IJDC | General Article

Data Curation: Introducing a Competency Framework for the Social Sciences

Kathrin Behrens
GESIS - Leibniz Institute for the Social Sciences

Tatiana Kvetnaya
Goethe University Frankfurt

Abstract

Research data management includes more than the question how researchers handle their data. In the sense of the FAIR principles, it is also about the sustainable safeguarding and organized reusability of research data. For social science, data-intensive research, research data centers and their data curating staff are therefore becoming increasingly important: data curators usually take on curation-specific tasks such as data preparation, securing research data in suitable archival environments, ensuring data accessibility, and the related control of the conditions of data re-use by third parties. Hence, they are specialized in the entire data curation process and, in particular, take on tasks of archiving and providing research data for reuse. Although the standards of comprehensive research data management are becoming more and more specific, this trend has not yet arrived in the corresponding training and further education measures. As a result, there is a gap between the growing demands on data curators and the development of competencies in the field of research data management with a focus on data curation. The competency framework presented in this article is intended to help close this gap: based on a Data Curation Lifecycle Model, a competency framework has been developed to support the development of targeted training and continuing education programs in the field of data curation, the formulation of learning objectives, and the evaluation of the corresponding trainings. The article points out the necessity to advance the development of competencies for this field, illustrates the schematic substructure of the data curation lifecycle, describes the development as well as the central core elements of the presented competency framework and discusses its perspectives. Overall, this competence framework is aimed in particular at (future) data curators, or as a schematic basis for the training of the relevant personnel. The focus is primarily on the data-intensive discipline of social sciences, although large parts can certainly be adapted for other disciplines and the corresponding data curation. The competency framework and this companion article are thereby intended to assist in advancing the sustainable professionalization of the previously understudied competency field of data curation.

Submitted date 31 May 2023 ~ *Revision received* 28 August 2024 ~ *Accepted* 24 February 2025

Correspondence should be addressed to Kathrin Behrens, Email: Kathrin.behrens@posteo.de

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction: Competency Development for Data Curation

Since technological progress and accompanying scientific areas such as research on big data, computational social sciences, or digitization and the related social changes are subject to steady growth within the social sciences research, data-based research as a whole becomes more prevalent. Accordingly, appropriate research data management (RDM) is becoming increasingly important for organizing this data-driven research and the corresponding research data. RDM determines all processes in the handling of research data, starting with study planning, data collection, data preparation and use, up to long-term storage (for example in repositories) and the subsequent usability of the research data by third parties. Accordingly, precise, sustainable and accurately executed RDM is an essential contribution to FAIR research data (Wilkinson et al., 2016), which describes the principles of findable, accessible, interoperable, and reusable research data. High-quality RDM and the corresponding documentation are a key element of responsible research, and are therefore increasingly considered a condition for research funding, for example by the European Research Council (European Commission, 2019).

This makes it even more important to identify specific RDM competencies in relation to various activity and training profiles. A schematic representation of competencies in the form of a structured framework helps in presenting activity-, task- or training-specific competency profiles that distinguish different levels of expertise.

As RDM topics are also increasingly addressed by the education and training sector, one can now also find more and more competency matrices dedicated to RDM qualification: Besides the competency profiles on the general approach to data literacy (e.g. Schüller, 2020), Blask et al. (2018) provide a research process-specific competency matrix for the introduction of RDM which is dedicated to the implementation of RDM structures at universities and non-university research institutions. Greci et al. (2018) explicitly focus on "safe data access professionals" and "it is intended to guide staff working in Safe Settings explore areas of professional development that will benefit them and add value to the services they work for, and the users they serve" (Greci et al., 2018, p. 2). Demchenko et al. (2021) provide a competence framework specifically for academically trained data stewards, that “should help implement existing FAIR and Data Stewardship competence frameworks in academic curricula.” (Demchenko et al., 2021, p. 9).

A recent example of a decomposition of RDM competencies is provided by Petersen et al. (2022): With their RDM-Learning Objectives Matrix, they aim to address a “generic approach that transcends subject and organizational boundaries, providing support for university lecturers, curricula managers, and individuals in the field of continuing education and training” (Petersen et al., 2022, p. 1 [translated by authors]).

In general, most of these descriptions of qualification profiles focus on RDM-topics related to the actual research process, for example, how to collect data properly, how to store datasets during the research process and how to secure individual rights and ethical aspects during data collection, which is of great importance in social science research. However, RDM goes beyond completing the data collection: To address the FAIR-principles (findable, accessible, interoperable and reusable data; Wilkinson et al., 2016) properly and to apply standards of good scientific practices, the collected data has to be prepared, stored, archived and finally it has to be made accessible for reuse in the long term.

Transferring research data to appropriate repositories or data archives is therefore only the beginning of a crucial RDM task process that makes it essential to ensure qualified handling of the transferred research data. Accordingly, the demand for qualified personnel in data curation positions has risen sharply, while at the same time curation specific RDM skills are often lacking, since RDM training for curating data has not been well established so far. Consequently, the need for target-group specific information and training measures for curation-focused RDM is also growing.

While the information and guiding resources for data curation related RDM in the context of FAIR data is already growing (e.g. Arguillas, 2022), the competency framework presented in this paper is intended to provide a supporting tool and a detailed scheme to fill the gap of curation

specific vocational trainings in the (social science oriented) RDM sector. This work is part of the Consortium for the Social, Behavioural, Educational and Economic Sciences (KonsortSWD)[1] within the National Research Data Infrastructure Germany (NFDI).[2]

The specific working group "Developing and exchanging RDM skills" [3] concerned the development of competencies in RDM, with a particular focus on employees in research data centres (RDC) being responsible for the curation of research data. With "RDM Compas – Research Data Management Competence Base",[4] a dedicated online platform has been set up for this purpose: by providing a RDM related knowledge base, "RDM Compas" aims to compile structured information and content on topics relating to social science data curation. Additionally, it will provide a specific training platform for employees in social science research data centers who deal with data curation topics.

To do justice to the fact that the curation of research data represents a complex profession (e.g. Kim et al., 2013), it makes sense to follow a basic scheme when developing the target group and activity-specific services of competency development. Just as for RDM skills tailored to researchers, a competency framework can be used as a schematic basis for the conception and design of competency development for data curators. It can be applied both as a general description of competencies as well as a basis for the formulation of learning objectives for concrete training opportunities. Hence, such a competency framework[5] serves as the foundation for a qualification profile, in order to provide proper services for RDM-competency development (Putnings et al., 2021, p. 256) for data curators.

