Letter of Intent

METHODS - National research data infrastructure for empirical research on human behaviour in Sports Science, Economics, Social Sciences, Psychology, and Medical Informatics

1  Binding letter of intent as advance notification of a full proposal

☒ Binding letter of intent (required as advance notification for proposals in 2021)

2  Formal details

Planned name of the consortium
National research data infrastructure for empirical research on human behaviour in Sports Science, Economics, Social Sciences, Psychology, and Medical Informatics

Acronym of the planned consortium

METHODS

Applicant institution
Leibniz University Hannover
Welfengarten 1
30167 Hannover
Prof. Dr. Volker Epping
Spokesperson
  • Dr. Markus Stocker, markus.stocker@tib.eu, Leibniz University Hannover, Institute of Distributed Systems
Co-spokesperson
  • Prof. Dr. Melanie Krüger, melanie.krueger@sportwiss.uni-hannover.de, Leibniz University Hannover, Institute for Sports Science
**Co-applicant institution**
University of Freiburg
Friedrichstr. 39
79098 Freiburg
Prof. Dr. Kerstin Krieglstein

**Co-spokesperson**
- Dr. Björn Grüning, grüning@informatik.uni-freiburg.de, University of Freiburg, Bioinformatik, Head of the Freiburg Galaxy Team, Technical systems management

**Co-applicant institution**
InfAI Leipzig
Institute for Applied Informatics (InfAI) at the University of Leipzig, InfAI Management GmbH
Goerdelerring 9
04109 Leipzig
Prof. Dr. Bogdan Franczyk

**Co-spokesperson**
- Dr. Sebastian Hellmann, hellmann@infai.org, University of Leipzig, Institute for Applied Informatics (InfAI), InfAI Management GmbH, Department Director

**Co-applicant institution**
University of Leipzig
Ritterstraße 26
04109 Leipzig
Prof. Dr. Beate A. Schücking

**Co-spokesperson**
- Prof. Dr. Roger Berger, berger@sozio.uni-leipzig.de, University of Leipzig, Faculty of Social Sciences and Philosophy, Institute of Sociology

**Co-applicant institution**
Ludwig-Maximilians-University Munich
Geschwister-Scholl-Platz 1
80539 Munich
Prof. Dr. rer. pol. Bernd Huber

**Co-spokesperson**
Co-applicant institution
University of Konstanz
Universitätsstraße 10
78457 Konstanz
Prof. Dr. Katharina Holzinger
Co-spokespersons
- Prof. Dr. Urs Fischbacher, urs.fischbacher@uni-konstanz.de, University of Konstanz, Department of Economics, Chair of Applied Research in Economics
- Prof. Dr. Susumu Shikano, susumu.shikano@uni-konstanz.de, University of Konstanz, Dept Politics and Public Administration

Co-applicant institution
TIB - Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Prof. Dr. Sören Auer
Co-spokesperson
- Prof. Dr. Sören Auer, auer@tib.eu, TIB - Leibniz Information Centre for Science and Technology
- Dr. Claudia Biniossek, claudia.biniossek@methods4nfdi.org, TIB - Leibniz Information Centre for Science and Technology
- Dr. Dirk Betz, dirk.betz@methods4nfdi.org, TIB - Leibniz Information Centre for Science and Technology

Participants
- Prof. Dr. Jeannette Brosig-Koch, Otto-von-Guericke-Universität Magdeburg
- Prof. Dr. Hans-Günther Döbereiner, University Bremen
- Prof. Dr. Sascha Füllbrunn, Radboud University Nijmegen, Netherlands
- Prof. Dr. Andreas Gerber-Grote, Zurich University of Applied Sciences, Switzerland
- Dr. Holger Gerhardt, University of Bonn
- Prof. Dr. Ben Greiner, Vienna University of Economics and Business, Austria
- Prof. Dr. Marcus Giamattei, University of Passau
Objectives, work programme and research environment

Research area of the proposed consortium (according to the DFG classification system: https://www.dfg.de/dfgprofil/gremien/fachkollegien/faecher/index.jsp)

109 Erziehungswissenschaft und Bildungsforschung
110 Psychologie
111 Sozialwissenschaften
111-02 Empirische Sozialforschung
111-03 Publizistik und Kommunikationswissenschaft
111-04 Politikwissenschaft
112 Wirtschaftswissenschaften
Concise summary of the planned consortium’s main objectives and task areas

