1st Conference on Research Data Infrastructure

Poster presentation II

https://doi.org/10.52825/CoRDI.v1i.341

Published: 07 Sept. 2023

When Data Crosses Borders – Join Forces!

Multidisciplinary Use Cases Within NFDI

Barbara Ebert[1][https://orcid.org/0000-0003-3328-6693], Sami Domisch[2][https://orcid.org/0000-0002-8127-9335], Christin Henzen[3][https://orcid.org/0000-0002-5181-4368], Jimena Linares[1][https://orcid.org/0000-0003-3847-2663], Kati Mozygemba[4][https://orcid.org/0000-0002-0326-1607], Bernhard Miller[5][https://orcid.org/0000-0002-4385-7245], Bernhard Seeger[6][https://orcid.org/0000-0002-9362-153X], Jörg Seegert[3][https://orcid.org/0000-0001-9357-2830]

1 GF Bio e.V.

2 Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin

3 Technische Universität Dresden

4 Universität Bremen, FDZ Qualiservice

5 GESIS - Leibniz-Institute for the Social Sciences, Mannheim and Cologne

6 Philipps-Universität Marburg

Abstract. A multidisciplinary use-case for integrating research data involves data, software and methods across research disciplines to address research questions that cannot be effectively addressed by a single discipline or method alone. Multidisciplinary use-cases thus contribute to NFDI's goal to make a significant contribution to answering novel interdisciplinary research questions. This paper presents some selected examples of multidisciplinary use-cases from different NFDI domains that illustrate the efforts needed to reap the additional benefits of these use-cases. It concludes by proposing cornerstones for the research data management of multidisciplinary use-cases.

Keywords: NFDI, research data management, use-cases, collaboration, interdisciplinary research, multidisciplinary research data

1. The Relevance of Multidisciplinary Use-Cases

1.1 Substantive Relevance and Goals of NFDI

Multidisciplinary use-cases for integrating research data involve data, software and methods across research disciplines to address research questions that cannot be effectively addressed by a single discipline or method alone. These use-cases require collaboration across multiple research disciplines and likewise the integration of heterogeneous data, e.g., experimental, observational, and computational data, and management as well as processing workflows. They also require communication about a common understanding of concepts and methods (e.g., the understanding of "interview" can differ between disciplines and methodological approaches). Multidisciplinary use-cases thus help developing a more comprehensive and nuanced understanding of complex phenomena, by leveraging the complementary strengths across research disciplines and analytic approaches including software and workflows, as well as data types. This is reinforced in the German research system by the goals underpinning NFDI: "NFDI will also make a significant contribution to answering novel interdisciplinary research questions with a high societal impact." [1, p.2] This is strongly tied to the German Council of Information Infrastructures' consideration that "perspectives for the development of transformed or enriched data sets or "data products" should be conceived to create opportunities for wider re-use in other disciplines, fields, and domains." [2, p.3]

1.2 Topic and data-driven perspectives

In principle, there are two perspectives to identifying multidisciplinary use-cases: one is based on research topics, the other is data-driven. The first approach is driven by the desire to tackle complex questions with scientific rigor. This strengthens the need to link data from different methodological approaches and disciplines to solve complex research questions. Examples include estimating the socio-economic impacts of climate change and biodiversity loss or assessment of conservation efforts. Climate change's overarching relevance as well as its complexity have brought, among others, researchers from the Earth System Sciences (ESS), the life sciences and the social sciences to join forces and approach the same questions from different angles. To do so it is necessary to create ways for the various types of data to be combined and used for different forms of analysis.

Likewise, disaster prevention and risk management resp. and esp. its aspects of vulnerability, preparedness and resilience also require a multidisciplinary approach, e.g., by linking georeferenced survey data with existing crisis-relevant contextual information of various disciplines.

The second approach is not motivated by known substantive research questions but seeks to enable the identification of relevant patterns through enhancing the data. A data-driven approach as envisioned in [3] will thus emphasize the machine-actionability of data from different sources and this facilitates their interoperability. Researchers – and artificial intelligence – will then, for example, be able to look for patterns – such as health effects (measured e.g., through treatment data) in brightly lit areas (measured using satellite data).

