# A Digital Research Process for FAIR Data and Metadata

**J. Anniés[1], E. Dobrokhotova[2], J. Sautter[3], A. Wuchner[4]**

1 University of Stuttgart, Institute of Human Factors and Technology Management IAT, ORCiD 0000-0003-3248-0443
2 Fraunhofer-Gesellschaft e. V., Institute for Industrial Engineering IAO
3 Fraunhofer-Gesellschaft e. V., Institute for Industrial Engineering IAO, ORCiD 0000-0002-5934-7979
4 Fraunhofer-Gesellschaft e. V., Information Center for Planning and Building IRB, ORCiD 0000-0002-5549-3434

Fraunhofer IAO

Fraunhofer IRB

University of Stuttgart
Institute of Human Factors and Technology Management IAT

With new specialisations such as Data Science driven by digitalisation, efficiency potentials of a digital transformation are raised in both empirical research and data governance processes. Here, one challenge is to establish open and interoperable datasets, recognising the FAIR criteria ([1]) as an ideal to strive for. Data – as well as metadata – should comply to this standard. However, traditional methodological research processes ([2], p. 28; [3], p. 119) lack the support of information technology which would lever the process into the digital age. Therefore, we propose a digital research process that closes ranks between the traditional process and the opportunities of a digital world.

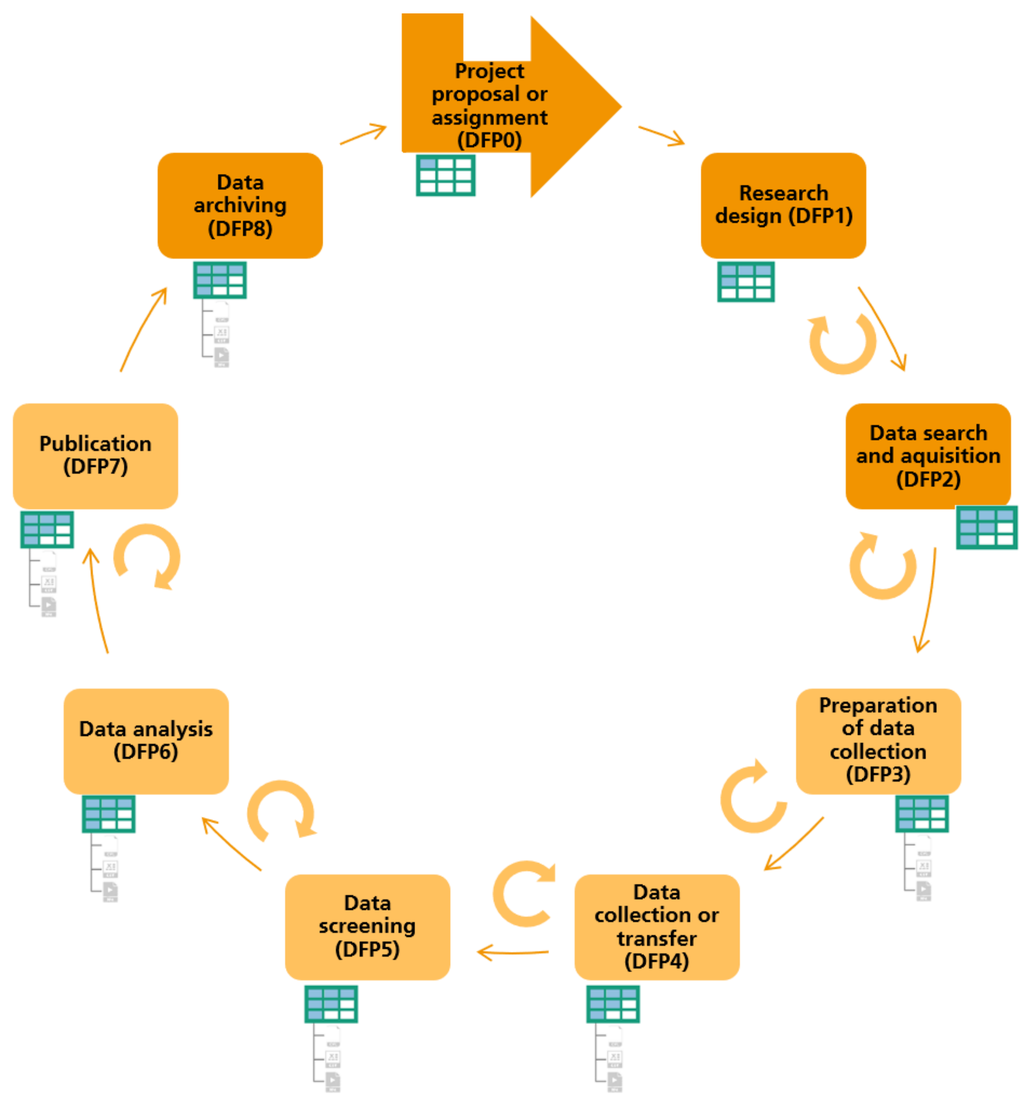
Based on practical experience and qualitative interviews in the research fields of qualitative and quantitative social sciences, building information modeling (BIM) and vehicle sensor data, we found that a data governance process depends less on the specific method and much more on a common cross-method research process. A digital research process thus needs to be highly adaptive to the purposes of different kinds of research domains ([4]). As a prerequisite to a digital research process, a concept for data governance ([5]; [6]), including corresponding roles as well as a data model was elaborated ([7]; [8]; [9]). Based on these requirements, we designed such a digital research process using the business process modeling notation (BPMN). This digital research process can serve as a blueprint for information systems or a data architecture for interdisciplinary research communities. In an iterative approach, we partly piloted the process within four research projects.

