Originally published in: Zenodo (2023), pp. 26
DOI: *https://doi.org/10.5281/zenodo.7688729*

Nationale Forschungsdaten Infrastruktur

White Paper: Interim Report Reference

Verabschiedet durch die Konsortialversammlung des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V. am 13.01.2023

Publikationsserver des Leibniz-Instituts für Deutsche Sprache
URN: https://nbn-resolving.org/urn:nbn:de:bsz:mh39-121144

White Paper: Interim Report Reference

Lisa Amelung, Verena Anthofer, Renita Danabalan, Évariste Demandt, Barbara Ebert, Elisabeth Elschner, Sara Espinoza, Jan Eufinger, Stefanie Fuchsloch, Barbara Götz, Christin Henzen, Johannes Hunold, Theresa Idda, Lukas Jansen, Ulrich Krieger, Cristina M. Rodrigues, Maria Meister, Bernhard Miller, Sarah Pittroff, Christian Popp, Aleksandra Pushkina, Nicole Schatlowski, Sonja Schimmler, Christiane Schneide, Thomas Schörner, Thorsten Schwetje, Jörg Seegert, Thorsten Trippel, Judith Weber, Elena Wössner, Wolf Zinke

This White Paper sets out commonly agreed definitions on activities of consortia within NFDI. It aims to provide a common basis for reporting and reference regarding selected questions of cross-consortial relevance in DFG's template for the Interim Reports.[1] The questions were prioritised by an NFDI Task Force as a result of discussing possible answers to the DFG template.[2] In this process the need to agree on a generalizable meaning of terms commonly used in the context of NFDI, and reporting in particular, were identified from cross-consortial perspectives. Questions that showed the highest requirement on clarification are discussed in this White Paper. As NFDI evolves, the Task Force will likely propose further joint approaches for reporting in information infrastructures.
While each of broad relevance, the questions addressed relate to substantially different aspects of consortia's work. They are thus also structured slightly different.

[1] Instructions and Template for Consortia Progress Reports National Research Data Infrastructure (NFDI). https://www.dfg.de/formulare/nfdi140/nfdi140_en.pdf

[2] The Task Force is indebted to many colleagues from the various consortia contributing a host of helpful comments during the feedback-phase in November 2022 but also during earlier stages of this paper. As many comments were made anonymously we cannot address our gratitude individually.

Table of Contents

Introduction

As part of the NFDI, all funded consortia dedicated themselves to perpetuate cross-domain synergies in Research Data Management (RDM). This necessitates both broad and deep cooperation between the consortia. New links between experts and institutions have to be developed that often run orthogonal to the varied disciplinary and RDM-expertise backgrounds. NFDI's bottom-up design – infrastructure tailored to the needs of research communities – supports this process. As a result, we see the shaping of new connections and structures of cooperation that represent a substantial share of the short-term success of NFDI.

As such, participants in NFDI also have shared stakes in accurately reporting such developments. It is out of this spirit of collaboration that we, a NFDI Task Force on Evaluation an Reporting (TFER, formerly Task Force Monitoring), set out to delineate agreed-upon approaches to the reporting on four broad topics: Community Engagement, Collaborative Work in NFDI, Services, and the Degree of FAIRness. These topics were prioritised because we believe that they require the highest degree of coordination between the consortia as well as the highest requirement on clarification as scientific culture between consortia varies. As NFDI evolves further, TFER will likely propose additional joint approaches to reporting on information infrastructures.

Particularly regarding the Interim Report for consortia funded in the first round of NFDI due in September 2023, this White Paper is supposed to establish a common ground for publishing structured overviews of joint activities. It contains agreements upon definitions and delimitations on terms used within NFDI and RDM concepts. One goal for the White Paper is to act as a reference document that allows outsourcing these shared definitions and concepts from the Interim Report. The individual Interim Reports will then be able to focus on exemplary or aggregate perspectives.

Community Engagement

Relevant Questions from the Template

This chapter will provide a common basis to facilitate addressing the aspects of section 1.1 of the Interim Report Guidelines. Below follows an excerpt:

> *1.1 Composition of the consortium and its embedding in the community of interest*
>
> *(...)*
>
> *1.1.2 Integration of communities of interest, relevance for the research system*
>
> *By which procedures have* ***communities of interest been given an active role in the consortium****? How have the* ***needs of the communities been identified*** *and how has the consortium reacted to changing needs? Please* ***describe qualitatively,*** *and whenever possible also by means of* ***quantitative indicators,*** *which* ***benefits the consortium has generated*** *for the communities that it addresses as well as for other disciplines and the research system at large.*

Introduction

Methods, tools, data and objects as well as the culture of questioning differ within the broad NFDI research range. Historically, the ways of collaboration between individuals, groups and institutions developed in great varieties according to the multiple scientific needs. Today, we experience an intuitive global understanding of terms and cultures of collaboration. At the same time, we see particularly different definitions in concrete actions and tasks. This chapter provides definitions and examples to illustrate how all consortia and their corresponding research communities follow similar or very different ideas when talking about community engagement.

These aspects appear to be of essential importance:

1) Delimitation of the terms *Community* and *Participants*
2) Identification of and reaction to the needs of the community
3) Procedure to give communities an active role
4) Naming the benefits generated by the consortium

While benefits are going to be addressed by specific developments such as services (see also chapter on *Services*), the first three topics will be discussed in more detail as follows. However, prior to this, the common term *Community* needs to be defined for specific use in NFDI context and distinguished or merged where appropriate from NFDI specific other terms such as *Participants*.

Definitions

Delimitation of the terms Community and Participant

While *Communities* are represented by individual disciplines or methods of subject-oriented groups, *Participants* can be considered equivalent to structurally bound societies, associations, institutions and other legal entities. *Participant* as a term defines a role within a NFDI consortium whereas the term *Community* is used to qualify the scientific nature or peer group of which participants may be part of. In this context, all NFDI consortia are part of the research data management (RDM) community. On the other hand, the term *Community* addresses scientific disciplines and domains and in this sense every consortium is dedicated to at least one – oftentimes several – specific scientific communities of interest.

The boundaries of these domains can either be 1) discipline-oriented, 2) method-oriented or 3) data-oriented.

