IJDC | *Conference Paper*

Professionalising Social Science Research Data Management with RDM Compas

Anja Perry
GESIS – Leibniz Institute for the Social Sciences

Ekaterina Chicherina
GESIS – Leibniz Institute for the Social Sciences

Abstract

It is not only since the publication of the FAIR principles, requiring the **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable organisation of research outputs, that the demands placed on researchers when conducting good scientific practices have increased. Universities and other research institutions have responded to this and offer increasing support, especially regarding research data management (RDM). This is to support their researchers in the FAIR organisation and thus transparent sharing of all their research outputs. However, digital curation and RDM are complex and demand an interplay of various skills and knowledge. So far, there are no established training programmes. Therefore, projects aiming to support researchers in RDM often face one major challenge: recruiting staff well trained in RDM.

In our project, we examined data curation skill needs and various approaches to define these needs. Paired with the DCC Curation Lifecycle Model and survey results from social science research data centres in Germany, we defined curation skill needs and learning trajectories and implemented these in an information and training platform for data curation in the social sciences, RDM Compas, with its content organised along the DCC Curation Lifecycle Model. A newly developed data curation module in a nationwide RDM certification course complements the training path toward professional RDM.

Submitted 30 January 2025 ~ *Accepted* 20 February 2025

Correspondence should be addressed to Anja Perry, email: anja.perry@gesis.org

This paper was presented at the International Digital Curation Conference IDCC25, 17-19 February 2025

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

It is not only since the publication of the FAIR principles (Wilkinson et al., 2016) that the demands placed on researchers when conducting good scientific practices have increased. The skills needed to do research expanded from pure research skills, ranging from the collection of data over their analysis to the description of the derived conclusions in a publication, to include additional skills in responsible handling of research data. These research data management (RDM) skills include aspects like IT, legal and ethical aspects, metadata, standards, and funding policies (Schultes, 2020). Universities and research institutions have responded to this and offer increasing support for researchers to fulfil these responsibilities (Böhmer, 2019; Cyra & Fingerhuth, 2022; Johnston et al., 2024; Teal et al., 2015; Wilson & Jeffreys, 2013). While the skill sets for digital curation and RDM are complex and demand an interplay of various skills and knowledge (Kim et al., 2013), the various projects aiming to support researchers in RDM often face one major challenge: recruiting staff well trained in RDM (Tang & Hu, 2019).

The increased demand for RDM support was also noticed in Germany (Pampel et al., 2010). Especially, the German social sciences experienced an early push in RDM, mainly coming from a network of European Social Science Data Archives (CESSDA).[1] CESSDA supported first the staff at data archives and soon established RDM training courses and resources for researchers[2] to ensure that their data meets the requirements for being published and reused. Like the data archives on the European level, starting in 2001, the German Data Forum[3] established a German social science research data infrastructure with a network of research data centres (RDCs).[4] Today, there are more than 40 RDCs throughout the country, representing a strong group of employers for RDM staff. Their focus is mainly on making highly sensitive data available for research that would otherwise be inaccessible.

Historically, data librarians have played a minor role in these data archives and RDCs. Instead, RDM professionals are often recruited from among university graduates or from research itself, for example, among PhD graduates. Especially researchers bring with them profound research experience and often a passion for data-related work. Common to both groups is that they are not specifically trained in data management and data curation. As RDM is not yet widely included in the social science university curricula, many skills and most of the knowledge, especially about metadata, standards, and archival practices, must be acquired on the job.

This makes recruitment for RDM positions difficult. There is no apparent career path for young graduates to enter this field, which makes it not an obvious career choice for them and difficult for recruiting institutions to reach potential candidates. The required skill sets are very diverse, and even the job titles vary a lot, making it difficult for candidates to identify open positions in the RDM field (Kim et al., 2013; Newbold et al., 2024).

