
- Applicant institution
University of Mannheim, Schloss, 68161 Mannheim
Prof. Dr. Thomas Puhl
- Spokesperson
Prof. Dr. Florian Stahl (florian.stahl@uni-mannheim.de), Chair of Quantitative Marketing and Consumer Analytics, Co-Director of Mannheim Center for Data Science

3 Objectives, work programme and research environment

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

www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtsperiode_2016_2019/fachsystematik_2016-2019_en_grafik.pdf

- 12 Social and Behavioral Sciences
- 112 Economics
- 112-02 Economic Policy and Public Finance

Deutsche Forschungsgemeinschaft

Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany
Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de



- 112-03 Business Administration
 - 112-04 Statistics and Econometrics
 - 112-05 Economic and Social History
 - 111 Social Sciences
 - 111-04 Political Science
 - 110 Psychology
 - 110-03 Social Psychology, Industrial and Organizational Psychology
- Concise summary of the planned consortium's main objectives and task areas
- Digitalization is part of the very fabric of economic, societal, and business life. According to estimates, between 80 and 90 % of all relevant business information resides in unstructured form such as text, image, voice, or video data. Existing data infrastructures are not prepared to handle this form of data. Novel standards to access and structure unstructured data and to integrate it with available structured data and infrastructures are to be developed for the future data landscape. Researchers in business, economics and related fields need efficient processes and tools, an efficient data infrastructure and a comprehensive implementation of pertinent methodological knowledge for their research and teaching. BERD@NFDI is a cooperation between the Universities of Mannheim, Cologne, Hamburg, and Munich, the Institute of Employment Research (IAB) as well as the ZEW, and is supported by community partners, such as the German Academic Association for Business Research (VHB). It brings together leading institutions in business, economics and social sciences, which intend to contribute their best resources in order to exploit the new types of data for evidence-based empirical research. The consortium will be supported by leading research institutions, such as the Research Center „Sustainable Architecture for Finance in Europe“ (SAFE), and infrastructure organizations, such as ZEW-FDZ, the ZBW, the University Library of Mannheim, and the Research Data and Service Center (RDSC) of the Deutsche Bundesbank. As a structural contribution to NFDI, BERD@NFDI aims to develop and disseminate transparent, open and innovative standards and tools to manage unstructured data and to combine and connect them with structured data in economics, business and related research fields. This involves transparent, open and innovative standards for infrastructures, protocols and metadata
- to search, to index as well as to process unstructured data,
 - to integrate various sources of structured, semi-structured and unstructured data,
 - to share structured, semi-structured and unstructured data,
 - to homogenize unstructured data,
 - to validate the accuracy, consistency and completeness of the data.

Task area 1: Supporting the BERD@NFDI community in accessing, aggregating, classifying and evaluating unstructured data

Standardized procedures and methods for accessing and aggregating unstructured data: Automated classification is needed prior to analysis and integration with other data sources. Useful classifications must follow the requirements of the intended insights purpose. While some are relevant for various questions of business and economics (e.g., sentiment analysis, fraud detection), others are highly domain specific (e.g., types of marketing messages). At the same time, automated data classification benefits from prior classifications of similar problems (transfer learning).

Standardized systems and methods for data classification and quality assurance: Managing unstructured data is not limited to data infrastructure, it includes standards for algorithmic data classification to allow a broad range of applications and leverage transfer learning. Ensuring open access and open collaboration to both data and classification approaches is essential to ensure knowledge building in business and economics in a reliable and verifiable manner.

Exchange and sharing of trained and standardized AI algorithms to improve the extraction of structure from unstructured data: The new types of data are meaningless without knowing the data generating context, e.g., of the consolidation level (user/individual level, company level etc.) and of the applied measurement units. Data must often be transformed before it can be interpreted by common data analysis programs.

Task area 2: Supporting the BERD@NFDI community to better interlink structured and unstructured data

Collection and provision of unstructured data: The methods developed in TA1 have to be tested and applied to relevant datasets in the subject areas of the communities addressed by the consortium. Therefore, the infrastructure partners of the consortium will collect datasets (e.g. via webscraping, APIs, cooperative agreements) that serve as models for the management of unstructured data and make them available to external researchers through remote access solutions.