1) To elaborate for **discipline-oriented domains** the DFG subject classification system[3] can be consulted, but may have to be modified depending on the community and related discipline/s.
   For example NFDI4Ing represents the "Engineering Sciences" in this classification, which in turn can be broken down into the research areas ("Fachkollegium") 41-45, each both part of the overarching "Engineering Sciences Community" as well as being a subject-oriented community themselves. NFDI4Culture addresses research area 103 "Kunst-, Musik-, Theater- und Medienwissenschaften" (art history, musicology, theatre, film and media studies) as a whole and architecture out of research area 410.
2) **Method-oriented domains** describe communities that share specific scientific methods or a set of methods that define the way they interact with the object(s) of their investigation. This includes e.g. the community applying qualitative methods in the social sciences, or the physics researchers employing heavy-duty, expensive, large-scale instruments.
3) **Data-oriented domains** include all those who work with a specific type of data. Image data management for example differs from other types of data and is taken care of by RDM in a variety of disciplines like biology, astrophysics and art history.

Accordingly, the communities addressed by a consortium may be represented by all three orientations and may include several sub-communities. A multidisciplinary consortium is based on a group of *Participants* that represents a cross-section of these communities. A separation or categorisation of communities might not be necessary as long as it is ensured that every researcher's needs regarding RDM are addressed.

[3] DFG Classification of Scientific Disciplines, Research Areas, Review Boards and Subject Areas (2016-2019). https://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtsperiode_2016_2019/fachsystematik_2016-2019_en_grafik.pdf

Examples

Identification of needs

A prerequisite for a successful involvement within the community is the identification of their needs. Although some needs can be clearly defined by the community, others are less easy to verbalise and therefore cannot be queried effectively. For example, the problem of a missing RDM may not be seen until the solution is already available. Accordingly, the consultation cannot only be carried out via methods such as surveys or helpdesk requests, but must also be identified and iteratively adapted by the experts in the field (in this case, the consortium itself). To give an overview, the following is a(n) (incomplete) list of methods that serve to identify community needs:

- Surveys
- Analysis of Helpdesk requests
- On demand: direct contact and/or interviews
- Delegated personnel (e.g. Data Stewards)
- Community presentations at events (e.g. forums and conferences)
- Community workshops/meetings, focus groups or permanent user feedback groups
- Ambassador Programs
- Community Boards (= expert panel of representatives from the respective specialist group)
- Exhibition booths at trade fairs, conferences, congresses
- Social media campaigns

Active (and passive) role in the community and its benefits

An explicit and serious interaction between communities and the consortia representing them is essential to the success of the NFDI as a whole. Ideally, collaboration is founded on the basis of a Letter of Commitment, of Support, or of Intent, a Memorandum of Understanding, or similar documents, unless a connection already exists by nature of the consortium organisation. This includes, for example, all consortium partners who have signed a funding agreement.

Some variants of the cooperation between communities and consortium in practice are listed below:

Community as governance

- Integration of the Community in Governance: permanent members are sent to a dedicated Board and operate as important multiplier in the community of interest
- Implementation of a Community Board: expert panel of representatives of the respective professional group

Community as participatory network

- Provision, and joint creation of data, code, tools and services
- Participation and joint organisation in recurrent events such as community forums, working groups, workshops/meetings, etc.
- Participation in Special Interest Groups: temporary, topic-related discussion platform and expert pool, Hackathon, etc.
- Participation of the *Community* (i.e. represented by individual community members) in internal meetings or services provided by the consortium, e.g. as subject matter experts in distributed helpdesks
- Funding via FlexFunds/SeedFunds in exchange for work package processing or best practices
- Collaborative work on proposals

Community as recipient

- Audience of interest
- Taking part in recurrent community events, like workshops and informative meetings, etc.
- Receive consulting in meetings or via helpdesk requests
- Usage of tools and services provided by the consortium
- Assistance by Data Stewards (in both directions)

The following table shows three approaches towards interacting with communities: *community engagement*, *community activation*, and *furthering interconnectivity in and between communities*. *Community engagement* in this context means establishing two-way communication between the consortia and their respective communities. *Community activation* subsumes activities aimed at empowering communities to self-regulate and further develop their RDM practices. Activities using the *furthering interconnectivity* approach aim at leveraging existing networks from the institutional level to the level of consortia and beyond. The table outlines some objectives relevant to each approach in a bit more detail, and includes an exemplary list of specific activities to achieve these objectives that are either planned or already in use in various consortia.

Table 1: Approaches, goals/benefits and related actions/measure/activities

Approach	Goal or benefit	Action, measure or activity
Community engagement	Keep the community informed about the NFDI and the services offered by the consortium/consortia	Single-Point-of-Information, common platform for services, Data Stewards
	Create awareness for the FAIR criteria and RDM in all phases of the data lifecycle	Events, Workshops, Trainings, Helpdesks and F2F consulting, Data Stewards
	Determine (i.e. by polling or surveying) the communities' requirements and needs regarding RDM	Surveys, integrating community representatives into governing bodies, calls for Seed/Flex Funds, Helpdesks and F2F counselling, user-centred design approaches

Approach	Goal or benefit	Action, measure or activity
Community activation	Prepare and implement processes to support standardisation	Publishing of guidelines and White Papers, making standards openly available, survey of community acceptance
	Teach and develop RDM skills	Creating training material and supporting RDM coordinators at the respective institutions, Data Stewards; Organization of workshops and training activities
	Incentivise and enable the community to participate in the (further) development of services	Hosting events (e.g. Hackathons), being as open as possible regarding project management and publishing of results (OpenSource, OpenAccess), calls for Seed/Flex Funds
	Incentivise and enable the community to use RDM services	Common platform for services, Data Stewards
	Get feedback on services and give the community the opportunity to help shape the service landscape through review (and usage behaviour)	Surveys, evaluation, accompanying research (particularly from social sciences), integration of new participants (e.g. via Memorandum of Understanding or Letter(s) of Intent/Support/Commitment)
Further interconnectivity	Cooperation with specialised bodies, associations, initiatives or research projects to exploit synergies and avoid duplicate work	Integration of external experts in scientific advisory boards, establishing "community boards", joint events
	Synchronisation with other consortia with overlapping communities	Participating in NFDI sections, "Special Interest Groups", Data Stewards
	Encourage and expand exchange within the communities	Providing areas for intensive exchange within the community, e.g. forums, chat platforms, social media (e.g. #4CultureHour on Twitter[4])
	Network different communities to discover shared challenges and synergies (e.g. Text+ & NFDI4Ing or specialist groups)	Joint events, joint projects, SIGs, Data Stewards
	Engage industry and/or other social groups, e.g. in the cultural sector	Offer for openly accessible events, publication and scientific communication, target group adjusted surveys, regular exchange

