Building the Next Generation of Data Savvy Biomedical Researchers

Jens Dierkes[1][https://orcid.org/ 0000-0002-0121-9261], Birte Lindstädt [2][https://orcid.org/0000-0002-8251-1597], Ulrich Sax [3][https://orcid.org/0000-0002-8188-3495], Canan Hastik[3][https://orcid.org/0000-0003-1729-4642], Julia Fürst[2], Tanja Hörner[6][https://orcid.org/0000-0003-3280-6941], Sebastian Klammt[7][https://orcid.org/0000-0001-7852-4769], Ines Perrar[8][https://orcid.org/0000-0002-2830-6322], Iris Pigeot [9, 6][https://orcid.org/0000-0001-7483-0726], Katja Restel[1], Carsten Oliver Schmidt[10][https://orcid.org/0000-0001-5266-9396], Aliaksandra Shutsko[2][https://orcid.org/0000-0002-7091-5084], Dagmar Waltemath[10][https://orcid.org/0000-0002-5886-5563], Atinkut Zeleke[10][https://orcid.org/0000-0001-7838-9050]

[1] University and City Library Cologne, University of Cologne, Germany

[2] ZB MED Information Centre Life Sciences, Germany

[3] Technical University Darmstadt, Germany

[4] University Medical Center Göttingen, Germany

[6] University of Bremen, Germany

[7] KKS-Netzwerk e.V., Germany

[8] Institute for Nutrition and Food Sciences, Bonn University, Germany

[9] Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany

[10] Institute for Community Medicine, University Hospital Greifswald, Germany

Abstract. Modern research data management in biomedicine requires data literacy skills. The NFDI4Health consortium, the national research data infrastructure for personal health data, addresses this need with several different training concepts for clinical and epidemiological scientists. Both the institutional anchorage of the training and the thematic focus vary and can be assembled from a modular system as required. The aim is to enable multipliers to adapt and use the training modules through open educational resources. In addition, a competency profile for data stewards is under development to support the choice of required training in research institutions. The results will be fed into the NFDI "Education & Training" Section and will contribute to the Data Literacy Alliance in the future.

Keywords: Data Literacy, Training, Life Sciences, Health Informatics, Promoting RDM

Introduction

Researchers, but also allied professionals in the biomedical and health fields need the skills to manage research data in a planned way that ensures transparency, quality and compliance with regulatory requirements. To equip them with the appropriate understanding and the necessary skills in research data management (RDM) and data science, the National Research Data Infrastructure for Personal Health Data (NFDI4Health, [6]) offers a modular, reusable training programme that is tailored to different audiences based on the FAIR principles for data stewardship [12]. Offerings range from a university-based, certified courses with multiple tracks for Data Scientists and Data Stewards (DS), to single units on specific topics or NFDI4Health services.

The design and delivery of trainings is accompanied by a DS pilot programme that specifically explores the role of academic libraries working with discipline-specific research institutions as multipliers in delivering NFDI4Health services to users [10].

The concepts and materials are designed to be reusable and adaptable, and fit with plans to build a more sustainable FAIR data literacy (DL) training environment across all NFDI consortia with the NFDI "Training & Education" Section (EduTrain). NFDI4Health will contribute to the related Data Literacy Alliance (DALIA) infrastructure, which is responsible for the continuous offering, demand-oriented, methodical, and technical development of learning and teaching materials in the NFDI e.V.

Methods

NFDI4Health follows a multi-level approach, targeting different audiences within the clinical, epidemiological and public health sciences: (i) graduate students, (ii) researchers and (iii) professionals such as technicians, study nurses or librarians. This approach leads to three strands of action:

1. Cross-domain graduate training programme for doctoral students (Data Train)
2. Modular training offers based on dynamic learning paths concerning biomedical RDM basics and specific services/tools provided by NFDI4Health, accompanied by a knowledge base (NFDI4Health training handbook)
3. Concepts and materials developed published as open educational resources (OER), which are didactically enhanced to complement the RDM train-the-trainer programmes

Training needs are also informed by user research and immediate responses of participants via feedback forms. A key finding of the first survey was the need for basic RDM training and training on specific aspects of RDM. Initially, training will be developed along these two lines, with a focus on the basic biomedical RDM. The materials will be bundled into problem-based, contextualised and applied learning journeys.

To enable scaling up, the course materials and concepts will be made available as OER. These will allow for self-directed study. Furthermore, the provided materials can be adapted and reused in other contexts. As part of the DS pilot, a competency profile for DS in health will be developed to serve as a template for future DS education.

Results

Cross-domain graduate training: Data Train

The U Bremen Research Alliance [11], with the support of the Federal State of Bremen, has established the cross-institutional and cross-disciplinary training program "Data Train" [2, 5] for doctoral researchers from member institutions. Data Train pursues the mission of strengthening basic competencies in DL, RDM, and data science, while offering doctoral researchers a platform to build an interdisciplinary and interinstitutional network. The nine NFDI consortia represented in Bremen participate in the development and operation of the training courses. For example, members of NFDI4Health delivered a couple of data stories. In return, courses are opened for the NFDI-communities whenever possible.

