ZBW *Publikationsarchiv*

Publikationen von Beschäftigten der ZBW – Leibniz-Informationszentrum Wirtschaft
Publications by ZBW – Leibniz Information Centre for Economics staff members

Schimmler, Sonja et al.

Conference Paper — Published Version
NFDI4DS Infrastructure and Services

Kontakt/Contact
ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der
Leibniz-Gemeinschaft

NFDI4DS Infrastructure and Services

Sonja Schimmler,[1] Bianca Wentzel,[1] Arnim Bleier,[2] Stefan Dietze,[2] Saurav Karmakar,[2] Peter Mutschke,[2] Angelie Kraft,[3] Tilahun A. Taffa,[3] Ricardo Usbeck,[3] Zeyd Boukhers,[4] Sören Auer,[5] Leyla Jael Castro,[6] Marcel R. Ackermann,[7] Thomas Neumuth,[8] Daniel Schneider,[8] Ziawasch Abedjan,[9] Atif Latif,[10] Fidan Limani,[10] Raia Abu Ahmad,[11] Georg Rehm,[11] Sima Attar Khorasani,[12] Matthias Lieber[12]

Abstract: NFDI4DataScience (NFDI4DS) is a consortium founded to support researchers in all stages of the research data lifecycle in order to conduct their research in line with the FAIR principles. The infrastructure developed targets researchers from a wide range of disciplines working in the field of data science and artificial intelligence. NFDI4DS contributes to systematically understanding the needs and challenges of researchers in various disciplines regarding data science and artificial intelligence, keeping in mind ethical, legal and social aspects. The identified needs will be addressed by support structures such as educational videos and challenges. Transparency, reproducibility and FAIRness will be improved by integrating existing and newly developed services into the NFDI4DS infrastructure, and by systematically adding all digital objects (articles, data, machine learning models, workflows, scripts/code, etc.) to the NFDI4DS research knowledge graph. This paper presents the goals of NFDI4DS, and gives an overview on what the consortium is going to contribute to the data science and artificial intelligence communities. It focuses on existing and newly developed services and their integration.

Keywords: NFDI; NFDI4DS; Research Data Infrastructures

1   Introduction

Making research available to other researchers does not only mean to publish articles. It means making the research lifecycle and all digital objects *findable*, *accessible*, *interoperable* and *reusable* (FAIR) [Wi16]. As data science (DS) and artificial intelligence (AI) continue to evolve and the methods used for data collection/creation, processing and analysis become

[1] Fraunhofer Institute for Open Communication Systems, Berlin, Germany

[2] GESIS Leibnitz Institute for the Social Sciences, Cologne, Germany

[3] Universität Hamburg, Germany

[4] Fraunhofer Institute for Applied Information Technology, Sankt Augustin, Germany

[5] TIB Leibnitz Information Centre for Science and Technology and L3S, University of Hannover, Germany

[6] ZB MED Information Centre for Life Sciences, Cologne, Germany

[7] Schloss Dagstuhl, Leibniz Center for Informatics, Trier, Germany

[8] Leipzig University, Germany

[9] Leibniz Universität Hannover and L3S, Hannover, Germany

[10] Leibniz Information Centre for Economics, Kiel, Germany

[11] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Berlin, Germany

[12] Technical University Dresden, Germany

more complex, it can be difficult to maintain transparency, reproducibility, and fairness in research.

The NFDI4DS consortium, part of the NFDI initiative to build a German national research data infrastructure, aims to provide comprehensive support for all phases of the complex and interdisciplinary *research data lifecycle* within the DS and AI communities. The research data lifecycle includes collection/creation of data, its processing, analysis, preservation, access, and reuse of digital objects.

2 Approach

To achieve its goals, the consortium plans to adopt an integrated approach towards research management. This involves making all relevant digital objects available and interlinking them to provide a comprehensive view of research (improving transparency). Additionally, tools and services will be offered to facilitate the management of these resources. These efforts will ensure that valuable resources are available and accessible for reuse (improving reproducibility), ultimately contributing to the growth and development of these fields.