The following section is devoted to the theoretical model for this competency framework. It sheds light on the procedural nature of data curation (section 2). Subsequently, the development of the competency framework and the resulting structure is described, and the underlying literature is excerpted in due course (section 3). Finally, the perspectives and limitations of the competency framework are discussed (section 4). For better readability of the article, the actual competency framework can be found in the appendix to the text (Appendix A).

The Data Curation Lifecycle

Data Curation is a highly complex profession involving all activities of "managing data from planning its creation, best practice in digitisation and documentation, and ensuring its availability and suitability for discovery and re-use in the future" (Abbott, 2008). Especially within the area of curating digital research data and the increasing significance of the FAIR-principles (Wilkinson et al., 2016), a sustainable and long-term oriented management of research data is of great importance in the domain of data-intensive science (Kim et al., 2013, p. 67). For instance, the social sciences undoubtedly represent a data-focused – and thus data-intensive – universe of different scientific sub-disciplines, such as sociology, political science, psychology, communication studies, and many others.

To cope with this discipline-specific high data intensity in the sense of sustainable handling and access, the role of data curators in the social science research data infrastructure is becoming increasingly important.[6] Therefore, it is crucial to advance the qualified training and professionalisation of data curators and thus to strengthen the quality of the social science research data infrastructure as a whole.

---

[1] KonsortSWD: https://www.konsortswd.de/en/
[2] NFDI: https://www.nfdi.de/en
[3] Working group: https://www.konsortswd.de/konsortswd/das-konsortium/services/rdm-skills/
[4] Online Access: www.rdm-compas.org
[5] The complete framework can be found in Appendix A.
[6] Although data curation is equally important in other disciplines, we focus on data curation in the social, educational, behavioural, and economic sciences because of our subject expertise and the orientation of our consortial affiliation. Explicitly, the (partial) adoption of relevant contents of the competence framework for other disciplines should not be excluded.

To contextualise the competence framework appropriately alongside further work, we focus on data curation personnel to emphasize the areas of competence involved in the entire process of data handling in research data centers, especially with regard to making the data available to third parties. Consequently, it is not directly about competencies required during the research process, although these are of course not completely irrelevant for the activity described (see further down in the article). There may well be overlaps in the competence descriptions with other professional profiles, such as data stewards, data managers, data professionals, etc. This is mainly due to the fact that, firstly, these job titles are not used uniformly in the labor market and, secondly, the job profiles overlap. However, the professional profile of data stewards in particular is becoming increasingly clear and prominent, for example thanks to the work of Demchenko et al. (2021) or Scholtens et al. (2022). To clarify that we are focusing in particular on the RDM training of employees in (social science-oriented) research data centers, we refer to data curation or data curating personnel in this paper and in the attached competence framework.

To be able to capture, describe, transport and, for example, translate the curation-specific management tasks into learning objectives in detail, the use of a conceptual basis appropriate to the complexity is of great value. To this end, a lifecycle model provides "a structure for considering the many operations that will need to be performed on data record throughout its life." (Ball, 2012, p. 3).

Research data lifecycles are most frequently applied to provide context for research-oriented competency profiles, and the competency framework presented in this paper applies the Data Curation Lifecycle Model (Higgins, 2008) by the UK Data Curation Centre (DCC).[7] Focusing on this curation-oriented lifecycle shapes the conceptual basis for RDM-related data curation tasks. In other words: Using the Data Curation Lifecycle Model (from here on: DCLM) provides a clear differentiation from the prominent research-centric perspective on RDM. Thereby, the focus is shifted towards an "high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt" (Higgins, 2008, p. 137).

The original impulse for the development of the DCC DCLM was the ever-increasing need to deal with digital objects, their curation specificity, and the proper way to handle them. Since the digital material in particular is sensitive to technological developments, the "curation and preservation activities undertaken, or neglected, in different stages of their management, can influence the ability to look after them successfully at subsequent stages." (Higgins, 2008, p. 135). In its original presentation, the DCC DCLM is a visual high-level overview of the different curation processes managing digital material. To achieve the goal of providing a competency framework for the services to RDC staff, we made slight adjustments and simplifications to the original DCLM. Figure 1 shows the simplified version of the DCLM used for this work.

In its centre, the DCLM focuses on a description of digital objects and databases to ensure a common understanding of its characteristics. As mentioned before, digital objects must be taken care of thoroughly, especially regarding technological development. It essentially distinguishes between continuous process steps (blue ring), such as "Description and Preservation Information", "Conceptualization & Preservation Planning", "Community Watch and Participation", as well as "Legal & ethical aspects", and sequential actions (orange ring), referring to actions like "Create and Receive", "Appraise and Select", "Ingest", "Preservation Action", "Store & Secure", "Access, Use and Reuse", and "Transform".[8] These two process levels are independent from each other and are also reflected separately in the competency framework with the corresponding process steps. The framework itself is then used to fill in the content of the various curation activities according to the criteria described further below (see section "Development and Structure of the Competency Framework for Data Curation", chapter 3).

[7] Digital Curation Centre (DCC): https://www.dcc.ac.uk/

[8] In its original version, the DCC DCLM also refers to occasional actions as "Dispose", "Reappraise" and "Migrate" that are only applied on a situation- or case-specific basis. Therefore, these aspects are not directly addressed in the simplified DCLM version of RDM Compas. For more information see Higgins (2008).

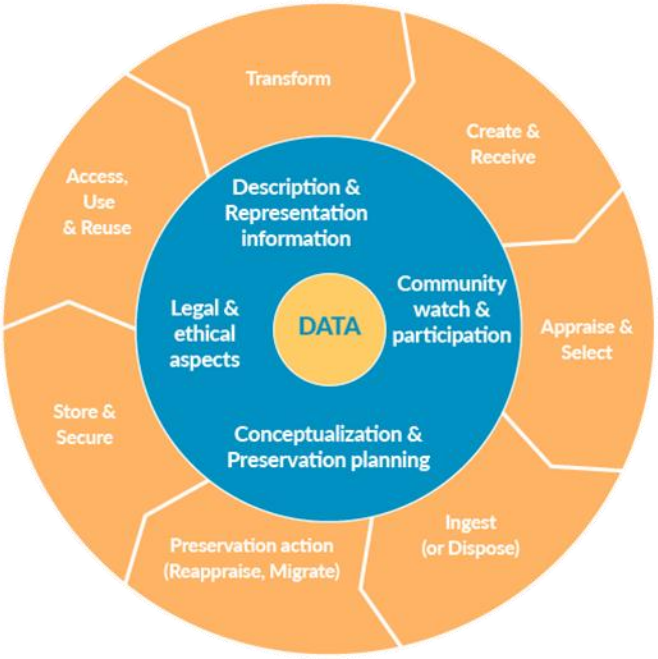

Figure 1. Data Curation Lifecycle by RDM Compas (following Higgins, 2008).

