NFDI4
MICROBIOTA

Foundation Guidelines

version 0.2

Contents

Motivation

NFDI4Microbiota acts as a central infrastructure hub for the national microbiology community by providing services for research data management, processing and publishing for scientists in the field. It relies in turn on already established infrastructure and services provided especially by de.NBI and NFDI4Microbiota associated partners for hosting and expertise in data processing, quality assessment and analysis.

These Guidelines shall act as basis for specific policies of all NFDI4Mircobiota projects and services. NFDI partners and other consortia are encouraged to use and adapt these guidelines according to their specific needs under the CC-BY license for their own purpose. NFDI4Microbiota associated partners aim to offer data handling and storage, bioinformatic tools and services as well as training within the scope of a centralized infrastructure to a broad scientific audience. In order to provide a consistent experience, permanent availability and long-term support following the FAIR principles and the Open science concepts, offered services and members of NFDI4Microbiota are supposed to adhere to the following guidelines, which are monitored regularly by TA leads and in Use Case reviews internally.

> *The following items are meant to be short and concise descriptions of community goals within NFDI4Microbiota that are communicated to partners and users (on the website). Expanded explanations in italics are meant as clarifications to NFDI4Microbiota members in help of fulfilling milestones and implementing services (not necessarily shown to end users or collapsed by default, to keep the guideline as brief as possible). Further details can additionally be found e.g. in the NFDI4Microbiota Knowledge Base.*

1 Availability

1.1 Providing software and documentation

Applications offered in the context of the NFDI4Microbiota are provided free of charge for download in case of offline tools. Online tools are provided for any researcher with reasonable quota of usage also free of charge. Links to the applications with a short description and accompanying documentation are available on www.nfdi4microbiota.de.

> *We aim to use and encourage the usage of free of charge tools to enable access for the broadest possible audience and anyone with a scientific interest. Pay-walls and license fees or unnecessary registrations may hinder the adoption and hence the re-usability considerably. Similarly, registration with a dedicated account may deter potential users. Instead services are encouraged to use existing Authentication and Authorization Infrastructure (AAI) such as LifeScience-RI.*

> Software that is designed to work on a local machine should be designed to work also without a network connection, where feasible and online-only tools should offer reasonable computation and storage quota for users to conduct meaningful analysis (not just example data).

1.2 Open source

Software in this context is developed under an OSI-approved license and the source code is published in openly accessible code repositories, along with appropriate documentation under an open license.

> From experience, tools developed in academia sometimes lack in the necessary documentation, are no longer maintained shortly after publication or not available at all after personnel left or institute's website links changed. In these cases, potentially useful tools become useless at best or even a frustrating waste of time for following generations of researchers. Openly hosted source code under a permissive license enables others to troubleshoot, adapt or even take over the development of a tool even in these cases. We recommend including the relatively concise MIT license. Publishing the actual source of a tool on a public code repository is necessary to prevent black-box results and preserve resources persistently. We recommend code hosting on github.com, gitlab.org, [gitlab.eudat.eu], [codeberg.org] or similar services for software projects. NFDI4Microbiota specific projects are hosted under the github.com/NFDI4Microbiota team account. Author should ideally also make use of the code deposition feature of Zenodo and make sure that the git repository is archived by Software Heritage. We also encourage to apply good research software development practices (including automated testing, semantic versioning, machine-readable software management plans, allocation of DOI to version, usage of Citation File Format (CFF) etc.).

1.3 Hosting

Online provided services are supposed to be accessible with minimal downtime. To assure the permanent availability, hosting containerized services in the de.NBI cloud infrastructure and re-using established high-availability services within the scientific community is encouraged.

> Even though we recognize that hosting web-tools on cloud services requires additional effort, we highly encourage this step instead of or additionally to hosting on local institute's machines to guarantee scalability for high demand of compute and storage resources and provide high availability. If done correctly, containerizing applications can prevent a lot of future work and headaches. Further de.NBI specific guidelines for hosted services apply.

1.4 Open access publishing

All types of scientific publications created in the context of NFDI4Microbiota are meant to be published in Open Access journals. In addition, we urge our users to use preprint servers like bioRxiv to make their findings openly accessible prior to publication. Researchers using the services of NFDI4Microbiota are highly encouraged to publish in Open Access journals as well.

> Research findings funded by public money should be made freely available in Open Access journals to maximize the potential audience and promote Open Science goals. Especially DFG funded projects have the obligation to publish their result under open licenses in Open Access and must contain a reference to the respective project number. In the case of publications enabled by NFDI4Microbiota this is "Deutsche Forschungsgemeinschaft (DFG) - Project number 460129525".

1.5 Open data

Data handled/produced in the context of NFDI4Microbiota must be made FAIR and openly available at the earliest possible time under a permissive license unless legal etc. restrictions prohibit a publication in commonly used repositories. This includes the raw data of the experiments and also the accompanying metadata in a commonly used metadata standard.