At the same time, the demands on data curators are constantly changing, due to a growing amount of data, new types of data, and changing technologies. Research-supporting infrastructures, therefore, need to not only train new staff on the job, but also ensure continuous education for their existing staff to keep up with the latest

[1] https://www.cessda.eu/
[2] https://dmeg.cessda.eu/Data-Management-Expert-Guide
[3] https://www.konsortswd.de/en/about/ratswd/
[4] Social science RDCs in Germany provide access to sensitive research data in secure environments. They are departments of research institutions or government agencies that process and provide data collected or recorded at their own institution and often also from researchers and institutions from in the same field: https://www.konsortswd.de/en/services/research/all-datacentres/

developments. Hence, there is a growing interest in the development of training offers for RDM professionals.

Professional Competences for Data Curation

Data literacy and RDM competences are fundamental for successful data sharing. That is why they are focused on by many initiatives aimed at the development of sustainable RDM within the German scientific community. Different projects are working on identifying RDM needs and competencies for various target groups (cf. Day et al., 2023; Pampel et al., 2010; Petersen et al., 2023).

Most projects cover generic knowledge by combining data competences and activities for several target groups. In contrast, the development of tailor-made offers is hindered by the heterogeneous needs and requirements of different target groups and the lack of sustainability and short-term nature of most projects (Lehmann et al., 2023). The main target group of most initiatives are researchers and other data scientists. Recently, there have emerged additional training offers for Bachelor's and Master's students (Petersen et al., 2023), since research data form the basis of most Bachelor's and Master's theses (Mertzen et al., 2023).

The research landscape has become increasingly complex in tandem with increasing demands for researchers regarding data management and data sharing, often also involving legal questions. This raises the need for institutional support for researchers (Davidson, 2015). This support can be provided by university libraries, data management services at universities, or research data infrastructures outside of universities.[5] Hence, besides trained researchers, "equally important is a well-trained scientific data curation workforce that is capable of solving problems in an increasingly complex, data-driven world" (Kelly et al., 2013, p. 256). While researchers as well as data curators agree about the vital role of ensuring data quality, thus contributing to a better data sharing process (Johnston et al., 2024), there is an overall lack of the sort of generic and data type-specific training targeted at employees in data curation and data stewardship that would ensure a systematic and sustainable learning process. This is not surprising, as the roles, tasks, and work areas of newly-emerged positions, such as data stewards, are often not clearly defined and are not standardised (Curdt et al., 2021).

Typical data curation activities include discussions with data providers, examination of research data (checking for duplicate files, reviewing documentation, reviewing metadata, etc.), and sometimes even development of the data governance framework. In addition to these tasks, data curators at institutional repositories often have to perform tasks that overlap with task areas of other repository staff—sometimes due to lack of resources for more staff—for example, developing the repository's services, marketing them, engaging with data providers and researchers, trying to meet their needs and directing them to an appropriate contact (Federer, 2018; Johnston et al., 2024; Lee & Stvilia, 2017). This may lead not only to work overload but also to a need for additional skills, such as project management skills and communication and marketing skills. A lack of resources in the form of certain abilities or competencies is seen by employees as one of the main limitations in their data curation work (Johnston et al., 2024). This results in a skills profile that is closer to that of a generalist rather than a specialist (Federer, 2018).

A first step towards a skills profile for data curation professionals is an inventory of data curation tasks. The Data Curation Lifecycle model provides a structured presentation of the research data curation process, including the following sequential steps: Store & Secure; Access, Use & Reuse; Transform; Create & Receive; Appraise and Select; Ingest (or Dispose); Preservation Action (Reappraise, Migrate), and is oriented particularly towards

[5] https://www.forschungsdaten.org/index.php/FDM-Kontakte

data curators (Higgins, 2008). Focussing on skills, Petersen et al. (2023) offer a learning objective matrix for RDM in general that encompasses generic learning objectives for four target groups: Bachelor's students, Master's students, PhD candidates, and data stewards. Their matrix addresses the following competence areas: professional competence, methodological competence, self-competence, and social competences. According to the offered matrix, the competence area of data stewards should include all competences and skills proposed under the learning goals of the other three target groups.