NFDI4DS aims to tackle the needs of various scientific fields involved in DS and AI by offering interdisciplinary as well as domain-specific services, accompanied with shared tasks and associated benchmarking datasets. Furthermore, support structures are being set up, including the circulation of learning materials and best practice guidelines, and the conduction of tutorials, workshops and challenges.
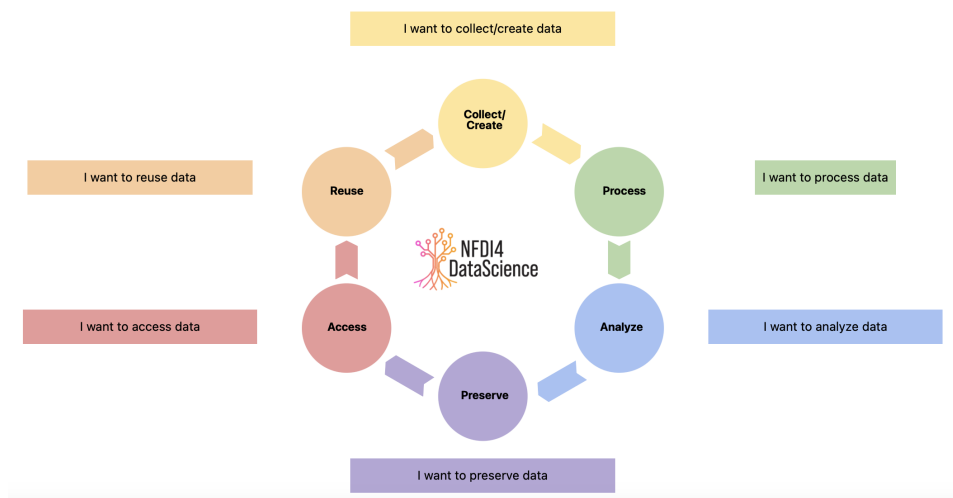

Fig. 1: Core Services Dashboard

3 Core Services

The core services focus on six main areas around digital objects: collection/creation, processing, analysis, preserving, access, reuse, and evaluation (see Figure 1). Digital objects include artifacts beyond articles, such as data, machine learning models, workflows, and scripts/code. Collection/creation, processing and analysis mainly deal with data, including data profiling [AGN15], cleaning [Ma19] and transformation[Ab16]. Preservation, access, reuse and evaluation covers a broader spectrum of digital objects as they refer not only to data. The NFDI4DS research knowledge graph forms the basis of the infrastructure, providing details about digital objects and their interrelation.

The NFDI4DS infrastructure is based on a number of already existing software components and already well-established tools and services, which target different phases of the research data lifecycle (see Figure 1). Some of these are detailed below:

- *The Open Research Knowledge Graph (ORKG)* [Au20, Ja19] is a service for semantically describing research contributions in a knowledge graph. The semantic descriptions of articles are crowd-sourced from authors and researchers leveraging NLP of articles.
- *The GESIS Knowledge Graph Infrastructure* consists of tools and pipelines for constructing actual research knowledge graphs of research information, metadata and primary research data, such as SoMeSci [Sc21], SoftwareKG [Sc22] or TWeetsKB [Di20].
- *The DBLP Computer Science Bibliography* is an open bibliographic data base, search engine, and knowledge graph on computer science publications. All of DBLP's curated data is available via its website, APIs, and as a data dump download [dt23].
- *The Data Management Platform Piveau* is developed within the scope of *data.europa.eu* [Ki19, Ki20]. Piveau provides services and pipelines for harvesting data and metadata from various sources saving it into a knowledge graph utilizing semantic linking. The knowledge graph can be explored and accessed via a Web interface, APIs and a SPARQL endpoint.
- *The General Entity Annotator Benchmark (GERBIL)* [RUN18] is a FAIR benchmarking platform for entity annotation and disambiguation, as well as question-answering systems [Us19].
- *GESIS Notebooks* [Bl22, Ra18] is an online reproducibility service for FAIR digital objects. Its main components are a BinderHub, a JupyterHub, and a place to publish, explore, try out, and learn about DS and AI methods. GESIS Notebooks is part of the BinderHub Federation powering mybinder.org.

