
NFDI4RSE:
National Research Data Infrastructure for Scientific Software
Non-binding letter of intent for the NFDI

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

Non-binding letter of intent (anticipated submission in 2020)

2. Formal details

Planned name of the consortium:

National Research Data Infrastructure for Scientific Software
(Nationale Forschungsinfrastruktur für wissenschaftliche Software)

Acronym of the planned consortium:

NFDI4RSE (formerly RSE4NFDI)

Applicant institution

Friedrich Schiller University Jena
Fürstengraben 1
07743 Jena

Spokesperson

Dr. Frank Löffler
Department for Mathematics and Computer Science
Friedrich Schiller University Jena
07743 Jena
nfdi4rse-orga@de-rse.org

3. Objectives, work programme and research environment

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

44 - Computer Science, Systems and Electrical Engineering

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

All scientific domains are currently facing a massive and conspicuous increase of observational data through increasingly digitised scientific workflows. Research software, with its key capabilities to operationalise, repeat, disseminate and document analytical procedures, has become one of the pillars of science in general. Usable and sustainable open software is a prerequisite for reproducible research. However, in contrast to scholarly communication and primary data sets, research software has long been considered to be a mere tool, or a second-class research product at most.

Software is often treated as “data”, i.e. a static digital object. While some aspects of software can indeed be handled this way, good practices for data fail to address numerous central properties of software and its development process: (1) Software requires maintenance in order to ensure its continuous usability in permanently changing hardware and software environments. (2) Due to necessary adaptations and continuous development, software usually evolves over time. (3) Such modifications will often - especially after the initial publication and successful adoption - come from external contributors. These desirable collaborations raise questions about acknowledgement, accountability and copyright. (4) While software projects have typically low requirements in terms of storage space, they critically depend on fine-grained change histories, version management, collaborative online platforms, and a scalable computational infrastructure for testing. (5) Most software builds on existing libraries or frameworks, which also change over time, thereby creating dependencies on a specific version. Therefore, long-term sustainability can only be achieved by taking these dependency graphs into account. (6) Software can be conveyed as source code or compiled binaries, leading to more complex licensing conditions. These interconnected - and often neglected - properties necessitate a distinct treatment of software to advance it to a truly sustainable and FAIR research product. While adequately addressing these needs will facilitate data-driven research at its full potential, a failure to do so will ultimately jeopardise fundamental scientific principles of transparency and reproducibility.

The challenges of sustainable research software have been recognised by a number of scientific groups and organisations, most prominently the constantly growing community of Research Software Engineers (RSEs), i.e. persons working at the interface of software development and science. Importantly, all of the problems stated above can be addressed on a technical and procedural level by existing solutions like distributed concurrent version systems, operationalised software development processes and long-term archiving repositories. There is however a clear need for a cultural change as well, i.e. the recognition of research software as a first-class research product and proper recognition of those developing and maintaining it. The acknowledgement must be supported by education of all stakeholders in science - across all career levels and roles - and an adequate infrastructure.

NFDI4RSE aims to lead and catalyse the changes described above and has therefore set itself the following objectives:

- Facilitate the efficient sharing, discovery and re-use of research software
- Safeguard the sustainability and FAIRness of research software
- Foster best practises of software engineering and maintenance

These objectives will be addressed in the following initial task areas:

- Host the NFDI Software Panel: A group of NFDI software experts, providing consulting to all NFDI consortia regarding all aspects of software sustainability, including best practices for development, technical infrastructure and legal questions.
- Foster software interoperability between NFDI consortia by recommending and providing feedback on data exchange standards and programmatic interfaces (APIs).
- Provide training and support for research software engineers (RSEs) and all scientists working with software (e.g. through Software Carpentry).
- Interlink NFDI internationally on topics of metadata standards, software registries and search engines for research software (as digital object), with RSE organizations and other stakeholders on European and international levels (e.g. Research Data Alliance, Force11, Software Sustainability Institute, Research Software Alliance).
- Create and promote NFDI guidelines for the development of software compliant with FAIR and Free/Open Source Software criteria.
- Build and maintain a federated software repository infrastructure complemented by overarching public platforms for discovery and testing. This will include acting as central negotiator for potential outsourcing of development to commercial entities.
- Identify and cooperate with existing repositories to implement solutions for publication, annotation, assignment of persistent identifiers and long-term archiving of selected versions of research software. This task area is complementary to 6., as its primary purpose is to review and archive, not to provide a development platform. This includes an offering to enhance scholarly communication with a review service for publications in regards to scientific computing aspects.
- General administration

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

NFDI4RSE has been in contact with the following consortia and discussed potential interfaces and collaborations:

- NFDI4_CS_4NFDI: NFDI4_CS_4NFDI and NFDI4RSE have one topic in common that is undeniably essential for the success of the NFDI as a whole: Software. Importantly, our two consortia have complementary views on the topic. While NFDI4_CS_4NFDI concentrates on research in computer science, NFDI4RSE focuses on the topic of software throughout all of the research landscape. The substantial overlap in common interests will inevitably lead to a close collaboration between the consortia, the nature of which we plan to discuss in detail over the summer.
- 2linkNFDI: With 2linkNFDI providing existing infrastructure that could be built upon and NFDI4RSE providing not only a substantial part of the user base of said infrastructure, but also experts that could guide these extensions, a strong link between those two initiatives would be beneficial for both. In fact, 2linkNFDI and NFDI4RSE are in close contact to organise meetings during the summer 2019 to discuss future collaboration options, explicitly including a potential fusion of consortia.
- NFDI4Ing: The engineering sciences have a similar mindset as RSEs, so it is not surprising that we share a lot of methods and data types. NFDI4Ing can benefit greatly from the network and platform NFDI4RSE provides to share experiences across disciplines and thus, also within the NFDI4Ing community. At the same time, it is that platform that would allow other members of the NFDI4RSE software panel to benefit from experiences within the engineering sciences.
- MaRDI: The field of mathematics uses research software to implement models, run simulations, and setup workflows using smaller tools. Similar interests follow along a research software cycle from creation, recombination, testing, execution and dissemination of research software. Naming one example: NFDI4RSE could learn from MaRDI the years-long experience running the software index swMath. But there are also common interests in training or (re-)using existing (common) infrastructures like code repositories.
- In addition to these subject-related consortia, NFDI4RSE has been in contact with numerous others. The following consortia explicitly recognise the importance of sustainable research software to meet their objectives and/or have expressed their interest in various services that NFDI4RSE plans to provide (with NFDI4RSE contact):
 - NFDI4AIRR (Christian Busse, DKFZ)
 - NFDI4Earth (Jan Philipp Dietrich, PIK)
 - NFDI4Health (Konrad Förstner, ZB Med)
 - NFDI4Life Umbrella (Konrad Förstner, ZB Med)
 - NFDI4Microbiota (Konrad Förstner, ZB Med)
 - DeBioData (Stephan Janosch, MPI-CBG)
 - CompeNDI (Frank Löffler, FSU Jena)

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.

- Fostering software literacy by providing training
- Policies and open standards concerning the entire software lifecycle
- Legal issues regarding software in science
- Reuse and adaptation of existing NFDI IT infrastructure
- Publication and citation of digital objects like software
- Meta-data handling for software (e.g., (semi) automated annotation, semantic search)
- Application of FAIR principles (to software and via software) for interoperability and transfer (e.g. policies, software, credit)
- Software as data
- NFDI governance

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

- Fostering software literacy by providing training: Workshops of “The Carpentries” (Software, Data and Library Carpentry), organising hackathons and learnathons; help to extend the Open Education Resources of the Carpentries to fill gaps.
- Policies and open standards concerning the entire software lifecycle: Use NFDI Software Panel to achieve common understanding, acceptance and international harmonisation.
- Legal issues regarding software in science: Provide experience via diverse examples that worked or did not (not legal advice), e.g. regarding licensing, copyright, liability.
- Meta-data handling for software (e.g., (semi-)automated annotation, semantic search): usage and promotion of DOIs, standard attributes and collaboration with libraries enabling findability and re-use
- Application of FAIR principles (to software and via software) for interoperability and transfer: Leverage the NFDI Software Panel for the provision of policies, support, and software sustainability, as well as the promotion of academic credit for research software
- Software as data: Provide the hub for the discussion when software can be treated as data and when it cannot, resulting in recommendations for good scientific practice regarding research software.

5. Annex

For each (co-)spokesperson listed above, please add a list of all persons and/or institutions with whom the (co-)spokesperson collaborated closely during the last three years.

The following list will be completed by a list of co-spokespersons and their close collaborations before August 16, 2019.

Frank Löffler

- Arizona State University, USA
- Georgia Institute of Technology, USA
- German Aerospace Center, German Remote Sensing Data Center, Oberpfaffenhofen, Germany
- German Aerospace Center, Institute of Data Science, Jena, Germany
- Jena University Hospital, Germany
- Louisiana State University, USA
- Max Planck Institute of Biochemistry, Germany
- Rochester Institute of Technology, USA
- University of Illinois at Urbana-Champaign, USA
- University of Jena, Germany
- University of Parma, Italy