Overall, this subdivision facilitates a process-oriented view of curation activities around digital (research) data, including those of the data curator. Now, the activities and tasks of data curators are not (always) the same or as comprehensive as those of a data curator profiled here. Many certainly adopt only parts of the activities described here, or are more flexible and dynamic in handling the actual data curation, rather than strictly following the process flow outlined here. However, as the DCLM "can be used to plan activities within an organisation or consortium to ensure that all necessary stages are undertaken, each in the correct sequence" (Higgins, 2008, p. 137), it provides solid support as a schematic basis for the presentation of a curation process-related competency profile. Last but not least, a (research-)data-centred approach allows, in perspective, the comprehensive integration of courses or course modules targeting different job profiles related to data curation: Thus, it can focus on, for instance, RDC personnel with different areas of curation responsibilities on the one hand, as well as on researchers interested in or wanting to work in the field of data curation on the other hand.

All in all, the DCC Curation Lifecycle Model by Higgins (2008) served as the most helpful basis for the competency framework targeting data curators since it has been specifically established to be used as a blueprint for the development of RDM and archiving services. As it appears to be widely known in the professional RDM field, we consider it to be a well-suited starting point for acquiring RDM skills. The next chapter highlights the resulting development and structure of the competency framework presented in this paper.

Development and Structure of the Framework

One of the central challenges for developing the competency framework presented here was to provide a structure which meets the following requirements:

- based on a (research-)data-centred data curation lifecycle
- geared especially towards data curators

- encompasses multiple cognitive and behavioural facets of relevant competencies in order to aid an effective learning experience.

The following section sheds light on how the proposed framework structure intends to include features which help fulfil these requirements. Furthermore, it gives insight into the process of reaching the proposed structure.

At the beginning of the process, we conducted a literature search guided by the requirements listed above: Since the (slightly adjusted) DCLM (as described in chapter 2) is at the heart of the competency framework, additional literature was selected that provides valuable content regarding curation-specific RDM to fill the different steps within the DCLM with the respective competencies.

More specifically, we selected journal articles, project reports as well as training manuals which provide definitions and descriptions of specific RDM tasks, preferably with examples. In addition, they should either be geared towards the main target group (data curators) immediately or include content which can be adapted to address curation specific job tasks, ensuring that the content is applicable for the target group.

Finally, we particularly considered sources which utilized a multi-dimensional competency definition, i.e., by differentiating categories such as cognitive, behavioural, and motivational features of competencies. These categories of competency features are not related to a specific scientific discipline or area of competencies. Rather, they represent a universally valid dimension added to the proposed competency framework, with high pedagogical significance: The modelling of competencies as multi-faceted is widespread in the educational research literature to describe the construct of "being competent" or "having competencies" in a comprehensive way.

One example is the Taxonomy of Educational Objectives, also known as Bloom's Taxonomy (Bloom et al., 1956), which is one of the most recognized learning frameworks in the field of education, and frequently used to create learning outcomes and assessments. It comprises three learning domains: the cognitive (knowledge), psychomotor (skills), and affective (attitudes) domains. Thus, it addresses both a pure knowledge dimension, a skills dimension, and an attitudinal dimension, which allows for a holistic approach to the description of competencies. Because of this substantial advantage, this multi-faceted taxonomy is often used when it comes to describing competency profiles or learning objectives – of course also in data management. For instance, Sapp Nelson (2017) adapted this taxonomy to design a pilot competency matrix for data management skills. Similarly, in a working paper of the Hochschulforum Digitalisierung ("Higher Education Forum on Digitization"), Schüller (2020) grouped data literacy competencies according to the three areas of knowledge, action, and motivation. Based on this theoretical framework, she describes which competencies data literacy education should address, and provides detailed examples.

We rely on this understanding of competencies for multiple reasons: Firstly, it is supposed that these different dimensions of competencies should be addressed in a subject-appropriate way in order to increase learning success. By learning success, it is meant, for example, that a prospective RDC employee not only understands the required content in theory but is also able to successfully apply it in practice and at the same time is sufficiently motivated to do so in the short or long term. Secondly, the description of exemplary competencies – specifically for each dimension – is also useful as a guideline for the development of future training material. In this way, it can be ensured that newly developing material for RDC personnel takes relevant competency facets into account. Therefore, the three facets of competencies, Knowledge, Skills, and Attitude, are included as columns in the competency framework, and illustrative descriptions and examples for each specific facet are provided in the framework lines. Consequently, this results in the following structure:

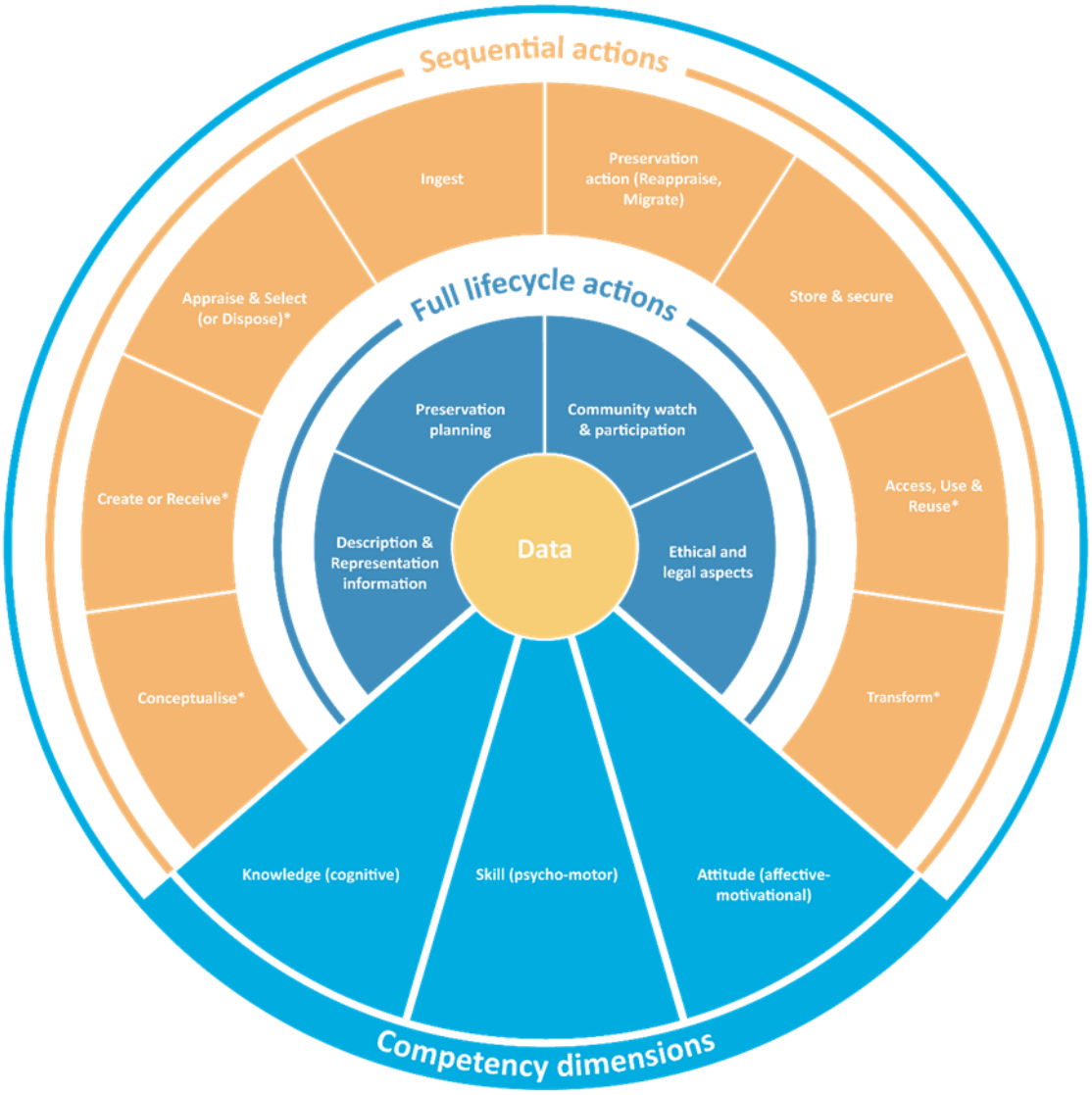

Figure 2. An extended illustration of the Competency Framework for Data Curation based on the DCC Curation Lifecycle Model. The inner circle contains the full lifecycle actions, which are relevant throughout the process of curating data (center) time-independently of the sequential actions (outer circle). The Competency dimensions (bottom triangle) represent different domains for successful acquisition of the described competencies (actions). These sequential actions are supplemented by a "research process" related competency description to support the advising of researchers.

The two rings of the lifecycle model map the specific competency areas of full-lifecycle-actions and sequential activities in data curation. The three light-blue "slices" refer to the respective competence dimensions that structure the contents of activity descriptions in our model. These general dimensions have been defined in previous educational research as aspects of successful learning, including illustrative descriptions and examples for each specific facet of a given lifecycle action in the full framework.

In a next step, an overview of the content within the selection of found published sources which was relevant for the aim was added. We mainly focused on suitable task descriptions and illustrative examples to complete the DCLM-oriented competency overview. For each source, an overview was created, that includes the specific curation-oriented content it contained.

When selecting the content, it was of particular importance that the descriptions were compatible with a research data manager's point of view. The curation-oriented content found in the identified sources was extracted and supplemented with additional competency descriptions when necessary. Proceeding in this way, the cells of the matrices were filled with the respective content according to the categories defined in Table 1. The full framework can be found in the Appendix A.

Furthermore, there is another distinctive feature of the competence framework: While data curators and research support staff are the primary target group, researchers and the research process remain important from their perspective in the context of curation tasks. Research support is often about advising and supporting researchers with regard to their handling of data and the research data management measures that are necessarily linked to the actual research process (e.g. data collection processes and their conditions).

Consequently, knowledge and skills related to advising researchers in research support are often in demand, making it a particularly relevant educational topic. Therefore, to the sequential measures identified as particularly relevant for researchers in light of the source literature, a row titled 'RDM services for researchers' has been added, describing the 'knowledge', 'skills' and 'attitude' facets for each competency that appear to be particularly relevant. These competency descriptions are thus related to the actual research process, but are highly relevant to the work of data curators, firstly because they should be aware of the relevant RDM requirements related to the research process itself, and secondly in order to be able to advise researchers on specific issues.

As a result, the competency framework combines valuable content from other frameworks with additional input and assembles the individual elements into a comprehensive activity profile of a data curator. It is not solely intended to function as an alternative to other existing, generic frameworks, but rather to substantiate the prospective services in the context of the RDM Compas information and training platform that has been designed on the basis of the same DCLM on the one hand, and providing a guideline for the scientific community developing similar training services tailored towards RDC personnel and data curators, especially within the field of social science research, on the other. The introduced framework is a first effort to achieve the initially mentioned requirements and remains open for future modifications and extensions.

Discussion: Perspectives and Limitations

The competency framework for data curation-based activities presented in this paper is designed to support data curators in different ways: It serves as structured framework for all (prospective) services provided by RDM Compas (with regard to the development of single course modules, their content-related connection to curation skills as well as of a comprehensive certification course and for the Knowledge Base), teaching all RDM-features relevant for curating digital research objects. The framework uniquely considers RDM competencies from a (social science) research-data-centred point of view, therefore being especially suitable for education of data-centred professions (like data managers, curators, stewards or RDC staff in general).

To this end, course profiles and/or training materials can be developed on the basis of the competence dimensions and their descriptions. The formulation of learning objectives, which can be derived from the competency framework, is a particularly important function in the training area. And finally, the explicit emphasis of the framework on the knowledge, practical, and motivational aspect on the one hand, and the researcher process perspective on the other hand, can help to ensure that these areas are considered and incorporated into the development of learning objectives to aid a holistic learning experience. In the long run, this framework can help to define and establish a professional profile for the work as a data curator.