4 Service Integration

Our service integration strategy is inspired by the EOSC Interoperability Framework [Co21], which is based on: (1) persistent identification using PIDs, such as DOI, ORCID

or authoritative URIs, (2) authentication and authorisation (AAI) adhering to common standards, (3) semantic interoperability using RDF, vocabularies and ontologies, and (4) API integration based on REST principles. Below we describe some exemplary integrations between the NFDI4DS core services, currently under development:

- *NFDI4DS Research Knowledge Graph.* The NFDI4DS research knowledge graph will entail automatically extracted metadata about resource relations, e. g. software mentions in scholarly publications [Sc22], highly quality-controlled manual annotations of scholarly resources [Sc21] and community-annotations of scholarly publications from the ORKG [Au20, Ja19]. A particular focus is on sharing data as well as methods and infrastructure for extracting and curating scholarly knowledge.
- *NFDI4DS Registries.* The consortium aims at providing registries for different digital objects, one of which is the DBLP computer science bibliography. DBLP and the ORKG started linking author and publication entities within their respective knowledge graphs. The integration makes use of WikiData [Vr12] as a central, community-curated ID hub; matching results will be mirrored there. More sources and digital objects will be integrated as needed.
- *NFDI4DS Gateway and Portal.* The portal will include a federated search engine and recommendation system. Through a unified and intuitive search interface, users are enabled to query a wide range of scientific databases such as DBLP, Zenodo, and OpenAlex. While the gateway queries APIs in an ad-hoc fashion, the portal, which will be based on Piveau, provides a harvesting-based service to account for larger research data dumps. The overall aim is to design an entry point that categorises and summarises multiple search results (such as researchers, articles, machine learning models, benchmarking results etc.) such that practitioners and researchers gain a swift overview of existing contributions.
- *NFDI4DS Notebooks.* The consortium aims at integrating different tools and services, including a Jupyter Hub instance. To facilitate the further analysis and visualisation of digital objects contributed by the participating services, they can be directly loaded into a new MyBinder Jupyter notebook data frame for further programmatic processing, e. g. for generating visualisations.
- *Compute Infrastructure.* While most of the NFDI4DS tools and services will be hosted in a cloud infrastructure located at various sites of the partners, some DS and AI tasks require access to specialized high-performance computing (HPC) resources such as GPU accelerators and large memory capacities. The goal is to make the transition from cloud to HPC as easy as possible for both service providers and users. On a Web interface users can specify their compute resource needs, which will be handed over to the batch system SLURM to launch JupyterHub on the HPC system.

5 Conclusion

Although the idea of portals around DS and AI is not new, there is still a lot of work to do in this area of such portals, especially when it comes to the inclusion of different digital

objects and the support of multiple interdisciplinary communities. One difference between the NFDI4DS portal and other related efforts (e. g., HuggingFace[13]) is the emphasis on FAIR metadata for digital objects. Proper and rich metadata enables not only FAIR but also further empowers *Linked Open Science* to improve best practices for conducting research including aspects such as transparency, reproducibility and fairness.

Acknowledgements

Bibliography

[Ab16]   Abedjan, Ziawasch; Morcos, John; Ilyas, Ihab F.; Ouzzani, Mourad; Papotti, Paolo; Stonebraker, Michael: DataXFormer: A robust transformation discovery system. In: 32nd IEEE International Conference on Data Engineering, ICDE 2016, Helsinki, Finland, May 16-20, 2016. IEEE Computer Society, Helsinki, Finland, pp. 1134–1145, 2016.

[AGN15] Abedjan, Ziawasch; Golab, Lukasz; Naumann, Felix: Profiling relational data: a survey. VLDB J., 24(4):557–581, 2015.

[Au20]   Auer, Sören; Oelen, Allard; Haris, Muhammad; Stocker, Markus; D'Souza, Jennifer; Farfar, Kheir Eddine; Vogt, Lars; Prinz, Manuel; Wiens, Vitalis; Jaradeh, Mohamad Yaser: Improving Access to Scientific Literature with Knowledge Graphs. Bibliothek Forschung und Praxis, 44(3):516–529, 2020.

[Bl22]   Bleier, Arnim; Erdogan, Kenan; Kahmann, Christian; Posch, Lisa: , GESIS Notebooks: Online Reproducible Computational Analysis for the Social Sciences, 2022.

