# RSE4NFDI

Safeguarding software sustainability in the NFDI

## Formal details

**Planned title of the consortium**
Safeguarding software sustainability in the NFDI

**Acronym of the planned consortium**
RSE4NFDI

**Lead institution or facility**
Friedrich-Schiller-Universität Jena

**Name and work address of a contact person**
Frank Löffler
Fakultät für Mathematik und Informatik
Friedrich-Schiller-Universität Jena
07743 Jena
rse4nfdi-orga@de-rse.org

**Potential members and participants as well as participating infrastructures of the planned consortium**

- Cluster of Excellence Matters of Activity, Alexander Struck
- Deutsches Krebsforschungszentrum, Christian Busse
- Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Stephan Druskat
- Fraunhofer FOKUS, Ina Schieferdecker
- Friedrich-Schiller-Universität Jena, Frank Löffler
- Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum GFZ, Martin Hammitzsch
- Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Uwe Konrad
- Helmholtz-Zentrum für Umweltforschung GmbH - UFZ, Thomas Schnicke
- Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik, Anke Lüdeling
- IT-Zentrum “Computer and Media Service”, Humboldt-Universität zu Berlin, Malte Dreyer
- Max-Planck-Institute for the Science of Human History, Robert Forkel
- Potsdam Institut für Klimafolgenforschung (PIK), Jan Philipp Dietrich
- Servicezentrum eSciences der Universität Trier, Marina Lemaire
- Sustainable Lifecycle Management for Scientific Software (SuLMaSS) Project: Karlsruhe Institute of Technology (KIT) (Institut für Biomedizinische Technik; Axel Loewe; Institut für Angewandte Materialien; Michael Selzer; KIT-Bibliothek, Robert Ulrich; Steinbuch Centre for Computing, Felix Bach), Universität Freiburg (Institut für Experimentelle Kardiovaskuläre Medizin, Gunnar Seemann)
- The European Virus Bioinformatics Center, Manja Marz
- TU Dresden, Center for Information Services and High Performance Computing (ZIH), Andreas Knüpfer
- University Medicine Greifswald, Dagmar Waltemath
- University of Freiburg, de.NBI-epi, Björn Grüning
- University of Münster - Institute for Geoinformatics, Daniel Nüst, and Zentrum für Informationsverarbeitung, Raimund Vogl
- Visualization Research Center (VISUS) / Institute for Visualisation and Interactive Systems (VIS), University of Stuttgart, Guido Reina
- Weizenbaum-Institut für die vernetzte Gesellschaft, Stefan Ullrich

- Westsächsische Hochschule Zwickau, Nico Herbig
- ZB MED - Information Centre for Life Sciences, Konrad Förstner
- 52°North Initiative for Geospatial Open Source Software GmbH, Simon Jirka

**Participants in the NFDI conference**

- Frank Löffler, Friedrich-Schiller-Universität Jena, frank.loeffler@uni-jena.de
- Ina Schieferdecker, Fraunhofer FOKUS, ina.schieferdecker@fokus.fraunhofer.de
- Martin Hammitzsch, Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum GFZ, martin.hammitzsch@gfz-potsdam.de