Generating scientific-use-files via record linkage: A new enterprise ontology will allow the linking and deep indexing of existing structured and unstructured datasets in a standardized way. BERD@NFDI will process data available at the partner institutions, and seek cooperation with other data providers (the Research Data Centers accredited by RatSWD, but also other research organizations) to generate scientific-use-files enriched with unified identifiers.

Development of a metasearch and an information portal for research data in business and economics: A specialized metasearch for business and economic data will improve its discoverability and enable specific queries according to scientifically relevant criteria. In order to make it easier for researchers to find and use existing resources, a central access point for the relevant

information must be established. It will combine user-friendly retrieval functionalities with information on national and international activities concerning research data management in business and economics.

Task area 3: Supporting the BERD@NFDI community to build up competences and capacities through on-site and online training modules and workshops: To avoid data graveyards, project-based workshops will provide professional training on the new forms of data in synchronous and asynchronous formats. Building on well-tested models in sub-spaces by the project partners, these training modules will be to the benefit of the entire community addressed by BERD@NFDI.

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

The planned infrastructure is based on the Business and Economic Research Data Center Baden-Württemberg (BERD-Center BW). The BERD Center BW is one of four Science Data Centers funded within the framework of the digital@bw, the digitalization strategy of the state Baden-Württemberg. The consortium of the BERD-Center BW consists of scientists from the ZEW and the Mannheim Center for Data Science, who are outstanding in their fields both nationally and internationally. They are joined by the infrastructure providers of the University and ZEW, namely the Mannheim University Library, the IT Services Center and the ZEW-FDZ, which bring specialist expertise in the field of research data, data processing and data-intensive computing.

Furthermore, BERD@NFDI builds on the following existing infrastructures:

- ZBW, which is the world's largest information infrastructure for economics. ZBW will bring in its entire portfolio for research data management, including the technologies developed within the DFG-funded project GeRDI as well as the entire portfolio of FAIR data implementation networks of the GoFAIR initiative. ZBW will provide its federation and harvesting technologies for research data repositories as well as its tools for metadata normalization developed within the GeRDI project. Latest developments within the FAIR data movement will be ensured through ZBW's leading role in this initiative.
- DFG Research Group on the impact of social media headed by the University of Hamburg and the University of Cologne, which generates, shares and works with both structured and unstructured data from various online channels.
- The BMBF-funded International Program in Survey and Data Science (IPSDS) established in collaboration between the University of Mannheim and the University of Maryland, with inputs from the Institute for Employment Research, the University of Munich and the Bundesbank. IPSDS will provide a platform and starting point for the asynchro-

nous professional training opportunities on the next types of data.

- The data and infrastructure services provided by the research data center of the University Library of Mannheim (see <https://fdz.bib.uni-mannheim.de/>), e.g. the Aktienführer Datenarchiv (<https://digi.bib.uni-mannheim.de/aktienf%C3%BChrer/data/index.php>), will also form a part of BERD@NFDI.
- The Coleridge Initiative (<https://coleridgeinitiative.org/>) – with a partnership to BERD@NFDI's co-spokespersons – successfully implemented a secure Administrative Research Data Facility which holds in addition to more structured records unstructured data as part of a joint data schema. The environment runs within AWS and has been successfully tested to provide a training platform and to allow remote access to researchers. The agility of this platform serves as a role model for the planned infrastructure.

BERD@NFDI will seek cooperation with other data providers (the Research Data Centers accredited by RatSWD, but also other research organizations) to create a comprehensive discovery service for relevant data sets and enhance their interoperability and standardization.

- Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration
 - Leading up to this Lol, BERD@NFDI had intensive discussions with KonsortSWD. There is a strong interest to collaborate on both sides. Synergistic activities are possible in particular around the task areas 2 and 3 (indexing and ontologies, training of working professionals).
 - ForumX and BERD@NFDI will strive for close collaboration. While BERD@NFDI deals with unstructured data from business, economics and related fields, ForumX focusses on the experimental method. In fields of intersection (conjoint analysis in marketing – factorial surveys in sociology; unstructured finance data and experimental finance data), both consortia will work closely together. Fields of collaboration are (1) metadata standards and models of metadata harmonization, (2) transdisciplinary research regarding methods and algorithms (AI), (3) the combination of similar experimental datatypes in different communities (factorial surveys in sociology vs. conjoint analysis in marketing).
 - The cross-cutting consortia BRIDGE4NFDI and 2linkNFDI follow promising approaches to solve cross-cutting topics, e.g. the linking of services from different NFDI consortia.
 - Similar to BERD@NFDI, Text+ will have a focus on analytics of unstructured text data, so that cooperation in this area is conceivable.
 - Possibilities of cooperation with AI4NFDI and NFDI Web will be evaluated.

4 Cross-cutting topics

- Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.
 - *Federation of data sources, interoperability:* Cross-disciplinary research data management is only possible if sophisticated federation methods are in place connecting data sources through a common protocol. These methods include normalization of metadata, homogenization of data etc.
 - *Support of the data life cycle:* Data life cycles, such as the life cycle from the UK data archive, call for several services (e.g. search, store, process, analyse) for a proper data management, regardless of the domain which uses these services.
 - *Licensing of data:* Research data in social sciences, but also in life sciences and other disciplines, often contain personal data. Therefore, appropriate licenses have to regulate the use of the data, depending on the conditions of the data collection and the degree of anonymization. Cross-disciplinary research data management connects data with different licenses to one another. This requires services for proper license management (e.g. overview of usage terms).
 - *FAIR data support:* The FAIR principles for research data are about to become a de facto standard for high-quality data. Measuring the “FAIRness” (i.e. the level of maturity) of research data is a topic, which cannot be solved within one discipline.
 - *Visibility in the European Open Science Cloud (EOSC):* Research data which is offered by research data repositories hosted in Germany should not stay isolated within NFDI but should be connected to the European data space and thus become visible in the EOSC.
 - *Training in data management:* The report of the German Council for Information Infrastructures (RfII) and the High Level Expert Group on the EOSC have identified a clear need for more building up new competences, more capacities and new curricula for research data management.
 - *Rules of Participation:* NFDI needs terms and conditions defining which data sources and services will be included to or excluded from the NFDI.
- Please indicate which of these cross-cutting topics your consortium could contribute to and how.

A strength of BERD@NFDI is the focus on the unstructured data from different data sources, thus BERD@NFDI will contribute to all of the cross-cutting topics mentioned

above. For instance:

- *Federation of data sources, interoperability:* The standards developed in task area 1 and the interlinking of data sets from various sources (task area 2) will significantly enhance the interoperability of existing data sets in business and economics. The tools and methods used to do this can be relevant for other disciplines as well. BERD@NFDI will also build on the DFG-project GeRDI, which provides a technology for the federation of data sources. GeRDI relies on DataCite, a generic metadata standard, which facilitates the normalization of metadata.
- *Support of the data life cycle:* GeRDI also offers services for each phase of the data life cycle.
- *Licensing of data:* When generating and providing its own scientific-use files, BERD@NFDI will work towards making the data available under open licenses as far as possible. Necessary restrictions on access rights are to be formulated in a standardized, machine-readable form in the metadata. BERD@NFDI will contribute to the further development of existing standards for the formulation of rights according to the needs of the economic and social sciences.
- *FAIR data support:* BERD@NFDI partners are drivers of the GoFAIR initiative, which aims at establishing the FAIR data principles across different domains. BERD@NFDI will be closely linked up to and collaborate with relevant implementation networks of this initiative.
- *Visibility in the EOSC:* BERD@NFDI partners have been involved in the development of the EOSC since its inception. Using GeRDI technology, data with relevance to BERD@NFDI will become visible to the EOSC.
- *Training:* The training concepts and modules developed in task area 3 can serve as a model for other disciplines.
- *Rules of participation:* NFDI can capitalize on the results from the Working Group “Rules of Participation” which has recently been established under the Executive Board of the EOSC. Through its close link to this as well as all other working groups BERD@NFDI will ensure a good knowledge transfer between EOSC and NFDI.