[Co21]   Commission, European; for Research, Directorate-General; Innovation; Corcho, O; Eriksson, M; Kurowski, K; Ojsteršek, M; Choirat, C; Sanden, M; Coppens, F: EOSC interoperability framework : report from the EOSC Executive Board Working Groups FAIR and Architecture. Publications Office, Brussels, Belgium, 2021.

[Di20]   Dimitrov, Dimitar; Baran, Erdal; Fafalios, Pavlos; Yu, Ran; Zhu, Xiaofei; Zloch, Matthäus; Dietze, Stefan: TweetsCOV19 - A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic. In: Proceedings of the 29th ACM International Conference on Information and Knowledge Management. CIKM '20, Association for Computing Machinery, New York, NY, USA, p. 2991–2998, 2020.

[dt23]   dblp team: , dblp computer science bibliography. Download, January 2023. Monthly RDF release of January 2023.

[Ja19]   Jaradeh, Mohamad Yaser; Oelen, Allard; Farfar, Kheir Eddine; Prinz, Manuel; D'Souza, Jennifer; Kismihók, Gábor; Stocker, Markus; Auer, Sören: Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. In: Proceedings of the 10th International Conference on Knowledge Capture. K-CAP '19, Association for Computing Machinery, New York, NY, USA, p. 243–246, 2019.

---

[13] https://huggingface.co/

[Ki19]   Kirstein, Fabian; Dittwald, Benjamin; Dutkowski, Simon; Glikman, Yury; Schimmler, Sonja; Hauswirth, Manfred: Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP. In (Lindgren, Ida; Janssen, Marijn; Lee, Habin; Polini, Andrea; Rodríguez Bolívar, Manuel Pedro; Scholl, Hans Jochen; Tambouris, Efthimios, eds): Electronic Government. Springer International Publishing, Cham, pp. 192–204, 2019.

[Ki20]   Kirstein, Fabian; Stefanidis, Kyriakos; Dittwald, Benjamin; Dutkowski, Simon; Urbanek, Sebastian; Hauswirth, Manfred: Piveau: A Large-Scale Open Data Management Platform Based on Semantic Web Technologies. In (Harth, Andreas; Kirrane, Sabrina; Ngonga Ngomo, Axel-Cyrille; Paulheim, Heiko; Rula, Anisa; Gentile, Anna Lisa; Haase, Peter; Cochez, Michael, eds): The Semantic Web. Springer International Publishing, Cham, pp. 648–664, 2020.

[Ma19]   Mahdavi, Mohammad; Abedjan, Ziawasch; Fernandez, Raul Castro; Madden, Samuel; Ouzzani, Mourad; Stonebraker, Michael; Tang, Nan: Raha: A Configuration-Free Error Detection System. In (Boncz, Peter A.; Manegold, Stefan; Ailamaki, Anastasia; Deshpande, Amol; Kraska, Tim, eds): Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019. ACM, Amsterdam, The Netherlands, pp. 865–882, 2019.

[Ra18]   Ragan-Kelley, Benjamin; Willing, Carol; Akici, F; Lippa, D; Niederhut, D; Pacer, M: Binder 2.0-Reproducible, interactive, sharable environments for science at scale. In: Proceedings of the 17th python in science conference. F. Akici, D. Lippa, D. Niederhut, and M. Pacer, eds., scipy.org, Austin, TX, USA, pp. 113–120, 2018.

[RUN18]   Röder, Michael; Usbeck, Ricardo; Ngomo, Axel-Cyrille Ngonga: GERBIL - Benchmarking Named Entity Recognition and Linking consistently. Semantic Web, 9(5):605–625, 2018.

[Sc21]   Schindler, David; Bensmann, Felix; Dietze, Stefan; Krüger, Frank: SoMeSci- A 5 Star Open Data Gold Standard Knowledge Graph of Software Mentions in Scientific Articles. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management. CIKM '21, Association for Computing Machinery, New York, NY, USA, p. 4574–4583, 2021.

[Sc22]   Schindler, David; Bensmann, Felix; Dietze, Stefan; Krüger, Frank: The role of software in science: a knowledge graph-based analysis of software mentions in PubMed Central. PeerJ Computer Science, 8:e835, January 2022.