To ensure a complete picture of skills needed in data curation, Behrens and Kvetnaya (2023) took further publications focusing on data management for data stewards and data repositories into account. More specifically, they incorporated work by Blask et al. (2018), who define relevant RDM competences along the entire research process. This matrix serves two essential functions: 1) it defines the tasks necessary for the implementation of a comprehensive institutional RDM concept; and 2) it specifies the associated competencies of the executing actors. Further competency models include Greci, Welpton, and Woods (2018), Sapp Nelson (2017), and Schüller (2020). In addition, Behrens and Kvetnaya (2023) incorporated handbooks, training materials, and best practice guides for data stewards and data repository staff, such as work by Corti et al. (2019), Curdt et al. (2021), Engelhardt et al. (2022), Lee et al. (2017), and Ludwig and Enke (2013). Doing so, the authors developed a competence matrix along the actions defined in the Data Curation Lifecycle and matched the necessary skills and knowledge to each curation task (Behrens & Kvetnaya, 2023). This work built the backbone for our project. The, to our knowledge, the most recent compilation of skills needed in a data curation profession is the Minimum Viable Skillsets (MVS, Whyte et al., 2024) by Skill4EOSC.[6] The MVS distinguish between data stewards and research infrastructure professionals (which can include data stewards) and list both technical skills and competences as well as soft and transversal skills. Unfortunately, this work could not be considered when deriving the competence model for our project in 2023.

Data Curation Skill Gaps and Needs

Regarding RDM skills development, KonsortSWD[7] focuses on the professionalisation of RDM. KonsortSWD is a consortium of the German National Research Data Infrastructure (NFDI)[8] and serves the social, behavioural, educational, and economic sciences. The focus on professionalising RDM originated in the needs that the KonsortSWD's RDCs identified when hiring professional RDM staff. The measure "Developing and exchanging research data management skills" initially runs for five years from October 2020 until September 2025 (Adena et al., 2020). The measure pursues three components on its way towards a professional RDM:

1. An online knowledge base that offers resources for knowledge acquisition to RDM professionals in the social sciences.

2. A training platform that collects and offers training courses targeted to the group of social science RDM professionals.

3. A certification for acquiring the necessary skills and knowledge for working as a social science RDM professional.

Our measure started out with a survey on the skills that supervisors in German social science RDCs perceived as highly relevant but often lacking in applicants to their RDCs.

[6] https://www.skills4eosc.eu/
[7] KonsortSWD is the social science consortium within the NFDI, https://www.konsortswd.de/en/
[8] https://www.nfdi.de/?lang=en

KonsortSWD supports a network of 41 RDCs.[9] The supervisors of all RDCs were invited to take part in our survey. Twenty-five of the 41 RDCs responded.

Besides asking about which skills are relevant for working at their RDC, we also asked which skills job candidates bring with them when applying to their RDC. We define a specific skill as lacking when ten out of 25 RDCs state that it is relevant but not sufficiently or only partly sufficiently existent in job candidates.

Lacking but relevant skills are, strikingly, almost all RDM-specific skills that were presented in the survey (Figure 1). These are especially legal aspects, such as data protection law and anonymisation, given the sensitive nature of KonsortSWD's data. But also, contract administration, data security, and metadata handling are among the most underdeveloped skills. An exception among RDM-specific skills is subject indexing, which also only a small number of RDCs consider to be a relevant skill. Two further skills that are highly relevant when working in an RDC are data checking and data processing. The view about whether these two skills are available in job candidates is mixed: While 11 RDCs state that they are not sufficiently or only partly sufficiently existent in job candidates in each case, 11 and 12, respectively, say that job candidates do have sufficient skills in these areas (Table A1).