Indicators for community engagement

While quantitative measurements can be easier to collect and compare, qualitative data can describe the variability and heterogeneous situation facing the consortia more accurately. Qualitative analysis within a community underlies many unsteady factors, like voluntariness, availability, time, etc. Still, some indicators can be:

- Number of community-contributed events
- Number of helpdesk requests

[4] #4CultureHour on Twitter. https://twitter.com/nfdi4culture/moments

- Number of applications/collaborations facilitated by a consortium/proposed in collaboration
- Publications (Data, FairDigitalObjects, Software, etc.)
- Surveys on quality of training and service, feedback forms
- Verified interconnectivity of existing systems and standards
- Number of social media followers and responses

There are indicators that seem only limited to address community engagement when compared between different consortia.

- Number of participants: While some consortia operate on a highly specified level regarding communities of interest (e.g. PUNCH4NFDI) and thus collaborate with a low number of participants, certain consortia with a broad spectrum of disciplines (e.g. NFDI4Culture) engage with a large number of participants.
- Number of software and service usages (training certificates, database usage numbers, etc.): The variety of disciplines addressed by NFDI consortia differ from huge scientific communities to small disciplines[5]. Usage numbers then only prove community size and are not able to indicate relevant engagement and must always be contextualised.

[5] Portal Kleine Fächer. https://www.kleinefaecher.de/

Collaborative work in NFDI

Relevant Questions from the Template

This chapter will elaborate on section 1.2 of the Interim Report template on "The consortium within the NFDI". This section elaborates on collaboration with related consortia, contribution to cross-cutting topics within the NFDI, participation in basic services and contribution to the NFDI governance structure.
In this chapter we propose to broaden the perspective from these four topics to collaborative work in NFDI as a whole. Whenever a consortium engages in the NFDI, this is by itself collaborative work. In addition, the primary mission of NFDI is bringing together the research data providers in Germany. Thus, collaborative work is a fundamental principle of work within the NFDI.

Introduction

A common approach to addressing questions in the Interim Report brings several advantages. The main one being the availability of a standardised assessment of measures and output through which consortia can clearly present their contribution within the NFDI. This would acknowledge that consortia are not only working to advance research data management, infrastructure and service development within their respective disciplines but also across the breadth of domains within the NFDI. Thus, demonstrating the relevance of the initiative as a whole.

Definition

Collaborations are defined here as the exchange of information on or development of common approaches to managing the research data of at least one domain. A necessary condition for any collaboration is that activities are *on behalf and in line* with the strategic aims of a consortium and are *not* activities by individuals within them only.

Collaborations can include consortia working within a single domain, cross domains.

1) Single Domain Collaborations: Collaborations with another consortium/institution/member of the NFDI e.V. within the same domain.
2) Cross Domain Collaborations: Collaborations with another consortium/institution/member of the NFDI e.V. on cross-cutting topics.

NFDI-wide cross-domain collaborations should be documented in one consolidated publication to be referenced by all consortia in their reports. The basis for this list are the attributes listed below.

Attributes

Both types of collaboration identified above can be described by a range of attributes which detail the collaboration. Some of these attributes are necessary to describe any collaboration, others provide additional information.

Collaborations by Domain-Coverage

The broadest collaboration within NFDI are the sections. In their working groups many cross-cutting topics are discussed by all or almost all consortia. This is most evident in the Sections *Common Infrastructures*[6] (135 members from 89 different NFDI member organisations) or the *Training and Education*[7] (105 members from 72 different NFDI member organisations) respectively. On the other hand, many consortia are currently establishing by which standards they wish to do RDM themselves. They might coordinate this across more than one domain creating a collaboration. On the other hand, single-domain collaborations will also make up a significant share of work in the NFDI context. Activities, like e.g. the BMBF-funded Competence Centres, can be part of either single-domain or cross-domain collaborations. Whenever they include partners outside of consortia, these should be provided.

Collaborations by Frequency and Duration

Frequency of Collaboration is closely linked to its output but can also illustrate the complexity of a given collaboration. When consortia, for example, work together to stimulate joint discussion in their fields on a topic to mutual interest (like a panel discussion at the 2022 JCDL conference) both frequency and duration will be lower than in the case of a Task Force. Some working groups in Task Force Tools, for example, had a high meeting frequency (monthly) but had achieved their desired results after, e.g., a year. Section working groups, like e.g. RDA working groups, work at different frequencies but most of them will exhibit a higher duration.

Collaboration by Output

Most importantly, collaborations are defined by their (intended) output. Consortia can collaboratively apply for a project grant aiming to fund some new service development. Sections will develop proposals for service development within Base4NFDI or propose White Papers for standards within NFDI. Likewise, two or more consortia will synchronise their agendas by agreeing on Memoranda of Understanding. Or a collaboration might aim to jointly organise a workshop. Publications across domains also are likely to be a common output of cross-domain collaboration.

Additional Attributes

In addition to these necessary attributes it will often be helpful to describe additional aspects, like if the NFDI Directorate was regularly part of the collaboration, whether the collaboration is institutionalised, i.e. is it part of a formal structure, if at least one partner of a collaboration

[6] Bericht der Section Common Infrastructures an die Konsortialversammlung für die Sitzung am 1.7.2022
[7] Bericht der Section Training & Education an die Konsortialversammlung für die Sitzung am 1.7.2022

is representing industry or other non-scientific domains, and where additional information might be available.