The main objective of METHODS, merged with the ForumX consortium, is to converge and connect existing empirical research communities, including Sports Science, Economics, Social Sciences and Medical Informatics, which share substantial research methodologies, to support FAIR processes of data creation and management within research communities, as well as data transparency across research disciplines. The involved communities are characterized by significant similarity in their methods despite the many differences in data types, standards, and tools. However, they also share the same trends towards increasing volume, heterogeneity of types and formats, and a massive increase in relationships between Digital Objects (DOs) encapsulating discipline knowledge. Most communities have developed a mixture of discipline-specific and method-specific tools and have also started to build repositories with idiosyncratic structures. These discipline-centric developments severely limit the potential of shared infrastructures. The necessary integration into commons-based workflows and best practices constitutes the big challenge since traditions and discipline-specific characteristics and flavors need to be preserved as widely as possible.

METHODS is a method-oriented open alliance of empirical research communities, specifically Sports Science, Economics, Social Sciences, Psychology, and Medical Informatics, that is determined to work jointly on a common and modular workflow-framework that:

1. Covers the entire data lifecycle of the involved disciplines.
2. Guides researchers through seamless, semi-automatic research data management.
3. Serves to put the FAIR principles into practice with focus on interoperability and reusability and not just findability and accessibility.
4. Is based on a FAIR Digital Object (FDO) ecosystem as developed by RDA and FAIR core members.
5. Helps to overcome the many inefficiencies in data-driven science that result in a tremendous waste of time. At the core, METHODS builds on a common workflow framework (CWFR) that automates processes wherever possible and is designed to minimize disruptions and losses at critical points in the research process. The CWFR framework is based on an in-depth analysis of research lifecycles and infrastructures. The conclusion is that while there are many commonalities in data creation, management, and processing across many disciplines from all scientific fields, researchers are stuck with habitual data creation and management practices and methods - practices that generally do not result in FAIR data. In addition, the analysis uncovered that, in general, data
collection, management, and processing practices in the Research Data System (RDS)\(^1\) are not yet adapted to take full advantage of digitization. At the same time, publishers are pushing communities to adopt standards and practices. It is becoming increasingly apparent that the local research environment in which most data are created, managed, and consumed will only change if commonalities are addressed that lead to greater efficiencies and enable new types of research. The analysis broadly confirmed the discussion results of the RDA Data Fabric Group and led to, among other things, the notion of FAIR Digital Objects (FDO).

(6) Aligns with the legal and ethical regulations that are relevant for the involved disciplines in a tailored manner.

**METHODS** is committed to strictly respecting and supporting the specificities of the disciplines and communities involved. The core tasks are to:

1. Develop the modular workflow-framework (CWFR) as introduced above, based on commonalities to facilitate cross-silo/discipline data-driven research.
2. Facilitate data management by reducing, as much as possible, the growing administrative and management burden imposed by the increasing number of regulations, documentation profiles (DMP) and overall data trends (volume, complexity).
3. Provide state-of-the-art and FAIR-compliant repositories as widely as possible. The FDO approach ensures, at the data management level, that all digital objects are treated equally. Docking existing tools and repository systems to the FDO-System via adapters reduces the complexity from an N*N task, where each system must be connected to every other, to an 1*N task, where each system is connected to a single common layer exactly once. Once generic frameworks are available, extended with discipline-based adapters, meaningful modular metadata sets and interfaces, researchers will not be burdened in their scientific work; instead, their work will be advanced to new frontiers, and greater efficiencies will give them time to refocus on content issues and not waste \(\frac{3}{4}\) of their time on data wrangling, as is currently the case\(^2\).

**METHODS** sees itself as a facilitator to allow heterogeneous groups of researchers and research communities to participate in the emerging FAIR-based data space, making them part of the NFDI. **METHODS** is based on an integrative strategy for the next decades and therefore feels obliged to add a methodology-oriented dimension to NFDI. It, therefore, fits into the long-term intentions of NFDI.

\(^1\) We are using the generic term Research Data System (RDS) to point to all institutions, departments, or laboratories that generate data with the help of observations, experiments, analyses and simulations, that manage data or that are processing and analyzing data.

