2nd Conference on Research Data Infrastructure (CoRDI): RDM in Action

Cross-disciplinary: RDM Infrastructure

eventual DOI

The industrial Data Mesh concept as blueprint for data publishing infrastructure in the engineering sciences

Mario Moser[1,*], Anas Abdelrazeq[1], and Robert H. Schmitt[1,2]

[1]Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Aachen

[2]Fraunhofer Institute for Production Technology (IPT), Fraunhofer Institute, Aachen

*Correspondence: Mario Moser, mario.moser@wzl-iqs.rwth-aachen.de

Abstract

With the rise of Research Data Management (RDM), increasing amount of scientific data is organised and published, making scientific results transparent and enable data reusage. However, future reuse purposes are often not known, and data is still scattered in different places, especially in sciences with high degree of heterogeneity, like the engineering sciences [1], [2]. Data can be provided in data publication journals, in repositories (institutional / generic / discipline-specific), or even web pages. York Sure-Vetter, director of NFDI, states 2023 in an interview with the German Süddeutsche Zeitung:

> ‘We are drowning in data, but cannot find it.’ There is a lack of interconnected data spaces for science, he says, meaning protected virtual locations that facilitate the exchange of data across disciplines. [3, translated by the author]
>
> Original quote: “„Wir ertrinken in Daten, können sie aber nicht finden.“ Es fehlten miteinander verknüpfte Datenräume für die Wissenschaft, sagt er, und meint geschützte virtuelle Orte, die den Austausch von Daten über Fachgrenzen hinweg erleichtern.” [3]

The *Data Mesh* paradigm from industrial data management appears as a potential solution approach: Unlike other approaches, distributed data is not integrated into a central system, but interconnected from its original publication place. This let data remain in their original (specialised) repository instead of building a new complex monolithic system. As socio-technical approach, underlying organisational structures and requirements are reflected, going beyond a purely technical solution. Data Mesh is driven by four principles (rf. Figure 1). First, Domain Ownership gives responsibility for the data to the creators instead of central IT teams, expecting to increase data quality by this. Second, product thinking is applied to data (‘Data as a Product’), leading to a certain form of standardisation w. r. t. comprehensibility (e. g. by metadata schema, data model) and accessibility (e. g. by APIs) for future data reusage, even in context different than initially planned. Third, a central Self-Serve Platform is the entry point to maintain and find registered datasets, e. g. in a data catalogue. Fourth, within a Federated Governance global interoperability is ensured by common standards, while domain-specific rules can be implemented locally. [4], [5], [6], [7], [8]

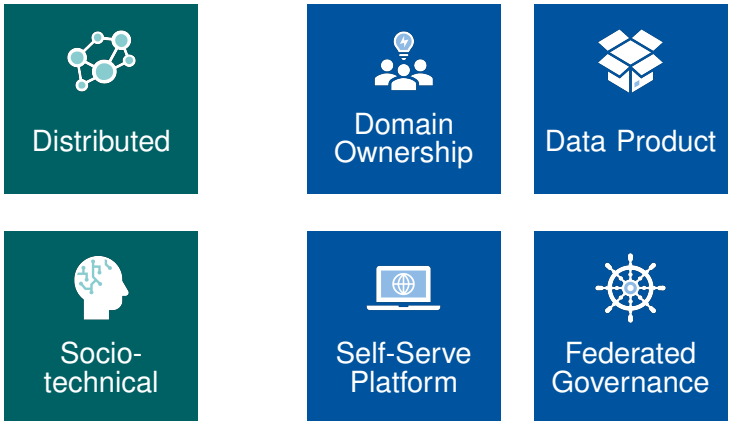

Figure 1. *Data Mesh characteristics (left) and principles (right) described by Dehghani [6]*

The Data Mesh for RDM is proposed and designed here as research data publication infrastructure in the engineering sciences with its heterogeneity regarding data formats, data structures, and disciplines (like mechanical / electrical / civil engineering, etc.). It contributes to making data FAIR [9]:

- Datasets can be **f**ound in (one or multiple) data catalogues as well as in their repositories.
- The Data Mesh redirects **a**ccess to the original source, keeping authentication and authorisation where required.
- Local and global governance specifies rules within domains and across the whole Data Mesh, fostering **i**nteroperability.
- Finally, governance and the data product approach make data **r**eusable.

Moreover, a Data Mesh for RDM provides the potential to support interdisciplinary data-driven research in engineering and beyond. Existing data from various sources will be explored and provided within such a Data Mesh [A]. Figure 2 depicts how existing components from RDM can contribute towards it. Architecture, concept, and use cases will be investigated in NFDI4ING's second funding period by the methodological-driven archetype Fiona [B]. The idea already is described in detail in a journal article [C].