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

Cross-cutting theme

**Participating research institutions**

See above

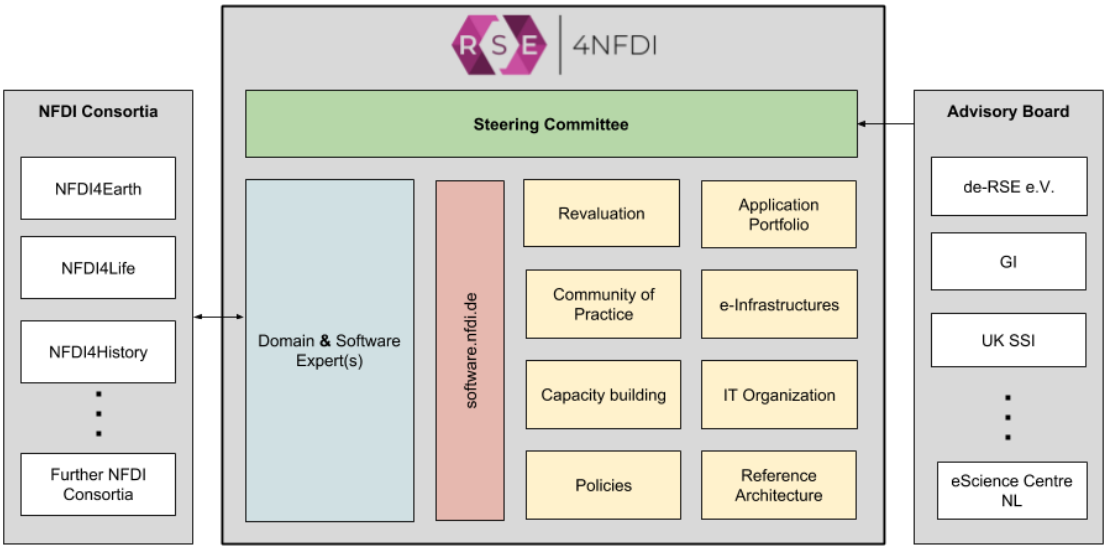
**Participating infrastructure facilities and/or potential information service providers**

See above

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

2019. Due to the cross-cutting nature of the consortium a close coordination in the starting phase with proposals and consortia in the cohort submitting 2019 is expected to increase impact of the planned activities.

**Overview diagram or organisational chart for the planned consortium**



# Subject-specific and infrastructural focus of the planned consortium

## Key questions/objectives of the consortium

**Software is critical to modern data-driven research.** Facilities, resources and services related to software are used ubiquitously by the academic science community to conduct research and foster innovation. As such research software is an indispensable part of the research infrastructure and underpins research endeavours, covering the whole range of data acquisition, data storage and data processing, data integration, data analysis, visualization, and modelling, from cloud-based collaborative word processing to high performance computing systems. Research software today supports science and engineering, social sciences, arts and humanities, and medicine.

Research software must meet the same rigorous requirements that researchers expect of their samples, equipment and infrastructures. One of these requirements is FAIRness (Findable, Accessible, Interoperable, and Re-usable). The **development of FAIR software** is only possible when technical and domain expertise are combined to harness the potential of computer science as well as library and information science. All of these aspects lead to widespread implications for sustainability of findings, trust and reproducibility of results, and the reuse of research outputs in academia and industry. However, scientists and research institutions today struggle to meet the increasing demands effectively. Also in the research community as a whole, measures and policies to address the crucial role of Research Software Engineering and software sustainability are still not adequately embedded in data-driven research.

Various initiatives, software development platforms, archival infrastructures, and the demand to better manage, understand, communicate, share, publish, and preserve research software and data are existing potential building blocks for a cross-cutting RSE4NFDI consortium. In Germany, these stem from funding agencies, e.g. the programme “Nachhaltigkeit von Forschungssoftware” by the German DFG[1], universities (e.g. local collaboration platforms such as GitLab[2]), research organizations (e.g., HGF HIFIS platform) and researchers themselves (e.g. as Carpentry Workshops instructors[3] or the FAIRDOMHUB[4]). Additionally, best practices for FAIR software development taking place outside of our scientific environment have to be evaluated. For example, the development of Spatial Data Infrastructures (SDI), Open Data Infrastructures, DARIAH or CLARIN established a set of principles and standards for the management of, e.g. spatial data, open data, and data from linguistics and the digital humanities on all levels of national and European authorities. Another example is the world-wide coordination board for the development of standards for computational modeling in biology, COMBINE[5].