To be sure, the first approach also benefits from richly annotated, machine-actionable metadata.

2. Benefits of Multidisciplinary Use-Cases and Examples

Data drives a better understanding of complex issues. Yet multidisciplinary use-cases generate particular benefits: 1) data can be contextualized in the knowledge of each participating field, which 2) given the likely relatively unconnected nature of research in the area offers new hypotheses that 3) can be tested by methods from each of the participating disciplines and

thus 4) will increase knowledge in likely more than one field. Multidisciplinary use cases therefore promise non-linear benefits as compared to single-discipline use-cases. The examples below illustrate that, in order to reap these benefits, some effort is required.

2.1 The *hydrographr* R-package

Freshwater ecosystems are relevant to both Earth Systems Sciences and especially to Hydro Sciences and - as a habitat - to biodiversity research. Data-wise they are uniquely characterized by their longitudinal, i.e., the up- and downstream connectivity between water bodies. Yet, this connectivity is largely neglected in spatial freshwater analyses on e.g., biodiversity. Since small streams to contribute by ca. 70% to the entire stream network length, a sufficiently high resolution is required resulting in large amounts of data. Any tool should therefore (i) efficiently harness the large amount of data, while (ii) allowing users to stay in their common environment, i.e., R.

The *hydrographr* R-package [4] was developed within a 4-month use case in NFDI4Biodiversity - on the basis of a pilot project of NFDI4Earth - that aims to lower the burden for potential users by offering easy-to-use functions within R. The package is tailored towards the high-resolution Hydrography90m [5] stream network data and capitalizes on fast and RAM-efficient open-source command-line GIS software (Geospatial Data Abstraction Library/ OpenGIS Simple Features Reference Implementation (GDAL/OGR), GRASS GIS [6] and AWK) within the Linux and Windows environments. Regardless of the operating system, users can hence create their workflows in R on their local computers employing high-resolution network data across large spatial extents, since the actual data processing is taking place outside R. Functions include wrappers to download, process, read and extract data, as well as to assess network distances and perform connectivity analyses. The package is hosted on GitHub and is available at https://github.com/glowabio/hydrographr.

2.2. FEdA BiodiWert Data

FEdA [7] was launched to fund interdisciplinary projects on the impact of nature conservation efforts in Germany. Projects are selected in annual calls. FEdA entered into cooperation on research data management with NFDI4Biodiversity. Workshops on FEdA's first call for proposals in FEdA's BiodiWert (valuation and preservation of biodiversity in politics, economy, and society) served to discuss data management plans (DMPs) and link up with domain-specific services. Most of the BiodiWert projects have multi-method research approaches, like the SLInBio project (Urban Lifestyles and the Valuation of Biodiversity – Dragonflies, Grasshoppers, Bumblebees & Co) [8] referring to biodiversity and environmental data (occurrence, molecular, ecotoxicological), geographical data, and empirical social research data (interviews, questionnaires, surveys, and documents), among other data.

To support this type of interdisciplinary project, a workflow was suggested across data centers from NFDI4Biodiversity and KonsortSWD using the Helpdesk service of the NFDI4Biodiversity's partner GFBio [9]. In the back office, experts and curators from GFBio's Data Centers [10] and experts from the KonsortSWD partners such as the Research Data Centers Qualiservice [11] and GESIS assist the creation of DMPs and prepare specific data types for archiving based on relevant community standards. Together with the FEdA coordination office, a FEdA data policy [12] was developed to create a guideline for the archiving and sustainable publication of research data.

The collaboration with the FEdA initiative constitutes a pilot project for consulting on multidisciplinary data management under NFDI's umbrella. For the partners involved, it required to create, re-think workflows, methodologies, and also enhanced the understanding of the different scientific practices in social sciences and biology.

3. The Way Ahead

Overall, the value of multidisciplinary use-cases for data sharing lies in their ability to facilitate collaboration between different scientific disciplines and methodological approaches. By leveraging heterogeneous data sources, researchers can gain new insights and make more accurate predictions about complex phenomena. Additionally, by supporting collaborations across different consortia, the NFDI can help break down barriers between research disciplines and promote a more holistic approach to science

In a first summary, we identified the following important questions for successfully detecting and processing multidisciplinary use cases across research data infrastructures.

