Information Infrastructures for Research Data

Frequently Asked Questions (FAQs)

The FAQs on the funding programme are continuously updated and are to be regarded as supplementary explanations to the programme guidelines[1]

Version date 07.04.2022

General Information

What are research data?

- Research data[2] are an essential basis for scientific work. The diversity of such data reflects the variety of different academic disciplines, epistemological interests and research methods. Research data includes measurement data, laboratory values, audiovisual information, texts, survey or observation data, methodological test procedures and questionnaires. Compilations and simulations can likewise constitute a key outcome of academic research and are therefore also included under the term research data.
- Research data in some subject areas is based on the analysis of objects (such as tissue, material, rock, water and soil samples, test specimens, installations, artefacts and art objects), so these must be handled with equal care.
- In the context of the funding programme “Information Infrastructures for Research Data”, the term “research data” does not include research software.

How does the DFG distinguish between infrastructure funding and research funding?

- Research funding is essentially aimed at generating new insights and knowledge. Infrastructure funding, on the other hand, is aimed at the (further) development of technologies, systems and tools with which scientifically relevant data can, for example, be collected, analysed, disseminated or secured in the long term.
- Infrastructure funding focuses on the establishment and further development of high-performance information infrastructures for science and the humanities.
- Infrastructure funding – which is what all LIS funding programmes are – does not cover basic research. For this reasons, the questions to be addressed in infrastructure projects should not be referred to as “research questions”. Under this programme, this also applies in particular to studies that are eligible for funding but do not fall under the basic research funded by the DFG. In the context of the programme, studies rather serve to clarify questions concerning research data management that are relevant to establishing and evolving research data infrastructures.
- Infrastructure funding cannot be used to advance academic qualifications.

[1] www.dfg.de/formulare/12_14/

[2] Further information from the DFG on the topic of research data
https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html

What are information infrastructures for research data?

- “Information infrastructures for research data” or “research data infrastructures” include both technical and organisational structures that enable science and the humanities to handle research data.

What is information infrastructure software?

- One aspect of the technical foundations of information infrastructures is information infrastructure software. This is predominantly based on a software stack. The individual hierarchical components of this stack are often reusable and established software (e.g. operating system, web server technology, runtime environment, programming languages, compilers, databases, frameworks, libraries and interfaces).
- The designation “information infrastructure software” clearly establishes the fact that information infrastructures on the provider and operational side are software systems whose development, implementation and functional further development can be funded under the funding programme.

What does vertical integration of a research data infrastructure mean?

- Vertical integration means the user-oriented, technical and, if necessary, organisational integration of an information infrastructure – starting at the researcher's workplace – in local structures and processes, at the regional (e.g. research data management state initiatives) and national level (e.g. National Research Data Infrastructure (NFDI)) through to the international level (e.g. European Open Science Cloud (EOSC)).
- Applicants are advised to engage with and network with national and international initiatives at an early stage.
- It is vital to strive for interoperability and the use of synergy effects, in particular with a view to achieving vertical integration but also in the interests of harmonising standards and organisational measures etc.

Questions on proposal submission/explanation of terms in connection with the programme

Can several applicants submit a joint proposal under this programme?

- The eligibility criteria are set out in section 2.1 of the programme guidelines. This makes it clear both academics and institutions are eligible to apply.
- A cooperative proposal submission process between infrastructure providers and users is explicitly recommended so that projects can be consistently geared to the benefits of researchers at an early stage.
- Individuals or institutions that are involved in, for example, NFDI, EOSC, specialised information services (FID) or another initiative can of course submit proposals under the “Information Infrastructures for Research Data” funding programme providing the formal requirements are met. Duplicate funding for the same project is not permitted.

What is a needs analysis?

- All projects must be demand-driven, i.e. they must meet the specific needs of relevant communities; these needs may exist in science and the humanities or at research infrastructure institutions. The needs analysis aims to determine these needs.
- The basis for the needs analysis can be the results of workshops, surveys, letters of support, etc. Existing information can be re-used.

What is an environment analysis?

- The environment analysis shows that there is not yet an adequate or sufficient solution available to meet the needs identified. This justifies the project proposal.
- With regard to the objectives of the project, the environment analysis shows which technical and organisational solutions can be used and how new solutions differ from existing ones.
- At the same time, the environment analysis shows the horizontal and vertical structures in which a project can or should be embedded.