This talk aims at spreading the vision in the RDM community beyond the engineering sciences and exchanging with other communities, in order to enable interdisciplinary research and identify components and activities that contribute to a Data Mesh for RDM.

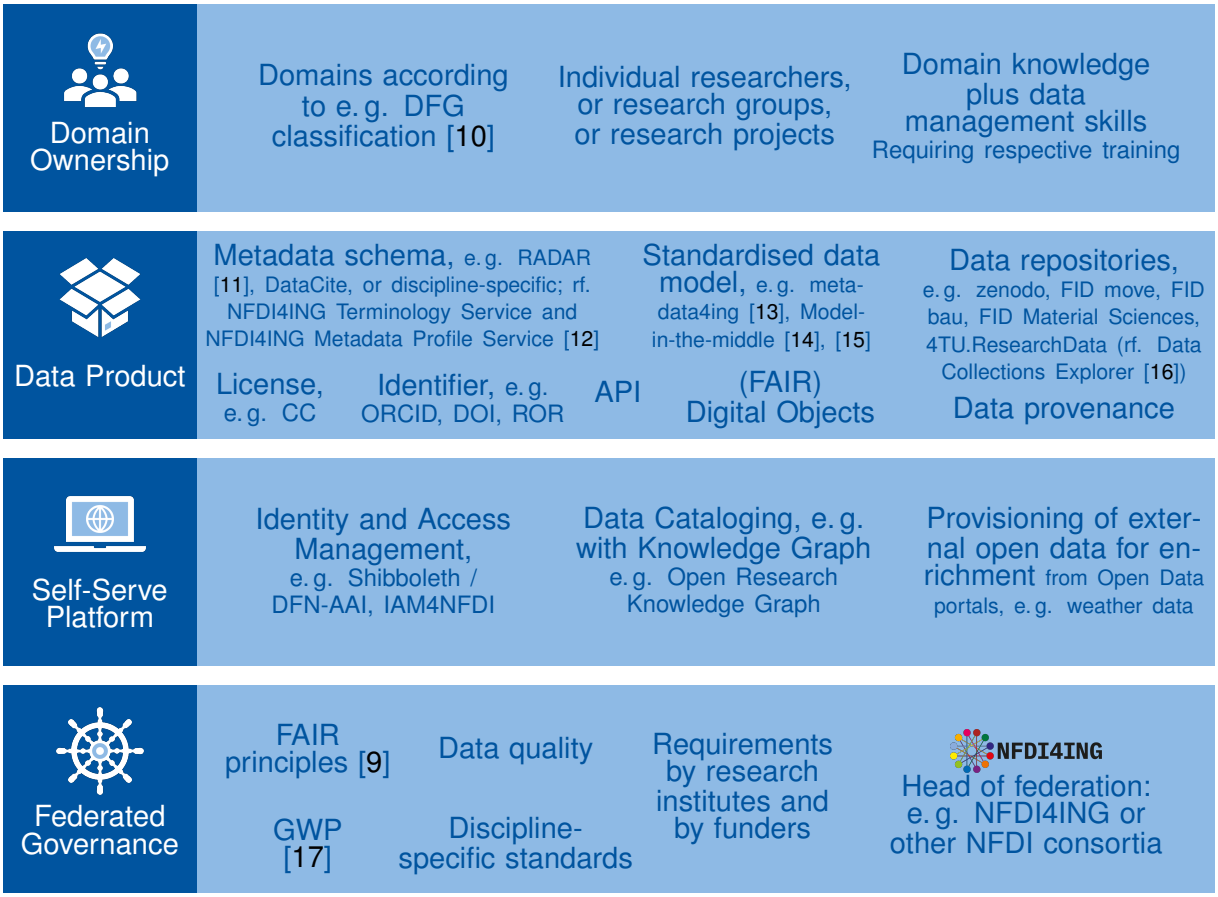

Figure 2. *Leveraging and connecting existing RDM components for a Data Mesh in the (engineering) research sciences*

Resources

[A] Introduction of NFDI4ING's archetype Fiona at the All-Hands-Meeting 2024. https://doi.org/10.5281/zenodo.15221815

[B] Webpage of NFDI4ING's archetype Fiona: https://nfdi4ing.de/home/25_services/25_manyparticipantsmanydevices/

[C] M. Moser, T. Hamann, A. Abdelrazeq, R. H. Schmitt, "Envisioning and proposing Data Mesh for Research Data Management in the Engineering Sciences", *ing.grid (Pre-Print)*, 2025, https://preprints.inggrid.org/repository/view/53/

Author contributions

M.M.: Conceptualization, Visualization, Writing – original draft, Writing – review & editing;

A.A.: Writing – review & editing, Supervision;

R.S.: Supervision, Funding acquisition

Competing interests

The authors declare that they have no competing interests.