[1] https://www.dfg.de/foerderung/info_wissenschaft/2016/info_wissenschaft_16_71/index.html
[2] E.g. gitlab.GWDG.de/explore, git.RWTH-Aachen.de/explore, gitlab.HZDR.de/explore and others
[3] https://carpentries.org/instructors-map/
[4] http://fairdomhub.org/
[5] http://co.mbine.org/

What is needed next is the elaboration of a common core set of software development principles, software components, services and enabling tools towards FAIR software. The **coordination and interoperation between the existing initiatives** needs to provide a seamless experience and common language to all those seeking to reuse research software, to improve their own developments, to acquire skills related to software, and to get help from experts for any of these needs. A key objective of the consortium is thus to **work towards a common understanding of the building blocks and skill sets supporting research software and tools**, across all participating disciplines. To this end, this NFDI initiative proposes to establish a research software community that jointly works **across disciplines covering an overall strategy and policies, reference architectures, research software components, tools and applications as well as support and education.**

## Known needs / current status

**Research data and research software form an inseparable unit.** A simultaneous build-up of infrastructures and expertise in both realms is the most pertinent change to advance science in a digitalised age. While both may technically be seen as digital artifacts, the range of required expertise to effectively collect, create and use them inevitably requires (a) teams of researchers, some of which must be Research Software Engineers (RSEs), and (b) research data and research software infrastructures. The **quality of data** is inextricable from the quality and documentation of processes linked to it, and these processes are built on and with software.

Identifying and constructing a national research infrastructure that supports software in and for research **leverages**

- **Integrating power of software** - Software can act as an integrator across research groups in the same discipline (intradisciplinarity), can lead to breakthroughs by transferring it from one domain to another (crossdisciplinarity), can be the manifestation of integrated knowledge from different disciplines (multidisciplinarity), can be improved by manifold perspectives and synthesise approaches from different user groups (interdisciplinary), and it can traverse or overcome disciplinary boundaries (transdisciplinarity).
- **Trusted research** - Reliable, well-tested software increases the trustworthiness of results.
- **Repeatable, reproducible and reusable research** - Software makes it easier for scientists to reuse their own workflows efficiently and transparently, others to repeat computations and reproduce results with the same data to verify and evaluate findings, and to leverage existing workflows with different data.
- **Extensible research** - FAIR research software, open workflows and free and open-source tools enable enhancement and extension for more effective and efficient science by adapting them to address further and different research questions.
- **Increased rate of discovery** - Reusable software is easy to adopt and extend. By building on existing software, researchers can invest more time into research and avoid wasting time by replicating software that already exists.

- **Increased return on investment** - Software can be very expensive to develop. Reusing software rather than replicating and a shared software infrastructure has the potential to save a significant amount of resources, which could then be invested into new research.
- **Safeguarded readability and improved usability of research data** - Data is meaningless without the software needed to read, process, and interpret it. Software that continues to function and is maintained according to current standards and paradigms allows continued access and reuse of research data.
- **Economies of scale** - Research data is always coupled with software for manipulation and interpretation, and providing a national research data infrastructure without a national research software infrastructure leads to a fragmentation in the software realm.

## Summary of the planned research data infrastructure

RSE4NFDI aims to safeguard software sustainability in the NFDI and across scientific disciplines on a national level. It addresses those in research who develop and use research software. The expertise, provided support and services will be complemented by the development and implementation of required and commonly accepted framing conditions. Taken together, the RSE4NFDI activities will ensure sustainable handling of research software supporting the FAIR principles.

For better use and the revaluation of scientific software, various complementary **activities** should be carried out in different areas. Pending further discussion at the NFDI conference, these activities may form the *areas of responsibility* in the nascent consortium. Some of these areas are very well suited to represent the *user perspective* in the consortium, namely community of practice, application portfolio, and revaluation.