What is a risk analysis?

- A risk analysis describes possible deviations from the project plan that may arise during implementation of the project. Furthermore, it sets out strategies on how risks can be minimised or how deviations from the plan can be dealt with appropriately.
- The risk analysis might concern personnel-related issues as well as technical, organisational, subject-related or other aspects.

Can direct project funding be applied for to cover legal advice?

- All legal issues that have a direct impact on the feasibility of the project must be clarified before the proposal is submitted and set out in the proposal.
- Legal issues that do not jeopardise the implementation of the project but are relevant to the handling of research data can be clarified and funded as part of the project itself.

Can investment funds be applied for to procure hardware?

- This is only possible if the project-specific need is convincingly presented. If permanent storage space is required, for example, this must be provided through Core Support.

Is it possible to obtain funding for the integration and curation of research data?

- Projects or work packages that focus solely on the integration of new data sets into an information infrastructure or solely on the curation of data without any direct link to the (further) development of the research data infrastructure are excluded from funding.

Design of the proposal or project

How can the range of services and functions of a research data infrastructure be designed?

- The design of the range of services and functions must always be geared to research needs. The following examples illustrate various aspects within the range of services and functions and can serve as a guide when planning a research data infrastructure. The list is not exhaustive.
  - **Information and planning** for the management of research data before, during and after research projects (e.g. planning of data management, recording of the data life cycle, compliance with funding guidelines, information services, etc.)
  - **Organisation and preparation** of research data (e.g. definition of a data curation profile, rights and access management, etc.)
  - **Description and documentation** of research data by means of structured information (e.g. use of metadata and metadata standards, controlled vocabularies, norm data, ontologies, etc.)
  - **Storage and connectivity** of research data for analysis (e.g. choice and set-up of storage locations, authentication and authorisation infrastructure, strategies and measures for data security and data backup, planning of interfaces with analysis and visualisation software, scientific computing, etc.)
  - **Publication and archiving** of research data (e.g. publication according to the FAIR, CARE or FACT principles in a suitable repository, use of persistent identifiers, measures for long-term archiving, choice of open and machine-readable file formats, etc.)

- **Findability and re-use** of research data (e.g. directories of repositories, metadata services and indices, data citation, etc.)
- **Rights and obligations** for handling research data (e.g. clarification and definition of copyright and data protection specifications for subsequent use, access options and publication of data as well as definition of usage licences, etc.)
- **Ethics and good research practice** (e.g. clarification of and adherence to ethics guidelines along the life cycle of data, development and implementation of disciplinary guidelines on the handling of research data, etc.)

How is the success of a prototype measured? How is the transition to a further development phase handled?

- The development of a prototype demonstrates the technical feasibility of a planned research data infrastructure. One aim of the project is to arrive at an assessment of whether the prototype can and should be developed into a reliable service and then operated.
- For this reason, proposals for prototypes must state the criteria according to which the suitability of the prototype can be determined. Criteria can be formulated for such aspects as technical functionality (scalability, etc.), benefit to users (usability: ease of use, accuracy of fit, etc.). These criteria can then be used to evaluate the prototype during the course of the project and at the end of the project; this is documented in the interim report or final report.
- The transition to a further development phase requires a renewal proposal accompanied by an interim report on the evaluation of the prototype. If the prototype and renewal proposal are found to be convincing in the review process, further development can be funded.

What is an implementation?

- In the context of this funding programme, an implementation is understood to be the conversion of a prototype or software design into a reliable service.

What is a functional further development of an information infrastructure?

- Functional further development is understood to mean the expansion of the range of functions and services of an existing information infrastructure. Further developments are geared to the changing needs of science and the humanities.

What is a purely technical enhancement of an information infrastructure?

- A purely technical enhancement is understood to be the integration of new or updated hardware or software components (for example software updates). This falls under the maintenance of an information infrastructure and is usually the responsibility of the operator. Purely technical enhancement is not eligible for funding.

What are digital services in the field of science and the humanities?

- Digital services include software and its organisational environment for academic use. The use of a service does not have to involve costs.
- Depending on the focus, digital services can be categorised[3] into:
  - Digital scientific service: Information technology service that provides such things as environments, tools and solution components for the scientific work of researchers and research groups.
  - Digital generic IT service: Infrastructure service for the purpose of identity and authorisation management and for the transfer, storage, processing, sharing, archiving and retrieval of data and information.