Table 2: Overview of attributes and suggestion for presentation (necessary attributes in bold)

Domain-Coverage	Number of consortia regularly involved and actively participating in collaboration	count
Frequency	Frequency of collaboration since start, predefined categories	▪ more than monthly ▪ monthly ▪ each quarter ▪ semi-annually ▪ annually
Duration	Start of non-interrupted collaboration (phase where activities took place with the frequency indicated)	date (yyyy-mm-dd)
Output	Type(s) of output planned or achieved, predefined categories, add if necessary	▪ Published Scientific Paper ▪ White Paper ▪ Memorandum of Understanding (MoU) ▪ Service ▪ Code ▪ Workshop / Conference ▪ NFDI Association (e.g. assuming responsibility in the Consortia Assembly) ▪ Standard Operation Procedures (SOP)
	Persistent identifier for output	e.g. doi
Involvement of Directorate	Was NFDI Directorate involved, predefined categories	▪ yes ▪ no
Institutionalisation	Is the collaboration institutionalised, i.e. is it part of a formal structure, and publicly (e.g. on web) recognizable as such?	▪ NFDI ▪ other (specify) ▪ no
Non-Scientific Domain Participation	Was at least one partner of a collaboration representing industry or other non-scientific domains?	▪ yes ▪ no
Further Information	Central source for more information on collaboration	e.g. doi, URL

The following table provides an example of how a joint, cross-consortia, document listing cross-consortial collaborations could look like. It should then be published as a separate document and later be referenced in the reports. It does not seem necessary that each consortium report on its within-domain activities in a similar fashion.

Table 3: Example of a document listing cross-consortial collaborations

Name	Domain-Coverage	Frequency	Started	Output	Involvement Directorate	Institutionalisation	Non-Scientific Domain participation
TF Tools	all consortia	monthly	2021-01-01	SOP	yes	no	no
Section *edutrain* (WP8 - Error Culture)	6 consortia	each quarter	2022-04-01	workshop	no	yes	no
Show & Tell: Social Media – Daten in der Forschungspraxis	4 consortia (4Culture, BERD KonsortSWD and Text+)	preparation: monthly Lecture and discussion series: 3	preparation: 2022-02-24 Lecture and discussion series:	3 lectures	no	no	no

Name	Domain-Coverage	Frequency	Started	Output	Involvement Directorate	Institution-alisation	Non-Scientific Domain participa-tion
		sessions	2022–05-13				
Interdisciplinary workflows	7 Consortia (4Ing, Mat-Werk, 4Cat, 4Chem, 4Culture, PUNCH and BERD)		2021-10-01	service (assess-ment of FAIRness in work-flow within	no	yes (availa-ble on the web)	no

Services

Relevant Questions from the Template

This chapter will elaborate on services. The definitions suggested here are meant as a reference to describe the consortial services as mentioned in

- question 2.4 in the Interim Report guidelines: "***Services provided by the consortium:*** *In describing the services currently provided by the consortium, distinguish clearly between services that consortium members provide as part of their institutional mission (Grundaufgaben) based on existing funding, and new services that have been established within the NFDI framework.*"
- or appendix section 1 in the progress report guidelines "*Please list the documents / **services** / procedures that you have produced and published briefly, but including at least the following data: title, year, and persistent identifier / web link.*"

Introduction

We have re-used existing definitions as much as possible (see references).

A special task to be solved for the DFG Interim Report is the **distinction between *Grundaufgaben (institutional tasks)* and *Projektaufgaben (project-based tasks).*** As can be seen from question 2.4 above, the DFG asks to distinguish contributions of the partners on the level of whole services. While it may be that new services (or service components) are being established within the NFDI, the funding may also serve to roll out an existing service component (funded by the institution) to a larger audience. In those cases, existing roles of partners, e.g. in providing institutional data archiving services, are extended to new projects. Clarifying such effects of the NFDI funding will be important to understand which funding and which provisions in terms of governance and mission will be needed to sustain NFDI services in the future.

Furthermore, **end-user services may be developed and provided collaboratively** by two or more consortium partners. In such cases, the actual service might be harder to describe, as the contributions of different partners have to be acknowledged and simply delineating different service components might not be sufficient. In such cases, existing models of roles and interactions in data service provisions may prove helpful, like those described in the reference architecture of the International Data Spaces Association (IDSA) and the German Council for Scientific Information Infrastructures - Rat für Informationsinfrastrukturen (RfII) position paper "Nutzung und Verwertung von Daten im wissenschaftlichen Raum". The roles described by RfII may provide further useful vocabulary as shown in the following table.

Table 4: Roles identified in an analysis of more than 40 scientific data services (Source: RfII[8])

Actors and their roles
▪ data subjects
▪ data producers
▪ data consumers
▪ sponsors/funders
▪ providers/operators
▪ technical service providers
▪ distributors and "brokers"
▪ curators

This chapter will provide an overview of roles in collaborative service provision and service categories based on the NFDI service definition. Taking the different roles in collaborative service provision into account, we first describe very broad categories into which services of NFDI consortia can be classified on a higher level. This is based on definitions from the German Council for Scientific Information Infrastructures - Rat für Informationsinfrastrukturen (RfII). These can help to classify the main mission of individual consortia. Afterwards a list of service components with accompanying definitions is given, inspired by the service categories defined in the German Bioinformatics network de.NBI. This should help consortia to group their individual services into a common scheme across NFDI. Naturally, given the diversity of the NFDI and RDM in general, the proposed categorisation is not exclusive and overlaps in the individual definitions are hard to avoid.

Definitions

General definition of a "NFDI service"

We propose to apply the **service definition used in the joint statement of NFDI consortia on basic services**[9] (Stellungnahme Basisdienste) wherever possible. Because the text has already been adopted by the consortia in early 2022, it can serve as a common - albeit very general - basis:

*"A **service** in NFDI is understood as a technical-organisational solution, which typically includes storage and computing services, software, processes, and workflows, as well as the necessary personnel support for different service desks."*

Roles in collaborative service provision

Service providers organise the space between data producers and data users, and several roles can be distinguished. These are often organised in a division of labour, but sometimes one organisation holds all roles. The role model is useful to get away from the question "which organisation does what?" and to analyse how reliable the different roles are fulfilled and

[8] RfII – Rat für Informationsinfrastrukturen: Nutzung und Verwertung von Daten im wissenschaftlichen Raum - Empfehlungen zur Ausgestaltung von Datendiensten an der Schnittstelle zwischen Wissenschaft und Wirtschaft (2021). Chapter 3.2: Akteure und ihre Rollen im Infrastrukturkontext, S. 44ff. (German only). https://rfii.de/?p=6961

[9] Konsortialversammlung des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V. (2022). Stellungnahme der NFDI-Konsortien zu Basisdiensten. Zenodo. https://doi.org/10.5281/zenodo.6091657

whether sustainability can perhaps be increased by finding another partner for certain roles. For roles involved in the provision of services, i.e. intermediary roles, consortium partners may have in making service components available, the following schema might prove useful.