METHODS has begun to build international networks and technical interfaces of relevant infrastructures and toolmakers to implement FAIR Digital Objects in canonical workflows. In parallel, METHODS participates in ongoing deep discussions in RDA (Data Fabric, FAIR Maturity), GEDE (global discussions about FDO details), and GOFAIR Implementation Networks (Driving FAIR implementation). In addition, a new RDA Working group has been initiated that will develop a common generic workflow approach (CWFR).

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

METHODS will demonstrate its flexibility by promoting FAIR compliance at its core, by also aligning with well-known open information systems such as OSF and existing infrastructures, software, and tools that span the entire research data lifecycle, including scholarly communication: Recruiting and conducting tools (e.g., PsychoPy, z-Tree, z-Tree unleashed, oTree, lab.js, classEx, Lioness, SoSci Survey), recruiting tools (e.g., hroot, ORSEE), the information platform for research data management forschungsdaten.info, workflows (e.g., Galaxy, YARD, CentraXX, CentraXX Med), archiving solutions (e.g., sport information portal SURF of the Federal Institute of Sport Science, Pavlovia, x-science.org, x-econ.org), scholarly knowledge solutions (Open Research Knowledge Graph, Databus). METHODS is supported by Elsevier BV and its AI / semantic technologies to identify research field or SDG category of datasets and publications according to internationally established thesauri.

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

METHODS is highly open to any form of organizational or substantive collaboration within the NFDI, with other consortia, and with national and international stakeholders (e.g., EOSC). The idea is to align the interests of all initiatives in terms of autonomy and financial resources with the interests of the NFDI community at large in terms of synergies, generic solutions, and factual collaboration.

Disciplinary vicinity can be found to BERD, KonsortSWD, and NFDI4Health. Transdisciplinary Vicinity can be found in NFDI4Phys. The speaker of NFDI4Phys is a participant of METHODS and several (co-)spokespersons are involved in existing NFDI initiatives, also in leading positions.

Structural proximity can be found in NFDI4DataScience, NFDI4Earth, NFDI4Ing, NFDI4Culture, NFDI4Chem, 2Link and DataPlant. Potential collaborations include, among others: PUNCH4NFDI on RDMO data management plans; NFDI4Ing, NFDI4Culture, NFDI4Chem on terminology services; NFDI4Ing on workflows; NFDI4DataScience on FAIR scholarly information.

Transdisciplinarity is a common vision we share with NFDI4Phys and foster together.
3 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.

Through the widely used services and tools constituents in METHODS, the FAIR Data Principles will be brought together automatically and in a low-threshold manner with both researchers and the entire NFDI, and connected to common low-effort FDM and archiving solutions.

As indicated above, we see two complementary paths to meet these requirements: (1) Forming an infrastructure, starting with the thematic focus on empirical research on human behavior in Sport Sciences, Economics, Social Sciences, and Medical Informatics, and expanding it from such a nucleus; or (2) Forming an infrastructure based on common methodologies being neutral to discipline cultures. Both have their advantages for NFDI, and both approaches are, as suggested above, necessary.

Up to now, a large fraction of research data are not addressed by any existing or proposed NFDI consortium, especially valuable data in the workspace. This is especially true for research data that exist in the local and often inaccessible Research Data System and which is far from being published in the "classical sense" but will need to be shared to implement Open Science and to propel Germany to a top destination for advanced data-driven research (see Figure 1). History of funding has shown that it would be wrong to expect that exclusively discipline-specific approaches will have a massive impact on other disciplines. Through its methodological focus, METHODS is able to address research data, researchers, and actors in the German research landscape which are not covered by any other existing or planned consortia. With its generic solutions, Canonical Workflows for Research (CWFR) and FAIR Digital Objects (FDOs), METHODS explicitly addresses data at all stages of the lifecycle, regardless of, e.g., their format, degree of structuredness or complexity, and its methodology will support the lifecycle optimally to enable a FAIR compliant data landscape. The support will start from planning an observation, survey, simulation, or experiment based on theoretical considerations and end with the registration of FDO’s in trustworthy repositories, including their metadata and Persistent Identifiers (PIDs). These can include sensitive data for which special protection and tracking mechanisms need to be implemented, which is facilitated effectively by the systematic application of the FDO approach.