[Us19]   Usbeck, Ricardo; Röder, Michael; Hoffmann, Michael; Conrads, Felix; Huthmann, Jonathan; Ngomo, Axel-Cyrille Ngonga; Demmler, Christian; Unger, Christina: Benchmarking question answering systems. Semantic Web, 10(2):293–304, 2019.

[Vr12]   Vrandecic, Denny: Wikidata: a new platform for collaborative data collection. In (Mille, Alain; Gandon, Fabien; Misselis, Jacques; Rabinovich, Michael; Staab, Steffen, eds): Proceedings of the 21st World Wide Web Conference, WWW 2012, Lyon, France, April 16-20, 2012 (Companion Volume). ACM, Lyon, France, pp. 1063–1064, 2012.

[Wi16]   Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1):160018, December 2016. Number: 1.

NFDI4DS Transfer and Application

Ekaterina Borisova,[1] Raia Abu Ahmad,[1] Georg Rehm,[1] Ricardo Usbeck,[2] Jennifer D'Souza,[3] Markus Stocker,[3] Sören Auer,[3] Judith Gilsbach,[4] Anastasia Wolschewski,[5] Johannes Keller,[5] Daniel Schneider,[5] Thomas Neumuth,[5] Sonja Schimmler[6]

Abstract: Due to the ever increasing importance of Data Science and Artificial Intelligence methods for a wide range of scientific disciplines, ensuring *transparency* and *reproducibility* of DS and AI methods and research findings have become essential. The NFDI4DS project promotes the *findability*, *accessibility*, *interoperability*, and *reusability* in DS and AI by developing an open integrated research data infrastructure in which all artefacts (e. g., papers, code, models, datasets) will be interlinked in a FAIR and transparent way. One of the key aspects is to build a bridge between NFDI4DS and other research communities which actively apply DS and AI methods. This paper describes the main actions taken to engage with the relevant (sub)communities.

Keywords: NFDI; NFDI4DS; Research Data Infrastructure; Data Science; Artificial Intelligence

1 Introduction

Most research artefacts nowadays are spread across repositories, digital libraries, and institutional databases. Such unsystematic and decentralised storage complicates the *findability*, *accessibility*, *interoperability*, and *reusability* (FAIR) [WDA16] of scientific data. The German National Research Data Infrastructure (NFDI)[7] initiative aims to interconnect and preserve interdisciplinary scholarly data in a FAIR way. NFDI for Data Science and Artificial Intelligence (NFDI4DS)[8] is one of the 26 NFDI consortia which promotes the idea of FAIRness, transparency and reproducibility in DS and AI. NFDI4DS's goal is to support all phases of the research data life cycle, including collecting, creating, comparing, processing, analysing, publishing, archiving, and reusing resources in DS and AI through innovative tools and services.

---

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Berlin, Germany
ekaterina.borisova@dfki.de, raia.abu_ahmad@dfki.de, georg.rehm@dfki.de

[2] Universität Hamburg, Hamburg, Germany, ricardo.usbeck@uni-hamburg.de

[3] TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany
jennifer.dsouza@tib.eu, markus.stocker@tib.eu, auer@tib.eu

[4] GESIS Leibniz Institute for the Social Sciences, Cologne, Germany, judith.gilsbach@gesis.org

[5] Innovation Center Computer Assisted Surgery (ICCAS), Leipzig University, Germany
anastasia.wolschewski@iccas.de, johannes.keller@iccas.de, daniel.schneider@iccas.de,
thomas.neumuth@iccas.de

[6] Fraunhofer FOKUS, sonja.schimmler@fokus.fraunhofer.de

[7] https://www.nfdi.de

[8] https://www.nfdi4datascience.de

To achieve this objective, it is important to engage with the relevant communities and to transform NFDI4DS's visions into concrete applications. We collaborate with four scientific fields, i. e., Natural Language Processing (NLP) and Language Technology (LT) including Semantic Web, Biomedical Research, Information Sciences and Social Sciences.