Funding

The authors would like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the framework of the NFDI4ING consortium. Funded by the German Research Foundation (DFG) — project number 442146713.

References

[1] E. Ramalli and B. Pernici, "Challenges of a Data Ecosystem for scientific data," *Data & Knowledge Engineering*, vol. 148, p. 102 236, 2023, ISSN: 0169-023X. DOI: 10.1016/j.datak.2023.102236.

[2] S. D. Urban, J. J. Shah, M. Rogers, D. K. Jeon, P. Ravi, and P. Bliznakov, "A heterogeneous, active database architecture for engineering data management," *International Journal of Computer Integrated Manufacturing*, vol. 7, no. 5, pp. 276–293, 1994. DOI: 10.1080/09511929408944616.

[3] C. J. Meier, ""Wir ertrinken in Daten"," *Süddeutsche Zeitung*, 2023. [Online]. Available: https://www.sueddeutsche.de/wissen/nationale-forschungsdateninfrastruktur-daten-forschung-nfdi-1.6016012.

[4] Z. Dehghani, *How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh*, https://martinfowler.com/articles/data-monolith-to-mesh.html, 2019.

[5] Z. Dehghani, *Data Mesh Principles and Logical Architecture*, https://martinfowler.com/articles/data-mesh-principles.html.

[6] Z. Dehghani, *Data Mesh*. O'Reilly Verlag, 2023, ISBN: 9783960107248. [Online]. Available: https://content-select.com/de/portal/media/view/62d68bd8-66c4-4aae-918f-0d688677ec64.

[7] I. Machado, C. Costa, and M. Y. Santos, *Data-Driven Information Systems: The Data Mesh Paradigm Shift.* E. Insfran et al., Eds., Information Systems Development: Crossing Boundaries between Development and Operations (DevOps) in Information Systems (ISD2021 Proceedings)., Valencia, Spain: Universitat Politècnica de València., 2021. [Online]. Available: https://aisel.aisnet.org/isd2014/proceedings2021/currenttopics/9/.

[8] I. A. Machado, C. Costa, and M. Y. Santos, *Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures*, Procedia Computer Science, 2020. DOI: 10.1016/j.procs.2021.12.013.

[9] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160 018, Mar. 2016, ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18.

[10] DFG Deutsche Forschungsgemeinschaft, *DFG Classification of Scientific Disciplines, Research Areas, Review Boards and Subject Areas (2024-2028)*. [Online]. Available: https://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/.

[11] K. Soltau, *RADAR Metadata Schema*, 2017. DOI: 10.25504/FAIRsharing.e26f92.

[12] M. Grönewald et al., "Mit AIMS zu einem Metadatenmanagement 4.0: FAIRe Forschungsdaten benötigen interoperable Metadaten," in *E-Science-Tage 2021: Share Your Research Data*, V. Heuveline et al., Eds. heiBOOKS, Apr. 2022, pp. 91–104. DOI: 10.11588/heibooks.979.c13721. [Online]. Available: https://books.ub.uni-heidelberg.de/heibooks/catalog/book/979/chapter/13721.

[13] D. Iglezakis et al., “Modelling Scientific Processes With the m4i Ontology,” *Proceedings of the Conference on Research Data Infrastructure*, vol. 1, 2023. DOI: 10.52825/cordi.v1i.271.

[14] W. M. P. van der Aalst, “Experiences from the Internet-of-Production: Using “Data-Models-in-the-Middle” to Fight Complexity and Facilitate Reuse,” in *Business Process Management Workshops*, J. De Weerdt and L. Pufahl, Eds., Cham: Springer Nature Switzerland, 2024, pp. 87–91, ISBN: 978-3-031-50974-2. DOI: 10.1007/978-3-031-50974-2_7.

[15] I. Koren et al., “Navigating the Data Model Divide in Smart Manufacturing: An Empirical Investigation for Enhanced AI Integration,” in *Enterprise, Business-Process and Information Systems Modeling*, H. van der Aa, D. Bork, R. Schmidt, and A. Sturm, Eds., Cham: Springer Nature Switzerland, 2024, pp. 275–290, ISBN: 978-3-031-61007-3. DOI: 10.1007/978-3-031-61007-3_21.

[16] P. Ost and R. Stotzka, *Data Collections Explorer – An Information System for the Engineering Sciences*, Poster at the 1st NFDI-MatWerk Conference on Digital Transformation in Materials Science and Engineering (2023), Siegburg, Deutschland, 27.–29. Juni 2023, 2023. DOI: 10.5445/IR/1000160844.

[17] Deutsche Forschungsgemeinschaft, *Guidelines for Safeguarding Good Research Practice. Code of Conduct*, 2022. DOI: 10.5281/zenodo.6472827.