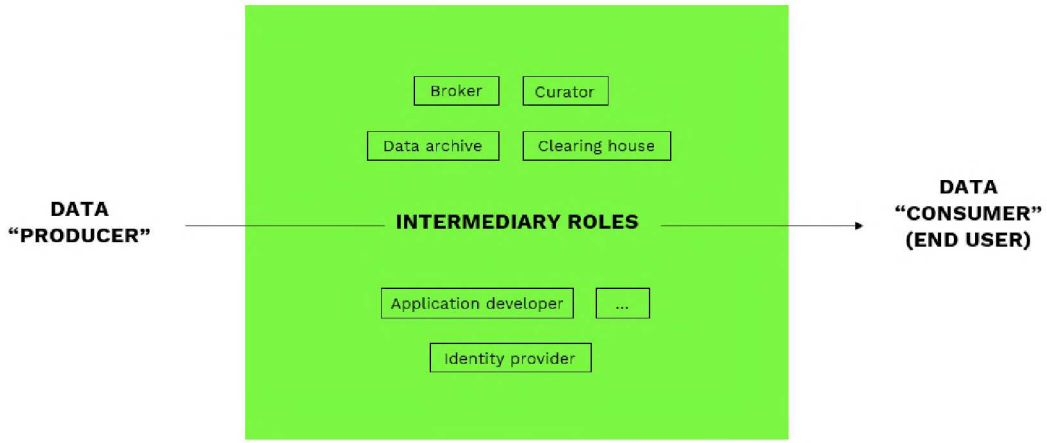

Figure 1: Simplified representation of intermediary roles; inspired by the International Data Spaces Reference Architecture Model.[10]

Categories of Services

Services can be categorised on different aspects. Here we provide an overview of definitions provided by the RfII.[11]

Data archives

The term *data archives* is used by the RfII to refer to services for data storage, publication, and discoverability. In most cases, the users themselves make their data accessible with metadata, and accordingly the licences issued for the individual digital objects can be heterogeneous. The basic scope of services appears largely standardised. Some providers offer special services, for which separate fees are charged in some cases.

Data platforms

Data platforms serve to exploit or market data in conjunction with analysis and visualisation tools or with infrastructure components such as computing and storage capacities. Users should be able to use the platforms as a tool for (their own) data processing. Data platforms are therefore similar to virtual research environments (VRE), which were also strongly supported by public money for a while. On the platforms, a data user can choose between different

[10] Otto, Boris, *et al.* International Data Spaces Reference Architecture Model, Version 3.0, Chapter 3.1 Business Layer, p. 20 ff. https://internationaldataspaces.org/download/16630/

[11] RfII – Rat für Informationsinfrastrukturen: Nutzung und Verwertung von Daten im wissenschaftlichen Raum – Empfehlungen zur Ausgestaltung von Datendiensten an der Schnittstelle zwischen Wissenschaft und Wirtschaft (2021). Chapter 3.2: Akteure und ihre Rollen im Infrastrukturkontext, S. 44ff. (German only). https://rfii.de/?p=6961

applications and gets access to computing and storage capacities. This makes the users independent of the technical equipment of the institution in which they work. The platforms also offer developers of applications the opportunity to exploit/market software.

Knowledge Bases

Knowledge bases aggregate data from different sources (literature, bibliographies, patents, data archives). Their main service to researchers is search and retrieval. Their contents are dynamic, i.e. they are constantly updated and supplemented. Depending on the technology used, the transitions between such information infrastructures and data platforms are fluid. Content is developed partly through manual curation and partly through the use of AI technologies. The services considered by the RfII, for example, make literature, patents, and data accessible for the life sciences and chemistry. They show some parallels to the systematically structured Specialised Information Services (Fachinformationsdienste, FID) in Germany. These emerged from the academic libraries' special collection areas, through which special literature was procured for participating subjects on the basis of a division of labour. With the introduction of FIDs, more and more digital resources have been procured and offered on a supra-regional basis since 2014, and growing amounts of different data have been acquired and made accessible for specific target groups. Acquisition and indexing are aligned with the needs of research.

Categories of Service Components

To describe the components of consortial services in a coherent way, we suggest starting with the **generic and more detailed service categories of the German Bioinformatics Network de.NBI**. de.NBI's service categories are generic enough to cover services across domains and can be extended, if necessary. Consortia can re-use this existing work also in terms of indicators: For each service category, the authors describe a set of key performance indicators and how to measure them[12]. In our revision, we have extended the original set of six de.NBI service categories by three more which seem relevant in the research data context. We have also renamed the categories as "service components", as the single items can be part of a larger service offering in NFDI, in line with the rather broad definition of "service" in NFDI (see above). Of course, de.NBI's categories may also be used in the original sense, i.e. to describe sets of services offered to end-users.

Last but not least, de.NBI's service categories can also be useful to structure answers regarding Grundaufgaben, as mentioned in question 2.4 of the Interim Report guidelines (see above). Larger services are sometimes organised in a collaborative way, with different components being provided by different partners.

Table 5: Items according to the list from de.NBI plus * = extensions to de.NBI list

Category	Joint understanding[13]
Databases	Software that provides large amounts of structured data from repositories and archives to the user. Usually the data can be uploaded, accessed, searched and/or downloaded via a web browser.
Libraries/API	Collection of pre-implemented functions for a specific task that can be accessed via a well-defined interface.

[12] Turewicz, Michael, *et al.* (2022). de.NBI service category-specific KPI selection and criteria. Zenodo. https://doi.org/10.5281/zenodo.6597826.

[13] All definitions in this column are cited from Turewicz, Michael *et al.*, unless noted otherwise.