2   Applications in Natural Language Processing

Like the other domains in scope of the project, researchers from the LT, NLP, and Semantic Web communities apply a wide variety of Research Data Management (RDM) methodologies. Although these communities follow similar research avenues, they utilise heterogeneous RDM techniques. This leads to inconsistency in documentation and sharing of digital artifacts (e. g., data, code, models) contributing to low transparency and reusability. Our goal is to bridge the gap between LT, NLP and Semantic Web by hosting different events (e.g., workshops, shared tasks, challenges) and by spreading knowledge about RDM best practices. We currently work on the following activities:

- Three shared tasks[9] which address problems in the area of scholarly information processing, i. e., software mention detection, leaderboard mining [KDA21, KDA23a, KDA23b], and research field classification. By organising these shared tasks we enable and encourage the sharing of RDM techniques, digital artifacts and evaluation measures across LT, NLP and Semantic Web communities.
- The ISWC 2023 Scholarly Question Answering challenge[10] allows researchers to compare their Knowledge Graph Question Answering Systems.
- We also work on compiling Open Science Best Practices in DS and AI, i. e., recommendations for ensuring a FAIR life cycle of digital artifacts. The guidelines will be published on the NFDI4DS website.

3   Applications in Biomedical Research and Clinical Decision Making

In the biomedical domain, Data Science faces obstacles due to 1. the safety-critical and ethically relevant nature of clinical applications, 2. complex heterogeneous and incomplete datasets, and 3. a lack of standardisation and data-privacy regulations. We plan to leverage the NFDI4DS infrastructure to foster data and knowledge transfer between the biomedical research community and DS. This ensures that the requirements of biomedical stakeholders can be adequately met within DS research. Our effort until 2026 includes:

- Host challenges dealing with the requirements of biomedical applications, starting with a competition related to reliable uncertainty estimation in 2024. The competitions will raise awareness for open research questions relevant to biomedical DS, flatten

---

[9] https://www.nfdi4datascience.de/docs/community-engagement/shared-tasks/
[10] https://kgqa.github.io/scholarly-QALD-challenge/

learning curves through tutorials and workshops, present overviews of the state of the art, provide benchmarks, and obtain novel solutions driving AI adoption in medicine.

- Promote the NFDI4DS infrastructure in the biomedical engineering, medical informatics, and DS research communities by giving talks and hosting focus sessions at relevant conferences (e. g., BMT[11], GMDS[12], and within ScaDS.AI[13]). These events will also help gather community feedback and requirements.
- Design tutorials and best practices for use cases from clinical prognosis and decision support using heterogeneous medical datasets. The tutorials will be deployed using the binder notebook service and involve topics including data preparation, synthetic data generation, dealing with missing data, and uncertainty estimation.

4   Application for Information Sciences

The Open Research Knowledge Graph (ORKG)[14] [St23, Au20, Ja19] is an NFDI4DS service for semantically describing research contributions published in scientific articles in a knowledge graph. Research contributions are semantic (i. e., machine-actionable) descriptions of published research findings together with the employed materials and methods and the addressed research problem. Semantic descriptions are crowd-sourced from authors and researchers. The ORKG also leverages NLP information extraction from articles to automate knowledge graph construction. ORKG supports numerous downstream services, including comparisons of research contributions addressing a common research problem, visualisations of compared information, leaderboards, etc. The ORKG is central for the NFDI4DS Application for Information Sciences, which has three goals:

- Deploy the most promising NLP models for information extraction from scholarly articles developed in NFDI4DS Shared Tasks in an infrastructure to further advance automation of scholarly knowledge graph construction and curation.
- Catalyse the adoption of the ORKG and, more generally, Scholarly Knowledge Graph infrastructure, in relevant conference series and journals. First experiments have been conducted in 2022 with ISWC and SEMANTiCS by actively guiding authors on how to include ORKG Comparisons in their submitted papers.
- Develop approaches to ensure expressions of research findings are produced machine actionable, and automatically flow into digital scholarly communication infrastructure such as ORKG. We currently test approaches for model performance evaluation conducted using Python or R computing environments to ensure that TDMS (Task, Dataset, Metric, Score) data published in articles automatically flows into ORKG benchmarks and leaderboards.