While the competency framework for data curation is developed to provide a basis for the creation of new services aiming to support in questions on RDM, it also has a few limitations. One limitation of the approach proposed here directly affects the use of the DCC DCLM: As mentioned earlier, it is important to note that operational activities of data curators may well differ from the very model activity outlined here. This means that, on the one hand, the process steps

described are not necessarily processed in their entirety by one data curator. On the other hand, it is possible that, depending on the position related service, not all curation process steps are (or can be) carried out by one respective employee.

In addition, it is possible to deviate (at least partially) from the schematic process by establishing "best practices" for data curation. Therefore, the competency framework provided in this paper is not to be understood as “rigid corset” of competencies. Rather, it should serve as orientation and descriptive model that identifies a set of competencies and qualities that individual professionals need in their striving for professional excellence.

Another aspect is important in this context: when we talk about competences in this article, we are aware that we are using an analytical construct to present a frame, that is meant to facilitate guiding principles for the professionalisation of data curating employees. As the framework can be used as schematic tool to describe and differentiate between various qualification dimensions and the related competencies, the analytical subdivision into competence categories is necessary. Depending on the actual characteristics of the day-to-day work of the target group under consideration here, the categories and competences may already describe highly specialized personnel.

A second aspect concerns the rather selective choice of articles relating to RDM competencies that were used to conceptualize the competency framework. The goal of the literature review was not to synthesize all existing RDM competency literature into an overarching meta-framework but to implement the most purposeful work to describe curation-focused RDM activities in RDCs. Therefore, it is both possible and desirable that in the future, other authors and RDM experts might supplement this framework with further relevant works and RDM activity descriptions.

Overall, the competency framework based on the DCC DCLM provides a guideline for structuring newly to-be-developed training content in a manner that maximizes learning outcomes. Consequently, RDM Compas uses the framework as “blueprint” or fundament for the online platform, containing all the (prospective) service features provided by KonsortSWD. As a detailed elaboration of competencies required for different curation activities, this offers an excellent basis for the structured representation of these operations. All sections of the RDM Compas-platform are structured along the DCC DCLM and the respective competency profiles addressed in the framework. In doing so, the framework simplifies the formulation of learning objectives or goals and thereby supports development of appropriate RDM-courses and -trainings. Finally, the detailed framework on the processual competencies could be relevant, in perspective, for evaluating the courses (or course modules) or further respective research on competency development in the area of curation-oriented RDM.

In summary, this paper and the provided competency framework are both an innovative guide and a helpful tool in professional competency development for data curation.

Acknowledgements

This paper is a result of a subproject from KonsortSWD, which is funded by the German Research Foundation (DFG) within the framework of the NFDI (project number: 442494171). Pawel Nuwaltzew (Leibniz Institute for Psychology (ZPID)) has provided us with excellent support in the work on this paper with his media design skills. Many thanks!

References

Abbott, D. (2008). What is Digital Curation? DCC Briefing Papers: Introduction to Curation. Digital Curation Centre. Handle: 1842/3362. Retrieved from https://www.dcc.ac.uk/sites/default/files/documents/resource/briefing-papers/what-is-digital-curation.pdf* Accessed January 4, 2023).

Arguillas, F., Christian, T.-M., Gooch, M., Honeyman, T., Peer, L., & CURE-FAIR WG. (2022). 10 Things for Curating Reproducible and FAIR Research (1.1). Zenodo. doi:10.15497/RDA00074

Ball, A. (2012). Review of data management lifecycle models (version 1.0). Bath, UK: University of Bath.

Blask, K., Förster, A., Lemaire, M. & Minn, G. (2018). Forschungsprozessspezifische Kompetenzmatrix für die Einführung des Forschungsdatenmanagements (FDM): Arbeitspapier des BMBF-Projekts PODMAN.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). Handbook I: cognitive domain. David McKay. Retrieved from https://scholar.archive.org/work/l7a35bfkqjge3ictjyt4cb2fsi/access/wayback/https://www.uky.edu/~rsand1/china2018/texts/bloom%20et%20al%20-taxonomy%20of%20educational%20objectives.pdf

Demchenko, Y., Stoy, L., Engelhardt, C., & Gaillard, V. (2021). D7.3 FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework) (1.0 DRAFT). Zenodo. https://doi.org/10.5281/zenodo.4562089

European Commission. (2019). H2020 Programme. AGA – Annotated Model Grant Agreement. Retrieved from https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf*(Accessed April 20, 2022).

Greci, C., Welpton, R. & Woods, C. (July 2018). The Safe Data Access Professionals (SDAP) Competency Framework. Retrieved from https://securedatagroup.files.wordpress.com/2018/07/sdap_competency_framework-01_00.pdf

Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 3(1), 134–140. doi:10.2218/ijdc.v3i1.48

Kim, J., Warga, E. & Moen, W. (2013). Competencies Required for Digital Curation: An Analysis of Job Advertisements. *International Journal of Digital Curation*, 8(1), 66–83. doi:10.2218/ijdc.v8i1.242

Petersen, B., Engelhardt, C., Hörner, T., Jacob, J., Kvetnaya, T., Mühlichen, A., Schranzhofer, H., Schulz, S., Slowig, B., Trautwein-Bruns, U., Voigt, A. & Wiljes, C. (2022). Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards. doi:10.5281/zenodo.7034478

Putnings, M., Neuroth, H. & Neumann, J. (2021). Praxishandbuch Forschungsdatenmanagement. De Gruyter. doi:10.1515/9783110657807

Sapp Nelson, M. (2017). A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs. *Journal of EScience Librarianship*, 6(1), e1096. doi:10.7191/jeslib.2017.1096

Scholtens, S., Jetten, M., Böhmer, J., Staiger, C., Slouwerhof, I., van der Geest, M., & van Gelder, C. W.G. (2022). Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise (Version 4). Zenodo. doi:10.5281/zenodo.7225070

Schüller, K. (2020). Future Skills: a Framework for Data Literacy - Competence Framework and Research Report. doi:10.5281/zenodo.3946067

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 160018. doi:10.1038/sdata.2016.18

Additional References for the Competence Framework (Appendix A)

CESSDA Training Team (2020). CESSDA Data Management Expert Guide. CESSDA ERIC, Bergen, Norway.

