1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.367

Published: 07 Sept. 2023

The Data Steward Service Center (DSSC)

FAIRagro RDM-expertise hub

Nikolai Svoboda[1][https://orcid.org/0000-0003-3860-4400], Lucia Vedder[2][https://orcid.org/0000-0002-8924-9800], Franziska Böhm[4], Markus Möller[6][https://orcid.org/0000-0002-1918-7747], Elena Rey-Mazón[5][https://orcid.org/0000-0003-4813-5927], Marcus Schmidt[1][https://orcid.org/0000-0002-5546-5521], Birte Lindstädt[3][https://orcid.org/0000-0002-8251-1597], and Ulrike Stahl[6][https://orcid.org/0000-0002-5659-910X]

[1] Leibniz Centre for Agricultural Landscape Research (ZALF) https://ror.org/01ygyzs83, Germany

[2] University of Bonn (UBN) https://ror.org/041nas322, Germany

[3] ZB MED Information Centre Life Sciences https://ror.org/0259fwx54, Germany

[4] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure (FIZ) https://ror.org/0387prb75, Germany

[5] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) https://ror.org/02skbsp27, Germany

[6] Julius Kühn Institute (JKI) - Federal Research Centre for Cultivated Plants https://ror.org/022d5qt08, Germany

Abstract. The Data Steward Service Center (DSSC) is the central institution within FAIRagro to develop data management tools based on the needs of the scientific community. The DSSC organizes the contin-uous exchange of RDM knowledge and experience with other institutions, channels user requests from the community, and transfers knowledge from the FAIRagro task areas to the FAIRagro data stewards. FAIRagro data stewards are experts in the field of RDM for agrosystems research super-vising and will train data curators in our community. Data stewards have core competencies in re-search data management (e.g., cross-scale from genes, phenomics, management to region; sensitive data, remote sensing, time series, plant, soil and related FAIRagro data). Knowledge and expertise is pooled to provide the full range of expertise to the community in one place to foster the coalescence of the community. The DSSC is headed by a coordinator and will house five data stewards, who are active in the community e.g. train data curators, give legal support. In the course of the project, further institutional or project data stewards will be integrated and the pool of experts will be further expanded. The network to the other NFDI consortia is continuously growing.

Keywords: Knowledge transfer, Help desk, DSSC, Data Stewards, Community, Network, RDM

1. FAIRagro

FAIRagro is a community-driven RDM initiative and focuses on the well-organized field of agricultural systems domain integrating important disciplines and scales needed to develop sustainable crop production and agroecosystems. FAIRagro provides researchers with FAIR and quality-assured RDM to generate, publish, and access relevant data, innovative RDM services, and modern data driven science methods to support and advance agrosystems research. FAIRagro is very well networked within the agricultural sciences and open towards NFDI and beyond.

1.1 The Data Steward Service Center (DSSC)

The data stewards networked in the DSSC are well trained in the area of general research data man-agement and have specialized skills to respond to the needs of agrosystem science. Specific RDM expertise is provided at DSSC by the FAIRagro partners: e.g. Guidance on long-term storage of plant breeding data, use of DMPs, implementation of Minimum Information About a Plant Phenotyp-ing Experiment (MIAPPE) (partner: IPK Gatersleben); FAIR data publication, data repository, stand-ardized metadata, DOI, long-term archiving, data acquisition (at ZALF, the FAIRagro coordinating institution); intellectual property rights in distributed information infrastructures, copyright, licensing, data protection law and IT security, legal advice (partner: FIZ Karlsruhe), curation of spatial data in spatial data infrastructures (SDI) considering quality aspects and provenance, enabling SDI interop-erability (partner: JKI Quedlinburg), management of large amounts of heterogeneous data originating from field robots and drones (Farming 4.0), knowledge transfer, data visualization (partner: University of Bonn).

The work of data stewards is mainly to support scientists as authors in RDM and especially in using the FAIRagro infrastructure and publishing data in qualified, domain-specific repositories.

1.2 Connecting the DSSC to the world: Our Online-Help Desk on the FAIRagro Portal

FAIRagro offers first level support consisting of training and information materials available via the FAIRagro Portal. The data stewards work in the area of second level support, easily reached via the help desk by several means such as e-mail, an online form, an implemented semi-automated chat-application or by phone. A third level support consists of direct contact with the developers in FAIRagro for very specific demands. Aim is to offer services quickly, easily accessible, in person and with a high degree of availability.

In addition, data stewards are temporarily assigned as local supporters in the community according to the principle of "Book A Data Steward" based on their skills and the need in the six FAIRagro use cases to implement professional data management from the beginning and act as trainers for the community.

1.3 Insight into the agricultural science community

FAIRagro and the DSSC largely build on the needs of the agricultural science community. As early as 2020, a large-scale survey was conducted to identify RDM needs and wishes of the target group. The results have been published [2] and scientifically evaluated [3]. They are the foundation of actions implemented in the DSSC. The results of an additional survey in 2023 are comparable [1], e.g. the urgent desire for a low-threshold, competent personal support.