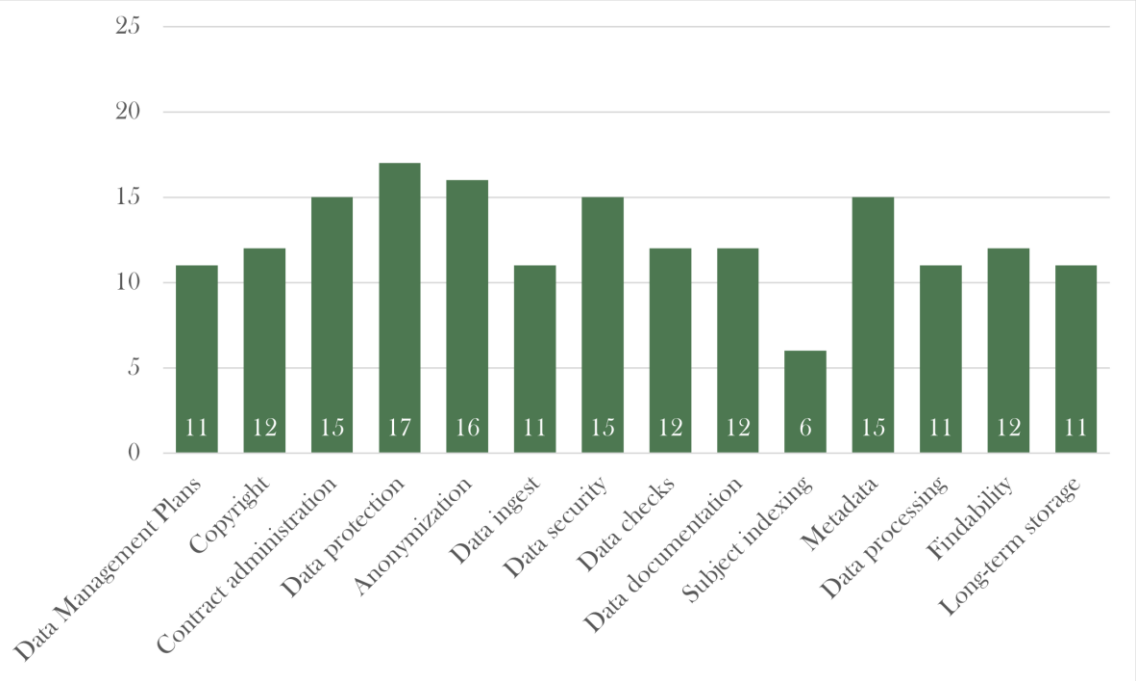

Figure 1. Number of RDCs that state that an RDM-specific skill is relevant but not sufficiently or partly sufficiently existent in job candidates.

Note: *N* = 25, number of cases, data collection in December 2021 and January 2022.

We also asked about skills beyond those specific to RDM, in the categories of research method skills, software and IT skills, and other skills. For these skill groups, we will here focus only on those skills that more than ten RDCs indicated as relevant here (Figure 2). For those relevant skills, considerable gaps are evident in R programming and in project management. Seventeen and 13 RDCs, respectively, say that these skills are often lacking. Further gaps, indicated by fewer (seven to nine) RDCs, are research planning skills, knowledge about quantitative data collection, data handling with Stata, consultation skills,

[9] https://www.konsortswd.de/en/services/research/all-datacentres/

and knowledge about research funding. All other skills are either not relevant for the RDCs or they are mostly existent in job candidates (see Table A2 for a complete overview).