- **e-Infrastructures** - Support of activities through the use of technologies, software components, tools, IT platforms, HPC systems, and Cloud services in the software development process, as well as support of the necessary human resources to exercise the software management related activities, such as collaboration platforms for software development, documentation, and testing (software.nfdi.de).
- **Application Portfolio** - Provisioning of applications based on specific requirements for the research software infrastructure, which ensure the further development and future viability of the scientific work
- **IT organization** - Provisioning, integration, networking and continuous adaptation of existing and new platforms, tools and services in a distributed, heterogeneous and differently mature system landscape in order to provide the operative basis for innovation-promoting, digital processes.
- **Reference Architecture** - Integration of complementary components of research infrastructures to a large extent, using blueprints and IT standards for data, software, interfaces and services as well as test sets to enable interoperable software extensions.
- **Policies** - Creation of framework conditions and establishment of processes through guidelines that cover an organisational model for an overall software sustainability infrastructure on top of existing building blocks, integrating existing services and commissioning new services required by the respective communities; creation of a governance model and delivery mechanisms for the services, legal and IPR questions as well as further issues regarding ownership and use.

- **Capacity Building** - Supporting those involved in developing software for research in acquiring and using the necessary skills by conducting training in basic concepts, tools, craftsmanship and best practices in software engineering, by teaching trainer skills, and by developing courses complementing - and in collaboration with - scientific disciplines and other NFDIs across varying technologies, levels of expertise, academic experience, and facilities.
- **Community of Practice** - Providing a forum for people who develop research software regardless of their job title and degree including scientists and academics, students, PhDs and postdocs, computer scientists, software developers, administrators and other stakeholders involved in research software development, regardless of the nature and extent of their activities and regardless of their experience or knowledge.
- **Revaluation** - Opening up of scientific software as a central building block in open science and improving the perception of the role of software in research, by integrating it into the scholarly value system, by undertaking knowledge and technology transfer, and by yielding them as assets in further activities to address new challenges.
- **Common Language** - Initiate an exchange between research software engineers involved in all NFDI consortia about needs and workflows to derive a common language that helps define the role and tasks of RSEs in science as well as research infrastructure development and maintenance, to enable efficient communication and collaboration between NFDIs, and to improve personnel development, career opportunities, as well as recruiting.
- **Long-Time Archiving and Software Heritage** - Over time the RSE4NFDI may contribute to build up a digital software heritage of German research as shown by the INRIA project in France.
- **Innovation** - Bring together software development and research requirements to develop innovative concepts which help researchers with their software development

## Description of data types and of underlying data processing / data analysis methodologies

All domain-specific NFDI consortia will use research software for the challenges of data processing and analysis methodologies. Unlike the analysis workflows, specific formats, and tools, the practices of software creation do not differ between consortia. A lack of coordination, which cannot be provided in passing by RSEs working in NFDI, has the danger of leading to disadvantageous parallel developments.

## Implementation of the FAIR principles

The use of FAIR principles in software can have implications that reach further than in data. The FAIR Data Principles play an important role to support Open Science and thus transferring their ideals to software is relevant in the context of all NFDI activities. The equivalent principles for software development and software preservation will be a core work area for RSE4FNDI.

- **Findable software** requires advances in software metadata and software citation, which will be contributed to and latest practices will be spread. We will evaluate suitable means to index software related to all NFDI consortia both as a showcase and platform for discovery.

- **Accessible software** from a developer perspective is closely linked to licensing, documentation, and archival. We will improve the understanding and application of Free and Open Source Software (F/OSS) across NFDIs, spread good practices for describing software, and improve software deposition in suitable repositories.
- **Interoperable software** is a huge challenge, as finding the right level of usefulness and abstraction for a tool is tough, especially in research settings where requirements are often evolving fast. The networking between developers and users will help to realise commonalities and potential for integration, e.g. synchronising recommendations of input/output data formats in respective domain NFDIs, and provide test methods and tools to safeguard the functionality and interoperability of software components.
- **Reusable software** needs researchers to follow community practices for developing, packaging and disseminating their work, e.g. using (language-specific) software repositories, which is greatly undervalued as a research output today. RSE4NFDI can foster the reusability with clear recommendations, education of both developers, funders, and evaluators, and with lobbying for long-term developments and arguing for the benefits of proper software project structures. Reusability goes both ways - researchers should reuse and contribute back more than develop new tools.