> Services that receive data should encourage the user to publish their data openly right away (e.g. by default) either within the service or recommended repositories. As a guide to sharing microbiome related data and analysis results please refer to Huttenhower et al. 2023. Where data is not made publicly available, data sharing with other users should be enabled with fine grained control over read/write and re-sharing permissions. Finally, all data and its accompanying metadata (following the minimal standards) needs to be deposited in the respective repositories latest upon publication of the research. An appropriate and compatible license (Creative Commons CC0 or CC-BY is recommended) needs to be supported by the service and chosen by the user. The data set should be properly referenced in the publications using the respective identifiers.

2 Support

2.1 Support contact

Each service requires a contact (e-mail address) deposited with NFDI4Microbiota for support questions via the help desk (e.g. e-mail support, online issue tracking).

Primarily, developers of a service are responsible for the maintenance, availability, documentation and supporting users with problems and handling feature requests. Since we aim to interlock the different services offered within NFDI4Mictobiota (e.g. in common workflows), we want to also offer central support to users who use NFDI4Microbiota as a platform. In order to address requests directly from the help desk or delegate issues to the respective maintainers, a contact address and link to the documentation is needed for each service.

2.2 Documentation in the Knowledge Base

Services maintain a short description with documentation on a sub-page in the NFDI4Microbiota Knowledge Base published under a permissive license.

As part of the effort to make a service available via the NFDI4Microbiota platform for better visibility, a short description (optimally with a screen shot) along with links to the service itself, documentation and code repository. This sub-page will also be used to summarize frequently asked support questions and answers by the maintainers and the help desk.

3 Interoperability

3.1 Compatible community standards

Software and services support widely used bioinformatics standard formats for input and output in order to make them compatible with other tools. Metadata formats adhere to widely used minimal standards and are required for any user uploaded data. Services and tools provided by NFDI4Microbiota which deal with metadata and relevant file formats support at least one of these standards as a default or recommendation for users.

While a myriad of file formats is generally available and it is easy to generate purpose-build output for your own applications, many are hard to comprehend and parse, not or barely specified and generally a hindrance to re-usability and interoperability. Additionally, neither metadata, nor provenance information is stored along the primary data more often than not. Therefore, services are supposed to adhere to widely-used and specified file formats for input and output. Summary statistics and logs with complete parameters are highly encouraged to enable provenance tracking. Web services which let users upload data should request metadata in a format compatible with at least one well-specified metadata standard by default.

3.2 Persistent Identifiers

Whenever possible, persistent identifiers are used within the services own databases and when refer-
ring/linking to external sources also in output of offline tools. The interoperability between offered
services that are relevant for each other is highly encouraged whenever possible.

> Databases have to make use of existing (persistent) identifiers when importing external resources
> (e.g. other databases). External identifiers should link back to the respective resource via perma-
> links or API. Additionally, databases may provide their own persistent identifiers, especially when
> cataloging novel entries. These identifiers should be exposed as permalinks or in an API.

4 (Re-)Usability

4.1 Terms of use

Services publicly state the terms of their use.

> Web services present their terms of use for visitors as part of their website to inform users about
> the conditions and restrictions of the service, specifying the following information: Who is offering
> the service under which license or jurisdiction and with which (limited) guarantees to which
> audience (in case of restricted access). Additionally, services that handle user data (login infor-
> mation or uploaded material) need to display a privacy policy (ger.: Datenschutzerklärung) at
> registration/upload specifying which data is stored where, the responsible data privacy office and
> terms for data retention, sharing and deletion in compliance with the law. Services may make
> use of generic Terms of Service and Privacy Policy offered by NFDI4Microbiota internally and
> customize them to their needs under legal advisement.

4.2 Accessibility

Services strive to be accessible by providing APIs (application programming interface) or supporting
machine-readable formats and support barrier-free accessibility.

> In order to maximize the usability of a given service, developers must think of different ways,
> user may interact with their interface, sometimes differently than intended. Instead of using a
> service just as a front-end to a database, developers should consider exposing an API or exporting
> machine-readable data dumps or releases that other developers can interact or make use of.
> Additionally, user interfaces should be designed with specific user constrains in mind, by offering
> alt-text for images (to be compatible with screen readers) and avoiding color-only encoding of

> information to accommodate red-green deficiency or color-blindness. Making a user interface rather static than dynamic (avoid hover-over menus) and navigable by using a logical tab order improves the usability of all potential users.

5 Appendix

Links to other consortia's policies and guidelines:

- **dataPLANT:** https://nfdi4plants.org/content/news/2021-11-08-dataplant-tools-and-services-development-principles.html
- **nfdi4health:** https://repository.publisso.de/resource/frl:6431645/data
- **NFDI4BIOIMAGE:** https://nfdi4bioimage.de/en/aims/
- **NMDC:** https://microbiomedata.org/data-management/
- **GHGA:** https://zenodo.org/record/6828131