Corti, L., van den Eynden, V., Bishop, L.; Woollard, M. (2014). Managing and sharing research data. A guide to good practice. Los Angeles, London, New Delh, Singapore, Washington DC: SAGE. Retrieved from https://www.ukdataservice.ac.uk/manage-data/handbook

Curdt, C., Dierkes, J., Helbig, K., Lindstädt, B., Ludwig, J., Neumann, J., Parmaksiz, U. (2021). Data Stewardship im Forschungsdatenmanagement - Rollen, Aufgabenprofile, Einsatzgebiete. doi:10.17192/bfdm.2021.3.8347

Engelhardt, C., Biernacka, K.; Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., ..., Zhou, B. (2022). D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions. doi:10.5281/zenodo.6674301

Lee, D. J. & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. In: PloS one 12 (3), e0173987. doi:10.1371/journal.pone.0173987

Ludwig, J., Enke, H. (Hg.) (2013). Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Glückstadt: VWH Verlag Werner Hülsbusch Fachverlag für Medientechnik und -wirtschaft.

Appendix

Appendix A. A Competence Framework for Data Curation

	Label	Description	Knowledge	Skills	Attitude
Data	Data	(Higgins, 2008) • Data, any information in binary digital form, is at the centre of the Curation Lifecycle. This includes: ▪ **Simple Digital Objects**: discrete digital items; such as textual files, images or sound files, along with their related identifiers and metadata. ▪ **Complex Digital Objects**: discrete digital objects, made by combining a number of other digital objects, such as websites. • Databases: structured collections of records or data stored in a computer system.	• Knowledge about common data types and structures. • Knowledge about kinds of digital objects (simple, complex). • Knowledge about what databases are and how they can be used. (Blask et al., 2018) • Knowledge about best practices regarding data structures, types and formats. (Greci et al., 2018) • Understanding of different (safe) data settings required for different types of data	(Greci et al., 2018) • The skill to practically support analysts of various data types (e.g. to answer data queries, to carry out SDC of analysts' outputs). • The ability to perform basic operations on data (i. e. data manipulation and preparation) in the designated software environment.	• Willingness to investigate, process, and edit data in a structured manner • Willingness to learn about and deal with technical questions and issues in regard to data (Greci et al., 2018) • Confidence providing advice about the use of different data settings.
Full lifecycle actions	Description & Representation information	(Higgins, 2008) • Assign administrative, descriptive, technical, structural and preservation metadata, using appropriate standards, to ensure adequate description and control over the long-term. • Collect and assign representation information required to understand and render both the digital material and the associated metadata.	• Know appropriate standards for adequate description of data and its long-term control. (Blask et al., 2018) • Knowledge about (best practices of) description of data (structures) with controlled vocabularies, ontologies and metadata. • Knowledge of the challenges of working with large datasets, for example that coded data can be handled in bulk and open answers cannot • Knowledge of the research process and information needed for data interpretation and accountability	• The ability to choose and apply relevant standards for data description. • The skill to collect representation information and assign preservation metadata. (Blask et al., 2018) • Ability to provide consultation and training on commonly used data and data description methodologies. • The skill to use/provide services and infrastructures on commonly used data and data description methodologies (vocabularies, ontologies, metadata). (Lee & Stvilia, 2017) • The skill to optimize existing data for search engines (SEO) to enhance findability.	• Willingness to comply with relevant standards for data description, to choose and apply them. • Research integrity (be transparent and accountable)

	Label	Description	Knowledge	Skills	Attitude
Full lifecycle actions	**Preservation Planning**	**(Higgins, 2008)** Plan for preservation throughout the curation lifecycle of digital material. This would include plans for management and administration of all curation lifecycle actions.	• Knowledge about the discipline-specific features and requirements of the handled data. • Knowledge about storage and secure back-up strategies and how to implement them **(Blask et al., 2018)** • Knowledge of relevant hardware/storage and computing capacity/technical solutions for data preservation. • Knowledge of current RDM guidelines from third-party funders and funding bodies.	• The skill to actively develop and follow an administrative / data management plan. **(Blask et al., 2018)** • The skill to select and assess suitable hardware/storage and computing capacity/technical solutions • To provide for / communicate with external service providers • The skill to assess possible software solutions for planned RDM activities. • The skill to follow and apply current RDM guidelines from third-party funders and funding bodies.	• Willingness and motivation to plan administrative processes in advance in a thorough manner, and follow the specified plans
	Community watch & participation	**(Higgins, 2008)** • Maintain a watch on appropriate community activities, and participate in the development of shared standards, tools and suitable software. **(Curdt et al., 2021)** • Networking: Public relations and community building (website, events), creating incentives, establishing information bridges between researchers and different infrastructural units or areas of the institution or external data management/specialist experts, stimulating interdisciplinary exchange of experience between researchers (e.g. in the form of interdisciplinary events). Stay up to date about current RDM-related developments and innovations.	• Know about the RDM community, i.e., which relevant information sources and channels to follow and where to gather and exchange relevant information. • Know about / be aware of new trends in the research field, for example, new data types, new funder requirements regarding OA, ... • Know consulting and training network structures to stay informed about current developments. • Knowledge about technical innovations / developments in the area of FDM. **(Blask et al., 2018)** • Know and use RDM info and training material in order to acquire and apply relevant knowledge and skills.	• Ability to actively participate in community activities • Ability to choose and use appropriate consulting and training materials for one's specific goals	• Willingness to engage in networking and interdisciplinary exchange • Willingness to consider the needs of the research community • Willingness to reach out to institutions and individuals one was previously unfamiliar with. • Willingness to advocate for Open Science and FAIR data sharing principles