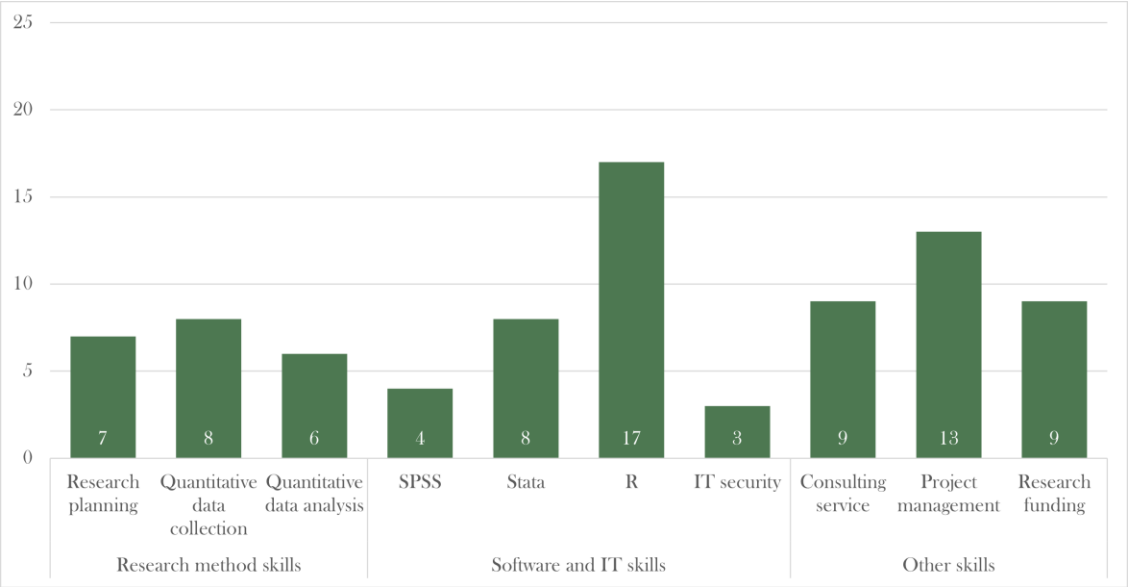

Figure 2. Number of RDCs that state that an additional skill is relevant but not sufficiently or only partly sufficiently existent in job candidates.

Note: *N* = 25, number of cases, only skills that at least ten RDCs indicate as relevant, data collection in December 2021 and January 2022.

In a further step, we asked about ways to decrease gaps in RDM-specific skills. One way to achieve this would be by providing opportunities to acquire these skills via a certified course, outside of learning on the job or through studies. The majority of RDC supervisors who responded (18 of 25) said that certified training in professional RDM was necessary or would at least be helpful for the hiring process. We also asked RDCs who they thought should certify the acquirement of RDM-specific skills. KonsortSWD (to which the RDCs belong) scores highest, followed by online badges, the NFDI, and ECTS credits. Nationwide actors outside of academia, such as the Chambers of Commerce and Industry or the Federal Employment Agency (both provide and/or certify professional trainings), do not play a role for RDCs (Figure 3).