1. How can we find meaningful multidisciplinary use-cases?
2. How can the needs of researchers, the various discipline-specific issues, and scientific methods be taken into account and integrated?
3. What are the ways and processes to systematically explore and exploit use-cases?
4. How can we encourage researchers working on multidisciplinary use-cases in the future?
5. What are the specific requirements in multidisciplinary use-cases? Are there limitations? Which obstacles have to be overcome for a good scientific service?
6. What challenges and opportunities arise from the consideration of discipline-specific or multidisciplinary use-cases for the development of (national) research data infrastructures?
7. How can multidisciplinary use-cases be stimulated or initiated?
8. How do we consider legal and ethical problems that occur by combining different data types?

Data availability statement

not applicable.

Underlying and related material

not applicable.

Author contributions

All authors have contributed to the **conceptualization** and the **writing – original draft**.

Competing interests

The authors declare that they have no competing interests.

Funding

This contribution is funded by grants from the German Research Foundation (DFG) within the framework of the agreement between the Federal Government and the Länder on the establishment and funding of the National Research Data Infrastructure (NFDI) of 26 November 2018:

KonsortSWD – grant no. 442494171

NFDI4Earth – grant no. 460036893

NFDI4Biodiversity – grant no. 442032008

Moreover, we acknowledge funding by the Leibniz Competition (J45/2018) to SD.

Acknowledgement

-

References

1. Gemeinsame Wissenschaftskonferenz von Bund und Ländern, Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018, November, 2018, https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf
2. RfII – German Council for Scientific Information Infrastructures, Discussion Paper on the Enhancement of Research Data Infrastructures, 2020, https://rfii.de/?p=4422
3. Tansley, S., & Tolle, K. M. (2009). The fourth paradigm: data-intensive scientific discovery (Vol. 1). A. J. Hey (Ed.). Redmond, WA: Microsoft research.
4. Üblacker, M. M., Grigoropoulou, A., Garcia Marquez, J., Torres Cambas, Y., Schürz, C., Floury, M., Tomiczek, T., Bremerich, V., Amatulli, G., & Domisch, S. (2023). hydrographr: Scalable Hydrographic Data Processing in R. R-package available at https://glowabio.github.io/hydrographr/.
5. Amatulli, G., Marquez, J., G., Sethi, T., Kiesel, J., Grigoropoulou, A., Üblacker, M. M., Longzhu, Q., S., Domisch, S. (2022). Hydrography90m: a new high-resolution global hydrographic dataset. Vol 14 (10), https://doi.org/10.5194/essd-14-4525-2022.
6. Neteler, M., Bowman, M., H., Landa, M., Metz, M. (2012). GRASS GIS: A multi-purpose open source GIS, Environmental Modelling & Software, Vol 31, pp. 124-130. https://www.sciencedirect.com/science/article/abs/pii/S1364815211002775
7. Linares, J., Ebert, B., Eberhardt, J., Frohne, K., Sauerland, K., Mozygemba, K., Miller, B., Collaborative Research Data Management Support for FEdA projects. Zenodo. https://doi.org/10.5281/zenodo.7624583.
8. SLInBio - Urban Lifestyles and the Valuation of Biodiversity – Dragonflies, Grasshoppers, Bumblebees & Co. FEdA - Biodiversity Initiative for the Conservation of Biodiversity. https://www.feda.bio/en/slinbio/ (accessed on 23.04.2023)
9. GFBio - German federation for Biological Data. https://www.gfbio.org/ (accessed on 23.04.2023)
10. Data Centers GFBio - German federation for Biological Data. https://www.gfbio.org/data-centers/. (accessed on 23.04.2023)
11. Qualiservice. https://www.qualiservice.org/en/. (accessed on 23.04.2023)
12. Taffner, J., Eberhardt, J., "FEdA Forschungsdaten-Policy" Zenodo. https://zenodo.org/record/7798305.