[3] cf. Digital services for science – where is the journey heading? DOI: 10.5281/zenodo.4301924, page 7

What are forms of organisation and networking?

- The organisation of an information infrastructure includes the design of all work processes that arise in the context of the development, expansion and operation of an information infrastructure for research data. When a form of organisation is conceived, responsibilities, rights and duties are defined relating to the design and use of the information infrastructure.
- Forms of networking involve the integration of the information infrastructure (to be developed) for research data in a specialist academic community and/or a developer or operator community. Different formats (conference papers, workshops, round tables, journal articles, public relations, etc.) can be used to promote the establishment of the information infrastructure.

What is an operating model?

- An operating model describes structures and processes that enable sustainable operation of the information infrastructure. Accordingly, regulations on the use of the information infrastructure can be made here (scope of use, for possibly different user groups, rights, obligations, possibly costs, data clearing house, etc.). Organisational procedures and responsibilities can also be defined (e.g. for applying for a storage quota, maintenance cycles, distribution of tasks between different institutions involved in the project, etc.).
- An operating model can also include a business model in which financial aspects such as user fees are regulated.
- The development of an operating model is eligible for funding under the “Information Infrastructures for Research Data” programme; it is important to note that where operating models are to be developed, all project results must also be made available for subsequent use under a free licence.

How can sustainability concepts be designed?

- A sustainability concept should be suitably adapted to the development phase of the information infrastructure. The more mature a project is, the greater the requirement to secure the long-term operation of the information infrastructure – technically, financially and organisationally.
    - The results of each development phase must be FAIR[4] (findable, accessible, interoperable, reusable). When developing a prototype, ensuring reusability may itself be recognised as a form of sustainability. It should be noted here that even in the case of a negative evaluation of a prototype, there is an obligation to secure and make available the project results in a well-documented form.
    - If a project leads to the operation of an information infrastructure for research data, the sustainability concept will focus on how continuous operation is to be organised. Here it is vital to provide a long-term concept that ensures sustainable operation of the information infrastructure. Different forms of organisation are conceivable. In addition to the responsibility of the applicant organisations, a transfer of responsibility to other institutions or organisations (such as NFDI consortia) is also conceivable, providing the latter make a binding commitment to organise continuous operation.
    - It is possible to seek to establish a developer community so as to ensure the continuous development of an information infrastructure.
    - Should it become apparent in the course of the project that a service will only be in operation for a foreseeable period of time, an end-of-life concept must be developed.
- In order to ensure sustainability, so-called lock-in effects, i.e. dependencies on commercial providers, are to be avoided.

What is a data curation profile?

- The aim of data curation is to compile coherent, reproducible and reusable data sets.

[4] The FAIR Guiding Principles for scientific data management and stewardship, DOI: 10.1038/sdata.2016.18

- The data curation profile describes the criteria for selecting data sets (e.g. quality, content, format, metadata, markup depth, etc.) and defines a description standard for these data (metadata schema, norm data, persistent identifier, markup format, etc.).

Integration of external services in the project

- If external services are required for the project, two comparative quotes must be submitted as part of the proposal.
- In principle, the recommendation is that the technical expertise for the establishment or further development of information infrastructures should be anchored within the applicants' institutions.
- If work is to be contracted out to third parties, please note that contracts for work and services are subject to approval by the DFG. See the funding guidelines for further details[5].

Licensing of data and software

- Licensing regulations for information infrastructure software are set out in the programme guidelines. In this context, open-source licences should be chosen that allow free subsequent use by third parties.
- Licensing of research data is recommended to be as free and open as possible.

After submitting a proposal

What criteria are used to assess proposals under this programme?

- The general review criteria for LIS programmes are applied. These are published on the DFG website[6].

How should the final report be structured?

- The final report is based on the Guidelines for Project Reports[7] in the field of "Scientific Library Services and Information Systems". Please address the issue of sustainability in particular.

[5] DFG funding guidelines https://www.dfg.de/formulare/2_00/

[6] http://www.dfg.de/formulare/12_107/

[7] https://www.dfg.de/formulare/12_02/