	Label	Description	Knowledge	Skills	Attitude
Full lifecycle action	Ethical and legal aspects	• Know and address personal data protection, copyright, research data licensing, and confidentiality throughout the curation lifecycle. (Blask et al., 2018) • Stay informed about current developments in relevant information and copyright law and ethical frameworks.	• Know which criteria characterise data that is particularly worthy of protection • Knowledge of data protection laws and which ones apply, also follow-up with developments/changes in the law, interpretations, court decisions etc. • Know which legal bases apply and how to check whether their requirements are fulfilled (often informed consent but there can be other rules that apply, for example Sozialgesetzbuch) (Schüller, 2020) • Knowledge of which data protection and data security requirements are necessary for the respective data situation. • Knowledge of which statistical and technical possibilities exist to protect and secure data. • Know characteristics of protection-worthy data, know security requirements for given data, know technical & statistical possibilities, detect violations of data protection.	(Schüller, 2020) • The skill to apply data protection and data security requirements in a specific situation. • The skill to implement data protection and data security by analogy if no clear regulation exists for a situation. • The skill to apply necessary anonymization and pseudonymization procedures to ensure compliance with law.	(Schüller, 2020) • Willingness to be guided by data protection and data security as well as ethical principles, even if there are no clear rules for the specific situation, or compliance with these rules is not checked. • Understanding data protection, data security and informational self-determination as values. • Willingness to follow the current development of new technologies for data management and to train independently. (Nelson, 2016) • The willingness to comply with institutional, disciplinary, and state and federal laws and regulations governing data.
Sequential Actions	Conceptualise	(Higgins, 2008) • Conceive and plan the creation of data, including capture method and storage options.	• Extensive knowledge of the research process, which serves as a preliminary and essential foundation for the curation process. • Knowledge about dataset structure and data formats	• Knowing which processes are important in which subsequent step of the curation life cycle • The ability to develop a course of action for a specific data curation task / choose relevant actions for an upcoming data curation task • The skill to anticipate possible problems in the course of a data curation task and their solutions • The ability to provide consulting for and interact with researchers.	• The ability to keep track of the conceptualization of the curation process and its sub-steps • The willingness to address potential problems early on in a solution-oriented manner

	Label	Description	Knowledge	Skills	Attitude
Sequential Actions	Conceptualise RDM Services for researchers	(Corti et al., 2014) • Discovery and planning. • Data collection. • Data processing.	• Define the research question. • Identify the data that will be collected or used to answer research question. • Be aware of the FAIR-principles. • Be aware of ethical and legal standards for your (discipline specific) research standards and prepare your research accordingly in the best possible way. • Knowledge about and consultation for writing a DMP (CESSDA Training Team, 2020) • Be aware of common metadata standards and their value.	(Corti et al., 2014) • Designing research. • Planning data management. • Planning consent for sharing. • Planning data collection, processing protocols and templates. • Finding and discovering existing data sources.	
	Create or Receive	(Higgins, 2008) • Create data including administrative, descriptive, structural and technical metadata. Preservation metadata may also be added at the time of creation. • Receive data, in accordance with documented collecting policies, from data creators, other archives, repositories or data centres, and if required assign appropriate metadata. (Schüller, 2020) • Identify data sources, evaluate their relevance, accessibility and usability. Access data portals, make queries/ enquiries. (Greci et al., 2018) • Data acquisition. • Quality assurance.	(Higgins, 2008) • Know appropriate standards, documented guidance and legal requirements. (Schüller, 2020) • Knowledge of possible data sources. • Knowledge of the characteristics of data based on the data collection method (e.g. survey, observation, measurement, interview). Knowledge of the new types of data sources created by digitisation ("Big Data"). • Knowledge of which standards and best practices are common in the respective field when data is collected, procured, and stored. (Greci et al., 2018) • Good understanding of analysis and analysts. • Good understanding of metadata. Reviews existing metadata, identifies gaps and creates new metadata where required.	• The skill to find and acquire appropriate data. (Schüller, 2020) • The skill to select and use appropriate strategies and for this purpose, when needed. (Greci et al., 2018) • The ability to support colleagues in negotiating for access to less complex data sources: to collect the relevant information (e. g. reasons why access is required, key legal gateways); to present a persuasive business case; to liaise confidently with analysts and data suppliers (F2F, telephone and email), and to ensure that key messages are understood and good relationships maintained. • The skill to investigate data quality.	• Willingness to actively contact institutions (as possible data sources) for data acquisition • Willingness to advise and cooperate with researchers and analysists. (Greci et al., 2018) • Appreciates what analysis is, its aim and why some analysis requires data of different levels of detail (from previous analysis experience).

	Label	Description	Knowledge	Skills	Attitude
Sequential Actions	Create or Receive RDM Services for researchers	• Create data and (preservation) metadata; adhere to documented guidance, policies, legal requirements.	• Knowledge about best practices in data documentation • Knowledge about tools for data documentation	(Corti et al., 2014) • Collecting data – recording, observation, measurement, • experimentation and simulation. • Capturing and creating metadata. • Acquiring existing third-party data. • Entering data, digitizing, transcribing and translating. • Checking, validating, cleaning and anonymizing data where necessary. • Deriving data. • Describing and documenting data. • Analysing data.	• Appreciation for Open Science and FAIR principles the importance of data documentation
	Appraise & Select (or Dispose)	(Higgins, 2008) • Evaluate data and select for long-term curation and preservation. Adhere to documented guidance, policies or legal requirements. • Dispose of data which has not been selected for long-term curation, securely destroy data (Schüller, 2020) • Check data quality, verify data. • Check the data quality regarding various criteria (correctness, relevance, representativeness, completeness). Documents the audit systematically. (Nelson, 2016) • Critique data sets based upon acknowledged best practices of quality control appropriate to their discipline.	• Knowledge about relevant appraisal / selection criteria • Knowledge of critical data that is subject to special requirements with regard to ethical and legal conditions (Ludwig & Enke, 2013) • Knowledge of the purposes, methods, and criteria for selecting data for further preservation. (Schüller, 2020) • Knowledge about possible quality problems and ways of dealing with them. • Knowledge about possible causes of data quality problems, e. g. in the form of collection, transformation, or data protection. • Knowledge of criteria for the evaluation of data quality.	(Schüller, 2020) • The skill to systematically identify quality problems, errors and gaps in the data and to remedy them appropriately. • The ability to provide consultation for researchers regarding data selection in the case of using secondary data.	• Willingness to reject further processing of data that does not meet the specified selection criteria. • Willingness to communicate quality issues to data provider. • Willingness to communicate the selection criteria transparently to the data provider and to justify the corresponding decisions. (Schüller, 2020) • Willingness to actively think through and communicate the possibilities and limits of working with provided data, and its possible transformations.