Affiliation with DFG subject groups [4] reflects the expected diversity of agrosystems sciences, with most participants in soil science (23%), plant breeding and crop production (18% each), and ecology (10%). Others were reported at 13%.

The future needs of the FAIRagro community point to a very broad range of data and specific support services and training opportunities for these. In particular, a need was expressed for omics, laboratory, remote sensing, and time series data. The types of data to work with vary from numeric data (31%), geographic data (18%), and text data (16%). It is important to note that a small portion will work with non-digital data and source code (7% each).

1.4 A tight Network within NFDI

FAIRagro is well connected within the NFDI and with the NFDI Directorate, and its partner institutions are members of the NFDI Association. The FAIRagro consortium supports the idea

of the Research Data Commons [5; 6], will network the infrastructures and services used and align them with this concept. FAIRagro is THE NFDI hub for agricultural research data and is networked with the broader agricultural science community. FAIRagro is involved in many NFDI consortia (NFDI4Earth, DataPLANT, NFDI4BioDiversity, NFDI4Health, NFDI4Objects, NFDI4DataScience, NFDI4Culture, NFDI4Memory, NFDIxCS, MaRDI, and NFDI4Chem), enabling direct information exchange towards the DSSC. FAIRagro has agreed with NFDI4Earth, NFDI4BioDiversity, DataPLANT, NFDI4Microbiota, NFDI4Health, NFDI4Objects, and NFDI4DataScience on cross-consortia activities and developments such as the data steward's service in research data management.

1.5 Outlook

Ultimately, the innovative structure of a DSSC aims to achieve a high level of RDM awareness in the agricultural-science community and beyond. With our Help Desk on the FAIRagro Portal, we will provide a low-threshold, user-friendly service hub where people can share and find data, learn about RDM through documents and in person and get support in all stages of the data life cycle. Personal support of the data stewards can be attained at any time for a large range of agrosystem-related topics from data-management plans to legal issues, big data, geodata, metadata and more. Our goal is to steadily develop greater expertise in specific RDM needs of agrosystem subfields and to create a network beyond our field and together with other RDM hubs which is easy to be navigated along for users of our data services.

Data availability statement

- FAIRagro survey 2021: https://doi.org/10.5073/20211013-105447 (data)
- DaKA community survey 2023: https://doi.org/10.20387/BONARES-STHV-TY43

Author contributions

Nikolai Svoboda: conceptualization, writing – original draft, writing – review & editing

Lucia Vedder: review & editing

Franziska Böhm: review & editing

Markus Möller: review & editing

Elena Rey-Mazón: review & editing

Marcus Schmidt: conceptualization, writing – original draft, writing – review & editing

Birte Lindstädt: review & editing

Ulrike Stahl: funding acquisition, project administration, writing – review & editing

Competing interests

The authors declare that they have no competing interests.

Funding

NFDI consortia are funded by the Deutsche Forschungsgemeinschaft DFG: FAIRagro grant no. no. 501899475.

Acknowledgement

Xenia Specka and the coordination and management team of FAIRagro.

References

1. Svoboda, N., Stahl, U., Möller, M., Everwand, R., Bracke, J., Duttmann, R., Kuhwald, M., Wamhof, T., Bock, D., Tapken, H., Herrmann, L., & Stiene, S. (2023). 2023 survey of agricultural science community needs: data use, data competency, support. (Version 1.0) [Data set]. Leibniz Centre for Agricultural Landscape Research (ZALF). https://doi.org/10.20387/BONARES-STHV-TY43
2. Senft, M., Stahl, U., Svoboda, N., 2021. Dataset: survey about research data management in agricultural sciences in Germany. https://doi.org/10.5073/20211013-105447
3. Senft M, Stahl U, Svoboda N (2022) Research data management in agricultural sciences in Germany: We are not yet where we want to be. PLoS ONE 17(9): e0274677. https://doi.org/10.1371/journal.pone.0274677
4. DFG subject groups: https://www.dfg.de/dfg_profil/gremien/fachkollegien/liste/index.jsp?id=207
5. Bierwirth, M., Glöckner, F.O., Grimm, C., Schimmler, S., Boehm, F., Busse, C., Degkwitz, A., Koepler, O., Neuroth, H., 2020. Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung.10.5281/zenodo.3895209.
6. Glöckner et al. (2020) Glöckner, F.O., Diepenbroek, M., Felden, J., Güntsch, A., Stoye, J., Overmann, J., Wimmers, K., Kostadinov, I., Yahyapour, R., Müller, W., Scholz, U., Triebel, D., Frenzel, M., Gemeinholzer, B., Goesmann, A., König-Ries, B., Bonn, A., Seeger, B., 2020. NFDI4BioDiversity - A Consortium for the National Research Data Infrastructure (NFDI).10.5281/zenodo.3943645.