Category	Joint understanding[13]
Workflows/ Pipelines	Software that combines multiple tools/applications. They may be used locally or remotely via the internet.
Tools/Applications	Software that can be downloaded and executed locally on the users' hardware.
Web applications	Software that is installed on a server and can be used by users via a web page and the internet, for example Software-as-a-service (SaaS).
Support/Consulting	Service with direct user contact for topics going beyond the support for specific single services.
*Data Curation	If not included in support/consulting: "The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials".[14]
*Training	Standalone training for self-study can be considered a technical service (usually a web application). Generally speaking, training materials often come in the form of specific measures or tutorials that are attached to a service and that are designed to improve the user's service experience. Our joint understanding of training as a stand-alone service, however, is not limited to the above and includes materials designed for education in all fields of research data management.[15]
*Storage	Provision of storage space for research data as a service to external users. Access is possible via web protocols.[16]

Examples for service components

Table 6: Examples for service components according to the list from de.NBI plus * = extensions to de.NBI list

Category	Examples
Databases	▪ Generic Repositories: RADAR ▪ Specific Repository: nmrXiv (NFDI4Chem) ▪ GFZ Data Services for data and software: https://bib.telegrafenberg.de/dataservices
Libraries/API	▪ B2FIND API for script-based metadata management: https://eudat.eu/services/userdoc/b2find-usage ▪ IÖR Monitor provides Open Geospatial Consortium access for data management: https://monitor.ioer.de/
Workflows/ Pipelines	Workflows listed in public repositories: ▪ https://snakemake.github.io/snakemake-workflow-catalog/ ▪ https://nf-co.re/pipelines ▪ https://gitlab.com/nfdi4culture/ta5-knowledge-graph/wikibase-deploy Other examples ▪ MOFA: Multi-Omics Factor Analysis: https://biofam.github.io/MOFA2/index.html ▪ PANGEO: https://medium.com/pangeo
Tools/Applications	▪ *Bioconductor-DESeq2:* https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html ▪ Swate workflow annotation tool for Excel: https://nfdi4plants.org/nfdi4plants.knowledge-base/docs/implementation/Swate.html
Web applications	▪ Online DMP Tools, e.g.: https://www.gfbio.org/dmpt or https://dmpg.nfdi4plants.org/ ▪ Diversity Workbench Arbeitsdatenbank als SaaS bei der GWDG ▪ Earth System Data Viewer: https://www.earthsystemdatalab.net/
Support/Consulting	▪ GFBio Helpdesk ▪ NFDI4Culture Legal Helpdesk ▪ NFDI4Earth User Support Network
*Data Curation	▪ PANGAEA curation services: https://wiki.pangaea.de/wiki/Curation_levels

[14] Definition from: CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020–2022 (v02_00-2020-2022). Zenodo. https://doi.org/10.5281/zenodo.3632563
[15] Own definition
[16] Own definition

Category	Examples
*Training	▪ GitLab repository for training and education materials run by NFDI4Ing partner organisations: https://git.rwth-aachen.de/nfdi4ing/education
*Storage	▪ de.NBI cloud https://www.denbi.de/cloud

Indicators to measure impact of services

Due to the diversity of the NFDI and the services provided by its consortia, a quantitative comparison based on indicators will mostly not lead to objective insights into the impact of the provided services. Still it may be useful to indicate selected indicators with each service to underline the reach of the services provided. If possible, these indicators should be put into context with other indicators such as the potential size of the community or by comparing to indicators of other established services.
For the listing of those selected indicators, we propose to base this on the Key Performance Indicators (KPI) developed by the de.NBI network[17] and to carefully assess whether and how those can be used to indicate the performance of individual services. KPI uptake can be selective, depending on the nature of a service. Regarding usage KPI, it has been suggested to differentiate between inhouse use of a service (i.e. by members of the providing institution) and external users (NFDI, national international). To assess acceptance, scope and potentially reputation of a service, KPI describing the user base would be helpful, but might be difficult to implement at this stage of development of the NFDI and cannot be mandatory for each service.

[17] Turewicz, Michael, *et al.* (2022). de.NBI service category-specific KPI selection and criteria. Zenodo. https://doi.org/10.5281/zenodo.6597826.

Degree of FAIRness

Relevant Questions from the Template

This chapter is intended to facilitate addressing question 2.3.1 of the Interim Report guidelines: "To what degree and in what way have you achieved implementation of the FAIR principles?"[18]. In the Interim Report guidelines, this question falls under the broader theme of question 2.3: "Implementation of the FAIR principles and data quality assurance". Our understanding is that implementing the FAIR principles, i.e., implementing Findability (F), Accessibility (A), Interoperability (I) and Re-Usability (R) of data, amounts to only one - albeit important - aspect of data quality. Other aspects of data quality not covered by the FAIR principles are, for example, curation or peer review of data and the protection of person-related information.

Introduction

The FAIR principles, hereafter referred to as FAIR data principles, were originally developed at a workshop in Leiden, Netherlands, in 2014 and have subsequently been published in 2016 by Wilkinson *et al.*[19]. They are guiding principles that aim at increasing the transparency and reproducibility of data and analytical workflows. The FAIR data principles are aspirational in nature and may be interpreted through the use of community-specific metrics[20]. As a matter of fact, more and more metrics are being published and are continually being refined (for a collection, see[21]), leading to a situation in which any one dataset or workflow is likely to score differently according to different metrics. Moreover, interpreting the metadata of any given dataset or workflow may involve guesswork in the face of ambiguities. This should make us cautious when using metrics for benchmarking or for obtaining binary judgements[22]. To reiterate the point, the FAIR data principles themselves are not a standard or specification [23], and there is no definite set of indicators that would facilitate implementation[24]. Notwithstanding the aforesaid, one metric, in particular, is referred to in the Interim Report guidelines[25].