**A Digital Research Process for FAIR Data and Metadata**

The digital research process we propose consists of nine activities, terminated by data filing points (DFPs). This process can serve as a portrayal of a digital transformation of traditional methodological research processes and can be adapted in any project handling research data. The *obligatory* DFPs consider projects that focus on data search, acquisition, and archiving only. The *optional* DFPs represent the process of obtaining new (research) data. Additionally, optional data analysis may play a role in projects that merely reuse existing data. The optional DFPs represent the adaptability of research objectives in humanities. Equally unique to the digital research process is the frequent update of metadata throughout the research cycle, to create FAIR metadata throughout the time frame of the research and data processing. Optional iterations allow to re-visit the (meta) data of previous DFPs in order to update information and files of the research project.

**The Process Phases: Data Filing Points (DFPs)**

At **DFP0**, there is the smallest amount of meta data on the project proposal or assignment, such as the organisation's project number and project partners.
**DFP1** enriches the research project's meta data with information on the research design envisioned.
**DFP2** specifies from which sources the research data is acquired, for example in projects focusing on preexisting data that is already FAIR.
At **DFP3**, data files are added to the repository for the first time. To prepare for the data collection, data files such as GDPR statements could be collected, or code books developed.
**DFP4** consists of the files and meta data after the actual data collection or transfer. A research data set may be created.




*The Digital Research Process*

Before **DFP5**, the data set is screened and updated in order to prepare for the data analysis.
At **DFP6**, the final data set is uploaded after the (statistical) data analysis.
In case the scholars decide for publication, **DFP7** consists of the research articles and other dissemination material of the project.
As one of the obligatory DFPs, **DFP8** collects the final data and meta data of the project, even if the other process phases have not applied to the project. The DFP depends on the organisation's, or repository's terms of data archiving.
During all the DFPs, previously added meta data is updated.

**Legend**

- meta data
- meta data and data files
- data filing point (DFP)
- optional data filing point
- optional iteration

Contact

Dipl.-Inf. Johannes Sautter
Fraunhofer IAO
Urban Data & Resilience
Nobelstraße 12
70569 Stuttgart
Tel. +49 711 970-2387
johannes.sautter@iao.fraunhofer.de
www.muse.iao.fraunhofer.de

[1] Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3*(1), 1-9.
[2] Brosius, H. B., Haas, A., & Koschel, F. (2016). *Methoden der empirischen Kommunikationsforschung: eine Einführung* (7. Aufl.). Wiesbaden, Deutschland: Springer VS.
[3] Friedrichs, J. (1990). *Methoden empirischer Sozialforschung*. Springer-Verlag.
[4] UK Data Service (n. d.). *Research Data Lifecycle*. Retrievable at https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx, last accessed 01/07/2020.
[5] DAMA (Ed.) (2017): *DAMA-DMBOK. Data management body of knowledge. With the collaboration of Deborah Henderson a. Susan Earley*. Data Administration Management Association. Basking Ridge, New Jersey: Technics Publications.
[6] Kouper, I.; Raymond, A. H.; Giroux, S. (2020). An Exploratory Study of Research Data Governance in the U.S. In: *Open Information Science 4*(1), p. 122–142. DOI: 10.1515/opis-2020-0010.
[7] Wuchner, A., Sautter, J. (2020). Data Governance und Data Stewardship in der Fraunhofer-Gesellschaft. Presentation at *11. DINI/nestor-Workshop: Data Stewardship im Forschungsdatenmanagement - Was ist das? Rollen, Aufgabenprofile, Einsatzgebiete* on 17/11/2020.
[8] Sautter, J., Wuchner, A. (2020). *Data Governance: Als Voraussetzung für Datenexzellenz an Forschungsorganisationen*. Poster presented at RDA Deutschland Tagung, Potsdam, 25–27/02/2020. Retrievable at http://publica.fraunhofer.de/documents/N-586577.html, last accessed 01/07/2020.
[9] Sautter, J.; Litauer, R.; Fischer, R.; Klages, T.; Wuchner, A.; Müller, E.; ... & Riess, S. (2018). Beyond Data Quality: Data Excellence Challenges from an Enterprise, Research and City Perspective. In *DATA* (pp. 245-252). Retrievable at https://doi,org/10.5220/0006912902450252, last accessed 15/11/2021.

HIFIS HELMHOLTZ FEDERATED IT SERVICES + SOFTWARE SERVICES DESY.