---

[11] https://bmt2023.de
[12] https://www.gmds2023.de
[13] https://scads.ai
[14] https://orkg.org

5   Applications in Social Sciences

Researchers from the Social Sciences handle very heterogeneous types of data, documented and shared in very diverse ways. The same applies to code written for data wrangling and analysis. Our goal is to encourage Social Science researchers to practice FAIR science and provide them with the knowledge and infrastructure. We currently compile the status quo regarding documentation, quality assessment and sharing behaviour of data and code. Subsequently, we will contribute to the implementation of our goals in the Social Sciences with an emphasis on making data and code available to the research community in a reusable way. Until 2026 we will be engaged in the following activities:

- Conduct a survey among researchers, investigating current data and code-sharing practices as well as quality assessment of digital research objects with a focus on relational social network data. For sharing behaviour a similar survey exists in the field of computational biology [CH22].
- Define use cases from an online access web-tracking panel under development at GESIS[15] to define requirements for data quality and sharing practices.
- Host a workshop for social scientists on open science practices, present NFDI4DS services and gather feedback from the Social Science community.

6   Call for Speedboat Projects

In addition to the four scientific areas mentioned in the last paragraphs, we are issuing a call for speed boat projects. These projects are meant to kick off collaboration with further domains, and to complement our overall service portfolio.

7   Conclusion

This paper presents NFDI4DS's main strategies to promote FAIR research data management across four scientific areas: 1. NLP and LT as well as Semantic Web, 2. Biomedical Research and Clinical Decision Making, 3. Information Sciences, and 4. Social Sciences. The current community engagement tools range from conducting various surveys and proposing best practices to organising shared tasks.

Acknowledgements

---

[15] https://www.gesis.org

Bibliography

[Au20]   Auer, Sören; Oelen, Allard; Haris, Muhammad; Stocker, Markus; D'Souza, Jennifer; Farfar, Kheir Eddine; Vogt, Lars; Prinz, Manuel; Wiens, Vitalis; Jaradeh, Mohamad Yaser: Improving Access to Scientific Literature with Knowledge Graphs. Bibliothek Forschung und Praxis, 44(3):516–529, 2020.

[CH22]   Cadwallader, Lauren; Hrynaszkiewicz, Iain: A Survey of Researchers' Code Sharing and Code Reuse Practices, and Assessment of Interactive Notebook Prototypes. PeerJ, 10:e13933, 2022.

[Ja19]   Jaradeh, Mohamad Yaser; Oelen, Allard; Farfar, Kheir Eddine; Prinz, Manuel; D'Souza, Jennifer; Kismihók, Gábor; Stocker, Markus; Auer, Sören: Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. In: Proceedings of the 10th International Conference on Knowledge Capture. K-CAP '19, Association for Computing Machinery, New York, NY, USA, pp. 243–246, 2019.

[KDA21]   Kabongo, Salomon; D'Souza, Jennifer; Auer, Sören: Automated Mining of Leaderboards for Empirical AI Research. In: Towards Open and Trustworthy Digital Societies: 23rd International Conference on Asia-Pacific Digital Libraries, ICADL 2021, Virtual Event, December 1–3, 2021, Proceedings 23. Springer, pp. 453–470, 2021.

[KDA23a] Kabongo, Salomon; D'Souza, Jennifer; Auer, Sören: ORKG-Leaderboards: A Systematic Workflow for Mining Leaderboards as a Knowledge Graph. arXiv preprint arXiv:2305.11068, 2023.

[KDA23b] Kabongo, Salomon; D'Souza, Jennifer; Auer, Sören: Zero-shot Entailment of Leaderboards for Empirical AI Research. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2023. 2023.

[St23]   Stocker, Markus; Oelen, Allard; Jaradeh, Mohamad Yaser; Haris, Muhammad; Oghli, Omar Arab; Heidari, Golsa; Hussein, Hassan; Lorenz, Anna-Lena; Kabenamualu, Salomon; Farfar, Kheir Eddine; Prinz, Manuel; Karras, Oliver; D'Souza, Jennifer; Vogt, Lars; Auer, Sören: FAIR Scientific Information with the Open Research Knowledge Graph. FAIR Connect, 1(1):19–21, Jan 2023.

[WDA16]   Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, et al.: The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data, 3(160018), 3 2016.