Providing a common basis to facilitate addressing question 2.3.1 of the Interim Report guidelines matters since the FAIR data principles represent key aspects of research data management. They are very promising when it comes to weighing the costs and benefits of

[18] Instructions and Template for Consortia Progress Reports National Research Data Infrastructure (NFDI) [10/21]. https://www.dfg.de/formulare/nfdi140
[19] Wilkinson, Mark D., *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018. https://doi.org/10.1038/sdata.2016.18
[20] Bahim, Christophe, *et al.* (2020). The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. Data Science Journal, 19: 41, pp. 1–7. https://doi.org/10.5334/dsj-2020-041
[21] FAIRassist. https://www.fairassist.org
[22] European Commission, Directorate-General for Research and Innovation, Aronsen, Jan M., *et al.*, Recommendations on FAIR metrics for EOSC, Publications Office (2021). https://doi.org/10.2777/70791
[23] Mons, Barend, *et al.* (2017). Cloudy, increasingly FAIR. Revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use 37: 1, pp. 49–56. https://doi.org/10.3233/ISU-170824
[24] Devaraju, Anusuriya, *et al.* (2021). From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. Data Science Journal, 20: 4. https://doi.org/10.5334/dsj-2021-004
[25] The FAIR Data Maturity Model of the Research Data Alliance (RDA). cf. footnote no. 2 of the Interim Report guidelines. https://doi.org/10.15497/rda00050

implementation[26] and they figure prominently in various founding documents of the NFDI (e.g.,[27], [28]). Making research data FAIR, however, is not synonymous with all aspects of research data management and it is our first-hand experience that problems of demarcation arise. Commonly encountered are the following misunderstandings (for more, see[29]):

- Implementing the FAIR data principles does not necessarily assure the scientific quality of data (see above).
- Openness is not a requirement of the FAIR data principles; access is to be provided under well-defined conditions[30].
- Implementing the FAIR data principles is not a one-sided enterprise that is geared towards either humans or machines. Quite to the contrary, implementation of the FAIR data principles is only successful if it improves both human and machine intelligibility of data[31].

Providing a common basis to facilitate addressing question 2.3.1 of the Interim Report guidelines further matters in view of metrics that supposedly ascertain the degree of FAIRness achieved in individual consortia. Our assumption is that, on the one hand, there is no definite set of indicators. On the other hand, most consortia are not advanced enough in their endeavours to apply any one of the FAIR metrics to the majority of their data. This includes applying the FAIR Data Maturity Model referenced in the Interim Report guidelines[32]. Instead of recommending a published metric, we suggest a broader approach outlined in the following.

Definition

Starting point and degree of FAIRness

When applying to become a consortium within the NFDI, consortia have written a chapter on "implementation of the FAIR principles and data quality assurance", usually indexed as chapter 3.2 in their respective proposals. This chapter might be a useful point of reference for the status of FAIR data at the beginning of the project and may therefore be used in order to lay out a consortium's strategy and progress. Accordingly, an outline of how your consortium has been working to achieve implementation of the FAIR data principles and what has been done in order to monitor progress should be given in the Interim Report.

As mentioned in the introduction to this chapter, the field of FAIR data assessments is, as of today, highly dynamic and very much evolving. Major inroads into the development of

[26] European Commission, Directorate-General for Research and Innovation, Cost-benefit analysis for FAIR research data: cost of not having FAIR research data, Publications Office (2019). https://data.europa.eu/doi/10.2777/02999
[27] RfII – Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland (2016). https://d-nb.info/1104292440/34
[28] Bund-Länder-Vereinbarung zu Aufbau und Förderung einer nationalen Forschungsdateninfrastruktur (NFDI) (2018). https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf
[29] Moser, Mario (2022). Die FAIR Prinzipien: Quiz und Übersicht zum Einstieg. Zenodo. https://doi.org/10.5281/zenodo.6647047
[30] Mons, Barend, *et al.* (2017). Cloudy, increasingly FAIR. Revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use 37: 1, pp. 49–56. https://doi.org/10.3233/ISU-170824
[31] Wilkinson, Mark D., *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018. https://doi.org/10.1038/sdata.2016.18
[32] The FAIR Data Maturity Model of the Research Data Alliance (RDA). https://doi.org/10.15497/rda00050, cf. footnote no. 2 of the Interim Report guidelines.

indicators are being made by various organisations and a continuous supply of both new and refined metrics is being published. As stated earlier, the fact that one particular metric is referred to in the Interim Report guidelines does not lead us to recommending this or any other specific one. Instead, we propose three alternative ways of ascertaining a consortium's degree of FAIRness:

- In view of the starting point of each consortium and the degree of FAIRness already achieved as described, respectively, in the consortia's proposals, we propose to lay out a consortium's strategy and progress by virtue of reference to their starting point.
- Following the assumption that most consortia are not advanced enough in their endeavours to apply any one of the FAIR metrics to the majority of their data, we propose to establish use cases in applying published FAIR metrics to a minority of data and analytical workflows. These use cases should then be evaluated and probed for scalability.
- Following the tradition of spelling FAIR letter by letter and thereby addressing in turn Findability (F), Accessibility (A), Interoperability (I) and Re-Usability (R) of data, we provide a table of applied concepts or building blocks that are widely acknowledged across organisations, thereby further breaking down the degree of FAIRness (see Table, below).

It is our understanding that the aforementioned ways of ascertaining a consortium's degree of FAIRness lay the groundwork for compliance with the FAIR data principles and may well be combined.

General Framework (Synthesis of Indicators)

The following table contains a synthesis of applied concepts or building blocks that are widely acknowledged across organisations. The building blocks are further elaborated upon through means of some notes and wide-spread examples[33,34,35,36].

FINDABLE

Table 7: Findability building blocks

Findability building blocks	Notes	Examples
Metadata	Metadata is information (context) that describes an object such as a dataset as richly as possible. It can grow in time. It makes the data understandable for other users and easy to find for both computers and humans.	▪ Title ▪ Creator ▪ Year ▪ Provenance ▪ Copyright ▪ Target group ▪ Indexed in a searchable resource ▪ URL
Persistent Identifiers[37]	Persistent identifiers (PIDs) are globally unique,	▪ DOI

[33] Rocca-Serra, Philippe, *et al.* (2022) D2.1 FAIR Cookbook. Zenodo. https://doi.org/10.5281/zenodo.6783564
[34] *FAIR Principles:* https://www.go-fair.org/fair-principles/r.org
[35] *SATIFYD Self-Assessment Tool to Improve the FAIRness of Your Dataset.* https://satifyd.dans.knaw.nl/
[36] *ARDC FAIR Data Self Assessment Tool.* https://ardc.edu.au/resource/fair-data-self-assessment-tool/
[37] *What are persistent identifiers (PIDs)? ORCID. Support:* https://support.orcid.org/hc/en-us/articles/360006971013-What-are-persistent-identifiers-PIDs-