At first glance, it might seem that the FAIR Data Principles can simply be applied also to software. However, software is different in its nature. What drives the change in data is different from what drives the change in software. It covers aspects - such as usability, maintainability, maturity, and reuse rather than reinvention - that are not captured by FAIR yet.

## Planned measures for user participation and involvement

Across disciplines and other NFDI consortia, the RSE4NFDI consortium aims to establish a national focal point for researchers, developers, and early adopters in sciences who create or contribute to software development, e.g., by writing documentation, testing, or organising workshops. The RSE4NFDI partners - distributed over the national scientific landscape - will work in partnership with research communities and software communities (around tools, stacks, or languages) for advancing the improvement and impact of research software.

Several broadly defined and individual overlapping **user types** are relevant for RSE4NFDI. Elsewhere in Europe and the world, people who develop research software are already referred to as Research Software Engineers (RSEs), and are grouped under that term regardless of their job title and degree. RSE is a collective term for developers and scientists who contribute to software development, regardless of the nature and extent of their activities, and regardless of their experience or knowledge. The term addresses software-developing researchers, PhD students and postdocs as well as computer scientists and other computer science professionals, scripting administrators and software project managers active in research - and all those who professionally contribute to software development in science and research.

Following the motto “Better Software, Better Research”, coined by the SSI, UK, **the research productivity shall be increased by imparting software engineering skills, by offering support, and by connecting peers. Acquiring aptitude, using services, and benefiting from assistance enables researchers to focus on and conduct their research** due to the time gained by the professional handling of software.

**Workshops and consultative advice, services on demand, collaborative partnerships and long-term engagement as well as outreach, guidelines and policies** would be key areas of RSE4NFDI to support research communities, elicit feedback, and adjust services and outputs.

## Existing and intended degree of networking of the planned consortium

Networking with other infrastructures is a core activity of the consortium. RSE4NFDI meets requirements and needs and can advise other NFDI consortia on all matters related to research software and the people developing, maintaining and using research software.

The consortium has strong links and will further close collaborate with other national and international organisations and initiatives such as: Software Sustainability Institute (SSI, UK); US Research Software Sustainability Institute (URSSI, USA); eScience Center (Netherlands); UK Research Software Engineers Association (UKRSE, UK); Netherlands Research Software Engineer community (nl-RSE, NL); Australian and New Zealand Chapters of the RSE Association (RSE AU/NZ, AU and NZ); Gesellschaft für Informatik, Schwerpunktinitiative "Digitale Information", Helmholtz Arbeitskreis Open Science, Eclipse Foundation, Open Knowledge Foundation, Standardization Bodies like W3C, OASIS; Free Software Foundation Europe (FSFE). In order to benefit from the knowledge and experience coming from these initiatives, organisations, and activities in other countries, these groups and further individuals will be invited for participating in an advisory board integrating evaluation and strategic advice to RSE4NFDI.

## Additional information

**While we support the formation of a *German Software Sustainability Institute* (SSI) akin to the efforts in UK (SSI), USA (URSSI) and The Netherlands (eScience Centre), this consortium is not such an institute.** Focussing on developers, maintainers, and users of the NFDI is on the one hand a narrower focus than a German SSI would assume, and on the other hand goes much deeper by providing concrete guidelines for software in an NFDI context and by providing support for NFDI consortia - something that an SSI, if it existed, would not be able to provide to the same extent as RSE4NFDI. This does not limit the outreach of RSE4NFDI, which will contribute towards the establishment of a German SSI, e.g., by improving knowledge about and acknowledgement for software in research. The expertise of existing institutions will surely be leveraged (see above).

It is currently unclear how *existing NFDI consortia include research software engineering* (RSEng) skills and services in their proposals. It is reasonable to assume that some will, though given the numerous challenges, we expect none of the consortia to approach this topic at the same depth and breadth as RSE4NFDI can. **The intended degree of networking is crucial for the RSE4NFDI concept to flourish and we see the NFDI Conference as a crucial start of interactions with all consortia.** How the interleaving will take place is intentionally not fully developed yet. This should be a collaborative process with other NFDI consortia.