Since 2021, the programme has gone through two passes with a total of 40 lecturers, 14 lectures, 17 workshops and data stories and 222 doctoral researchers participating. Since the programme was offered virtually, more than 2,600 participations were registered. Data Train will now be offered annually. It covers the entire data value chain and makes an important contribution to DL training for science as well as for the private sector [3].

Modular training offers

The NFDI4Health team has designed several RDM modules for researchers. These are regularly offered at conferences and in cooperation with individual research institutes (> 15 events). The courses cover RDM basics, biomedical aspects of RDM, and NFDI4Health-specific services.

The NFDI4Health digital services are being developed against a set of six representative use cases [7, 8]. Modules are designed, tested, and consolidated by DS of the pilot in close cooperation with the specialist researchers from the use cases. NFDI4Health will create a network of DS with different and complementary areas of expertise.

OER Material

A handbook detailing the above developments in training, the curricula and the main lessons learnt is in preparation.

All the underlying materials are designed as OER and will be published under a CC-By licence, initially in the Repository of Life Science [9] and Zenodo.

Outlook

To enable DS or multipliers to adapt and implement the trainings and to train new DS, NFDI4Health will publish a DS concept alongside the training handbook, based on the experiences of the DS pilot. These also feed into the work of EduTrain, which aims to sustainably improve DL in science across the consortia [4].

Once DALIA, a knowledge base for the use and provision of FAIR data [1], is up and running, the NFDI4Health OERs can be integrated into it. NFDI4health, DALIA, and the Section "Ethical, Legal, and Social Aspects" (ELSA) are connected via a privacy working package, which will enable to sort and offer OER along distinct DL careers with the future option of working towards certificated or other qualification artefacts (e.g. micro-credentials). NFDI4Health will also be part of DALIA's network of curators and experts.

Author contributions

JD: conceptualisation, supervision, project administration, writing – review & editing

BL: conceptualisation, supervision, writing – review & editing

US: conceptualisation, writing – review & editing

CH, JF, TH, SK, IP, IPe, KR, COS, AS, DW, AZ: writing – review & editing

Competing interests

The authors declare that they have no competing interests.

Funding

Acknowledgement

References

1. DALIA: Knowledge-Base für „FAIR data usage and supply" als Knowledge-Graph. https://www.fst.tu-darmstadt.de/forschung_fst/zusammenarbeit_in_der_forschung/dalia/dalia_ueberblick.de.jsp (26.4.2023)
2. Data Train - Training in Research Data Management and Data Science. https://www.bremen-research.de/data-train/ (26.4.2023)
3. Garbuglia, Federica, Saenen, Bregt, Gaillard, Vinciane, and Engelhardt, Claudia. (2021). D7.5 Good Practices in FAIR Competence Education (1.2). Zenodo. https://doi.org/10.5281/zenodo.6657165
4. Herres-Pawlis, Sonja, Pelz, Norbert Kockmann, Roger Gläser, Manuela Richter, Johannes Liermann, Jochen Ortmeyer, et al. "Sektionskonzept Training & Education zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.v.". Zenodo, April 21, 2022. https://doi.org/10.5281/zenodo.6475541; https://www.nfdi.de/section-edutrain/?lang=en
5. Hörner, Tanja, Frank Oliver Glöckner, Rolf Drechsler, und Iris Pigeot. 2021. „Disziplinübergreifendes Modell Zur Ausbildung Von Forschungsdatenmanagement Und Data Science Kompetenzen: ‚Data Train – Training in Research Data Management and Data Science'". *Bausteine Forschungsdatenmanagement*, Nr. 3 (Dezember). German:56-69. https://doi.org/10.17192/bfdm.2021.3.8343.
6. NFDI4Health. https://www.nfdi4health.de/en/ (26.4.2023)
7. NFDI4Health Task Areas. https://www.nfdi4health.de/en/about-us/task-areas.html (26.4.2023)
8. NFDI4Health TA5 „Use Cases". https://www.nfdi4health.de/en/about-us/task-areas/ta5-use-cases.html (26.4.2023)
9. Publisso - Repository for Life Sciences. https://www.publisso.de/en/publishing/repositories/repository-for-life-sciences (26.4.2023)
10. Shutsko, Aliaksandra, and Birte Lindstädt. 'Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten – NFDI4Health: Pilotprojekt zu Bibliotheken und Forschungsdatenkompetenzzentren als Multiplikatoren („Data Steward")'. *GMS Medizin - Bibliothek - Information* 20, no. 3 (22 December 2020): Doc27. https://doi.org/10.3205/mbi000484.
11. The University of Bremen Research Alliance. https://www.uni-bremen.de/research-alliance (26.4.2023)
12. Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (15 March 2016): 160018. https://doi.org/10.1038/sdata.2016.18.