Findability building blocks	Notes	Examples
	long-lasting labels to an object. PIDs belong to the metadata and must be clearly stated. URLs are not PIDs as they can change over time, resulting in broken links to the object. PIDs are permanent, fixed, and citable.	▪ ORCID ▪ URN ▪ Handle
Standardised Terms	For metadata to be truly understandable and serve its purpose, a uniform terminology is required. Standardised terms make metadata, as well as the data itself, easier to understand for both machines and humans, and as a result, more usable for indexing. Standardised Terms make data easier to find.	▪ Controlled Terminology ▪ Ontology ▪ Taxonomy ▪ linked open data ▪ Semantic web
Registry / Repository	A data registry or repository is a searchable source which ensures findability on the Internet. An object must be registered or indexed in such a source in order to be findable.	▪ Zenodo ▪ PANGAEA ▪ Institutional Repositories
Additional information	Additional information helps users to assess the relevance, usability and content of the featured data and to get started using it.	▪ README files ▪ Wiki ▪ Data structure

ACCESSIBLE

Table 8: Accessibility building blocks

Accessibility building blocks	Notes	Examples
Protocol	The retrieval of metadata should be mediated via Standard Communication Protocols (SCP), a set of formal rules that describe how data can be transmitted or exchanged across a network.	▪ tcp ▪ HTTP ▪ SFTP
Metadata	Metadata accessibility should NOT be mediated by proprietary protocols, tools or special communication methods and should remain available even if the object no longer is. Metadata is accessible if it can be retrieved by machines and humans at any time.	
Access Rights	Accessibility does not inherently mean "open" or "free". Accessibility means that the exact conditions under which the data can be accessed are specified in such a way that a machine can automatically understand the requirements to access the object.	▪ Open Access ▪ Restricted Access ▪ Registration ▪ Read-Only

INTEROPERABLE

Table 9: Interoperability building blocks

Interoperability building blocks	Notes	Examples
Metadata	Metadata should be readable by machines. Custom translators, mappings, or algorithms are not required to ensure interoperability with applications for analysis, storage, and processing. Wherever possible, always provide machine-	

Interoperability building blocks	Notes	Examples
	readable cross-references. Metadata is machine-actionable.	
Standardised Terms	The standardised terms used in the metadata and to describe the corresponding object must be easily identifiable and accessible to anyone.	
Standardised Formats[38,39]	When selecting file formats, the formats need to be non-proprietary, unencrypted, and widely used in the research community. Only in this way, interoperability between different platforms or applications can be ensured. Selecting the right format also guarantees that the information can be read in the future.	▪ JSON ▪ CSV ▪ RTF ▪ HTML

RE-USABLE - THE ULTIMATE GOAL

Table 10: Re-usability building blocks

Re-usability building blocks	Notes	Examples
Metadata	Metadata should not only include the context in which the content in question was created, but also help the machine or human to determine whether the object is actually useful for a particular purpose.	▪ Rationale for creation ▪ Limitations ▪ Software ▪ Explanation of variables ▪ Digital file on conclusions
Provenance	Provenance should be included in the Metadata. It is a record documenting where an object originally originated confirming its authenticity. Provenance ensures trust, credibility and reproducibility.	▪ Origin ▪ SOPs ▪ Citations of reused data ▪ Processing history ▪ Workflow (machine readable)
Licensing[40,41]	A licence governs the scope of use or distribution of digital objects. The author/creator determines the type of licence. Without a licence, all rights remain with the author/creator and reuse by the research community is unfeasible.	▪ Creative commons ▪ Open Government Licence (OGL) ▪ Open licence
Domain Standards	It is more likely that other researchers will reuse objects if the metadata contains well-established and sustainable domain-specific standards. If there are community standards or best practices for archiving and sharing, they must be followed.	

Outlook

NFDI undertakes various collaborative efforts towards implementing the FAIR data principles while building a joint Infrastructure. For example, NFDI cooperates with Gaia-X in the FAIR Data Spaces project and creates a framework for a cloud-based data space for industry and research in compliance with the FAIR data principles[42]. Also, a lot of work is being done in so-

[38] *DANS File formats.* https://dans.knaw.nl/en/file-formats/
[39] Publications Office of the European Union: E-learning Module - Choosing the right format for open data. https://youtu.be/zTq1clni3z8
[40] *RDMkit Your tasks: Licensing.* https://rdmkit.elixir-europe.org/licensing.html
[41] Ball, A. (2014). How to License Research Data. DCC How-to Guides. Edinburgh. https://www.dcc.ac.uk/guidance/how-guides/license-research-data#x1-8000 How to License Research Data | DCC
[42] nfdi Nationale Forschungsdaten Infrastruktur. FAIR Data Spaces. https://www.nfdi.de/fair-data-spaces/

called sections[43], where consortia come together and tackle a number of critical points in order to advance implementation of the FAIR data principles. Within the section *Common Infrastructures*[44] for instance, work package A5 advances Findability through Persistent Identifiers (PIDs). In order to ensure a functioning internal organisation, technical components such as Identity Management, Data Integration and Long-Term Archiving are planned in working groups consisting of representatives of the consortia and experts, who contribute with their knowledge to ensure the access to, respectively the integration of, data and metadata. This will ensure data exchange between users and services and thus promote Accessibility. Of course, metadata *per se* promotes all FAIR data principles and is therefore the focus of the section Metadata & Terminologies[45].

Another important aspect of strengthening the FAIR data principles is to provide targeted training, thereby building the required skills and raising awareness in the context of research data management. Targeted training will be facilitated through the section *Training and Education* which also addresses the cultural practice of handling errors in science[46].

In summary, the collaborative efforts towards implementing the FAIR data principles at the level of NFDI are representative of a broad approach to FAIRness. They are not only undertaken in support of building a common infrastructure, but also aim at bringing about cultural change in the handling of research data.

[43] nfdi Nationale Forschungsdaten Infrastruktur. Sektionen. https://www.nfdi.de/sektionen
[44] Diepenbroek, Michael, *et al.* (2021). Sektionskonzept Common Infrastructures zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. Zenodo. https://doi.org/10.5281/zenodo.5607490
[45] Koepler, Oliver, *et al.* (2021). Sektionskonzept Meta(daten), Terminologien und Provenienz zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. Zenodo. https://doi.org/10.5281/zenodo.5619089
[46] Herres-Pawlis, Sonja, *et al.* (2022). Sektionskonzept Training & Education zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. (2.0). Zenodo. https://doi.org/10.5281/zenodo.6475541