	Label	Description	Knowledge	Skills	Attitude
Sequential Actions	Appraise & Select (or Dispose) ← RDM Services for researchers		(Ludwig & Enke, 2013) • Transparent selection of data and its documentation.	(Corti et al., 2014) • Capturing and creating metadata. • Describing and documenting data.	
	Ingest	(Higgins, 2008) • Transfer data to archive/repository/ data centre, adhere to documented polities (Schüller, 2020) • Automatically reads data in various formats, integrates it and documents the integration	(Schüller, 2020) • Knowledge of how data can be integrated into existing infrastructures. (Ludwig & Enke, 2013) • Knowledge of the usual procedures necessary for transferring research data for long-term preservation, e.g., to a data archive. • Knowledge of commonly used checks, homogenizations, data enrichment procedures.	• The skill to implement necessary steps to integrate data into an existing infrastructure. • The skill to transform data to common formats suitable for long-term preservation.	• Willingness to adhere to documented policies in a thorough manner. • Willingness and ability to document the procedures.
	Preservation action (Reappraise, Migrate)	(Higgins, 2008) • Undertake actions to ensure long-term preservation and retention of the authoritative nature of data. Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its integrity. Actions include data cleaning, validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures or file formats. (Schüller, 2020) • Cleans data, corrects errors, imputes missing values, standardises, and transforms data, filters relevant data for a given question, links data	(Schüller, 2020) • Knowledge about possible quality problems and ways of dealing with them. • Knowledge about possible causes of data quality problems, e.g., in the form of collection, transformation, or data protection. • Knowledge of criteria for the evaluation of data quality. • Knowledge about possible technologies and tools for transforming data. • Knowledge about data formats and other characteristics of data (e.g., scale levels, degree of structuring). • Knowledge about possibilities to transform data by standardising or linking it. • Knowledge about how transformation processes change the information content of data	• The skill to systematically identify quality problems, errors and gaps in the data and to remedy them appropriately. • The skill to transform data e.g., by standardisation, recoding, aggregation, combination, or other methods. • The skill to link different data sources. • The skill to identify redundant information in the data and to eliminate redundancies. • The skill to provide data in various forms or aggregated data for specific purposes. • The skill to anticipate future or alternative uses of the data (Schüller, 2020) • The skill to anticipate future or alternative interpretations of the data and its transformations.	(Schüller 2020) • Willingness to actively think through and communicate the possibilities and limits of working with the collected data and its possible transformations. • Willingness to invest time and effort in elaborate and unexciting processing steps to increase the relative value of the data. • Understanding of the value of data; understanding that storage space and analysis capacities are scarce resources and willingness to use them sparingly, e.g., in terms of working time and energy consumption. • A basic attitude of expediency and user orientation, i.e., adherence to the effectiveness of the work steps (e.g., when imputing missing values)

	Label	Description	Knowledge	Skills	Attitude
Sequential Actions	Store & Secure	(Higgins, 2008) • Store the data in a secure manner adhering to relevant standards.	(Ludwig & Enke, 2013) • Knowledge about relevant storage options (possibilities, storage media) / Knowledge of current technical developments and offers from experienced storage service providers. • Knowledge about relevant file protection technologies (passwords, encoding) / procedures that minimize the chances of data loss during long-term storage. (Schüller, 2020) • Knowing which criteria characterise data that is particularly worthy of protection. • Knowledge of which data protection and data security requirements are necessary for the respective data situation. • Knowledge, which statistical and technical possibilities exist to protect and secure data.	(Schüller, 2020) • The skill to apply data protection and data security requirements in a specific situation. • The skill to implement data protection and data security by analogy if no clear regulation exists for a situation.	(Schüller, 2020) • Willingness to be guided by data protection and data security as well as ethical principles, even if there are no clear rules for the specific situation, or compliance with these rules is not checked. • Understanding data protection, data security and informational self-determination as values. • Willingness to force information loss for ethical or legal reasons. • Willingness to follow the development of new technologies for data management and to train independently.
	Access, Use & Reuse	(Higgins, 2008) • Ensure that data is accessible to both designated users and reusers, on a day-to-day basis. This may be in the form of publicly available published information. Robust access controls and authentication procedures may be applicable. (Greci et al., 2018) • Support. • Infrastructure for secure data access for very sensitive data, like secure data center / safe rooms, remote access	• Knowledge of common methods for increasing the discoverability of data. • Knowledge of legal requirements and restrictions on access to data. • Knowledge about relevant technological implementations of access controls, secure transfer, and authentication procedures.	• The skill to allocate persistent identifiers. • The skill to implement adequate authentication procedures. • The ability to manage contracts in regard to the access, use and reuse of data • The ability for output control (Lee & Stvilia, 2017) • The skill to optimize data for search engines (SEO) to enhance findability. (Greci et al., 2018) • The skill to answers a range of data queries, including more complex ones (by reviewing the documentation, analysing the data and/or contacting data suppliers).	• Understanding of FAIR Principles, open data, findability, accessibility and reusability as values and important goals in the data management and curation processes • Willingness to increase the findability, accessibility and reusability of data and research results • Ability to balance the benefits of accessibility, reusability and open data with legal and ethical data protection requirements (Nelson, 2016) • Willingness to explain the potential problems that could arise for new research because of data management decisions that were enacted in the pre-existing research data set.

	Label		Description	Knowledge	Skills	Attitude
Sequential Actions	Access, Use & Reuse	RDM Services for researchers	(Engelhardt et al., 2022) • Being able to reuse data and analyse secondary data can not only save time and energy for researchers, it can also fast track scientific discoveries with shared resources and perspectives, while adhering to the FAIR principles.	(Engelhardt et al., 2022) • Know the importance of data discovery and reuse. • Recognise the concept of "secondary data" vs collecting primary data. • Know how to discover published data sets in one's own discipline. • Know how to correctly cite data. • Know how to develop a strategy to search for data	(Engelhardt et al., 2022) • Can recognise the provenance of data they intend to use. • Can recognise the importance of the terms and conditions of data reuse. • Can recognise the importance of data citation when reusing data.	(Nelson, 2016) • Justify selection of data based upon the research topic or data types specified in the research design.
	Transform		(Higgins, 2008) • Create new data from the original, for example: • By migration into a different format. • By creating a subset, by selection or query, to create newly derived results, perhaps for publication.	• Knowledge about commonly used data transformation purposes, procedures and workflows • Knowledge about the specifics of statistical programs and how this affects data transformation	• The skill to transform data to common formats suitable for long-term preservation.	• Willingness to extract value from data through migration and transformation procedures (Schüller, 2020) • Willingness to actively think through and communicate the possibilities and limits of working with the collected data and its possible transformations. • Willingness to invest time and effort in elaborate and unexciting processing steps to increase the relative value of the data.
	Transform	RDM Services for researchers	• In case of use of secondary data, data might need a transformation process to fit for own research purposes.	• Know the technical conditions for data transformation. • Know about the legal conditions for data transformation.	• Be able to match data from different sources or to combine primary and secondary data.	