---

3 Experts agree that the "classical publishing" procedure will not be the primary option for sharing and reusing data and software. Much shorter cycles are required to be at the cutting edge.
Methods will implement an open methodology-oriented approach which is flexible and open further communities. Hence, methods needs to adhere to strict neutrality with respect to the discipline-approaches and traditions and respect the diversity of community methods and ethical standards. The diversity of methodological approaches within methods implies a universe of heterogeneous ethical standards. Methods and its partners, having roots in high-quality data from randomized controlled trials (RCTs), respect the opposing views on deception (psychology vs. experimental economics), incentives for participants, dual-use issues in medicine, or anonymity in data publications in education research and qualitative, hermeneutic approaches.

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

Which cross-cutting topics do you consider relevant for your consortium and the NFDI as a whole?

- Technical infrastructure and concepts: Research Data Commons, (Meta)Data; Findability; Terminologies; Infrastructure/Interoperability/Interfaces; Quality Management and Assurance.
- Community (User) involvement: User-driven Development.
- Collaborative governance and general framework: Common Vision and Strategy; Governance and Sustainability; Internationalization.

To which of these cross-cutting themes could your consortium contribute and how?

Methods will contribute its expertise, concepts, and partners to the collaborative NFDI working groups for cross-cutting topics and develop generic solutions based on a division of labor, in particular: Topic 1: Research Data Commons; Topic 2: (Meta)Data, Findability; Topic 3: Terminologies; Topic 5: Infrastructure/Interoperability/Interfaces; Topic 6: Quality Management and Assurance; Topic 8: Ethical-legal Aspects (person-related); Topic 9: User-driven
Development; Topic 11: Common Vision and Strategy; Topic 12: Cultural Change; Topic 13: Governance and Sustainability; Topic 15: Internationalization. METHODS will closely coordinate its focus on cross-cutting topics with NFDI4Phys due to the overlap in content.

An integrated national data infrastructure is not equal to the set of all services offered by the variety of discipline/domain-based approaches. An integrated infrastructure needs to be based on a few key standards and components everyone adheres to. Successful large infrastructures such as the electric grid, the Internet and the Web are based on a few simple standards that not only have an integrative function but the capability of picking up momentum. We believe that such a common infrastructure will be needed to make NFDI a success. With FAIR DOs, such a core standard has been suggested, and it now needs to be implemented broadly to demonstrate its capabilities on the one hand and fine-tune its approach on the other hand. METHODS will achieve this by being open to including other infrastructure initiatives in the developments and evolving testbeds. This NFDI initiative will also confirm the leading role of German data professionals in global discussions and implementations.

NFDI needs to include all scientific disciplines and local research environments working with data-driven approaches. Up to now, NFDI funded consortia are driven by domain-competence and coherence. However, this only covers a limited range of disciplines and the overall impact on further communities could be limited as they need to address their challenges and remain widely within their traditions. To include the many other disciplines, one has to turn to a method-oriented approach that enables the easy integration of different solutions without requiring researchers to change their habits abruptly, but which offers the chance to adopt FAIR-compliant practices incrementally. For sports psychologists, for example, the relevant, meaningful community is not only their discipline "sports science," but also the methodological approaches of their discipline of origin, psychology.

It is agreed that large infrastructures such as NFDI and EOSC should put the FAIR principles into practice, which, according to their creators, needs to go beyond a sole focus on findability and accessibility. Anticipating the challenges of the coming decades, turning FAIR principles into practice actually means to be prepared to make all data science artifacts (data, code, metadata, do-files, software, etc.) interoperable and reusable in the sense that machines know what to do. This is an ambitious goal and can only be realized in steps. Currently, FAIR DO’s are the only known suggestion that makes it possible to implement such a landscape if openness is of relevance. Current stakeholders of METHODS have world-class expertise and competence that is invaluable to the NFDI as a whole. The NFDI needs the FDO-technology to not only be prepared for the coming challenges but to become one of its global drivers. FAIR DOs will be a core technology to connecting meaningful method-oriented communities across the disciplines. However, more important for researchers is the implementation of a cross-disciplinary workflow framework based on FDOs and other technologies as well as common metadata frameworks such as CERIF (the Common European Research Information Format), DDI / DDI-CDI (DDI - Cross Domain Integration), Dublin Core, etc. and globally available PID solutions.

Up to now, many universities have fairly general approaches to RDM and data silos that are not connected to specific structures. METHODS will implement generic toolboxes, and shops that can help to overcome this actual inefficient situation.