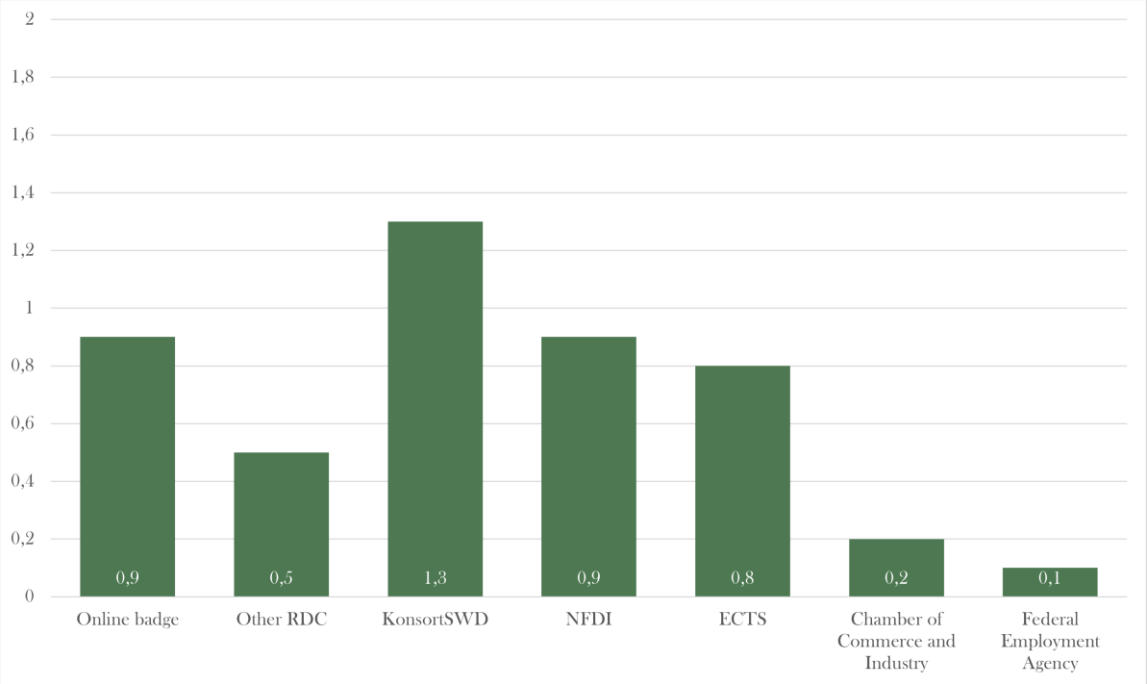

Figure 3. RDCs' opinion on who should certify the acquirement of RDM-specific skills.

Note: *N* = 20, scale from 0 to 2, derived from the response options irrelevant, helpful, and necessary, potential issuers of certifications ranked by hierarchy, data collection in December 2021 and January 2022.

RDM Compas—Professional Competencies for Research Data Curation

The goal of this KonsortSWD measure is to fill these skill gaps and provide training offers for existing RDM staff as well as potential applicants to social science RDCs and other employers in the RDM field. Its core element is RDM Compas,[10] an information and training platform, which aims at approaching RDM issues on a professional level, particularly competencies relevant or required for data curation activities at research data centres. To organise the information and training content on both platforms, Behrens and Kvetnaya (2023) suggest a competence model which is aimed at filling the gap in data curation competencies for FAIR and sustainable RDM. The authors base their competence model on previous work discussed in the professional competences section above. The competence matrix and both platforms are organised along the DCC Curation Lifecycle (Higgins, 2008), and therefore particularly address the needs of RDM professionals in their daily work processes. This is in contrast to other RDM training platforms that address researchers and are predominantly based on the research data lifecycle (Humphrey, 2006).

For the technical implementation of RDM Compas, we used WordPress[11] for general navigation through the information and the training platform, including a search function. Courses and material on the training platform are accessible via the learning platform Moodle.[12]

[10] https://rdm-compas.org/
[11] https://wordpress.org/
[12] https://moodle.org/?lang=en

Knowledge Base

One of the essential elements of RDM Compas is the Knowledge Base, an information platform which offers information on each step of the DCC Curation Lifecycle (Higgins, 2008). In addition to materials and information on generic research data management, the Knowledge Base also aims to give in-depth insight into the curation process in general and specifics of different data types produced or used in the social sciences, such as physiological data, corporate data, or social media data. This offer is suitable for job starters to gain an overview of curation activities at research data centres, as well as experienced curators wishing to gain further insight into a specific topic by following further links. RDM Compas therefore serves not only the onboarding process for new employees but also continuous needs of RDM staff to look up information, due to the emergence of new data types, changes in legal aspects, and technical changes.

The generic part of the Knowledge Base covers all topics relevant to comprehensive and sustainable data curation at RDCs, regardless of the data type or the discipline within the social sciences. The topics discussed encompass various aspects across all phases of the data curation life cycle, such as data acquisition, selection, ingest, etc., emphasising organisational and technical issues of each of these phases. Within some of these phases, for example, in appraisal and selection, the Knowledge Base also addresses legal considerations, such as data protection and contractual matters. In addition to general information, users will find specific examples and cases from actual practice. For example, they can access contract samples, a topic RDCs often request help with.

The second part of the Knowledge Base is dedicated to specifics of research data curation dependent on the data types that curators at RDCs typically work with. RDCs work with various research data types, which strongly vary both across and within each social science discipline. Most of the currently available information and training services are differentiated according to subject-specific characteristics and do not emphasise data curation nuances related to different types of data. RDC staff with a specific disciplinary or methodological focus have gained specialised expertise and offer this to researchers who wish to submit data to their RDC.

A prominent example is Qualiservice,[13] which offers support in managing qualitative data (Blätte et al., 2022). Although there are basic steps in the data curation process related to all data types, with some curation specifics shared between different data types (e.g., qualitative data and big data: Mannheimer, 2024), still there are certain curation procedures that are typical or essential for certain data types. For instance, ethical and legal challenges may be more difficult to solve for qualitative data and therefore need constant adjustment and more attention, especially when data was collected from vulnerable social groups (Mannheimer, 2024). Unlike quantitative data, which are often numerical or categorical and can be easily organised, qualitative data often consists of interviews, observations, or texts and may also be available in other formats, such as audio, video, or photographic data. Therefore, RDM of qualitative data requires specific approaches in relation to metadata standards, the implementation of appropriate storage and archiving solutions, the use of specialised tools for data organisation and anonymisation. Typically, consultation and guidance start well before the actual data collection. In addition, the concept of context plays a more central role in the documentation of the qualitative data to transmit thorough information to the researchers who plan to reuse the data (Heuer et al., 2020).

For corporate data, in turn, a particular feature is that the original purpose for which it was collected was not research. This might create specific challenges, such as the need for changes in the data structure, for example, when working with longitudinal data. Besides this, legal aspects play an important role in the curation of corporate data. Data curators must pay special attention to contractual agreements with all cooperation partners when

[13] https://www.qualiservice.org/en/

receiving and curating research data. These agreements often govern the handling of sensitive business data and define the rights and obligations for data use, storage, and sharing.

Thus, it makes sense that RDC employees benefit from information about data type-specific characteristics. Currently, the knowledge base offers information about five data types: qualitative data, corporate data, survey data, physiological data, and health data. In the longer term, this list will be expanded in accordance with general developments, such as the emergence of new data types, and particular interests of and requests from our stakeholders.

Training Center

While the Knowledge Base covers information needs, the RDM Compas Training Center provides further in-depth training for social science data curators. Like the Knowledge Base, the Training Center covers all aspects of the DCC Curation Lifecycle (Higgins, 2008) and is the result of a comparison between the identified competence model (Behrens & Kvetnaya, 2023) and what supervisors at social science RDCs specified as lacking in applicants for RDM positions. Available training courses are organised along the DCC Curation Lifecycle (Higgins, 2008). For example, the Ingest (or Dispose) phase encompasses courses on data pseudonymisation, as well as training on metadata schema and standards, tutorials for data documentation tools, and data curation training for data repository staff.

As our survey results highlighted a need for competences and skills beyond research data curation activities, the RDM Compas Training Center also curates and offers links to cross-disciplinary courses. These are, for example, trainings on project management, according to Figure 2, a competence that scores second in terms of being needed but often not available in job candidates.

The Training Center currently offers a variety of courses and will be expanded in the future. Example courses are introductory courses on all topics along the data curation life cycle and in-depth training courses on specific topics, such as introductory courses on software like R, or training on specific models and standards used for data archiving, including the DDI Standard. The courses are primarily designed as self-learning courses. In the future, there is an intention to offer courses in blended-learning formats as well. The materials used in the courses also vary and include videos and webinars, text documents, and PowerPoint presentations. Registration for the courses is free and available for everyone on the Moodle platform. A participant can register on the Training Center platform or access the courses as guest.

Our intent with the RDM Compas Training Center was not to create new training content but rather to offer navigation to existing trainings and training material for our target group of data curators. RDM Compas staff therefore research existing training offers. Whenever existing courses meet the required learning objectives, the Training Center links to them. RDM Compas therefore includes links to courses by external providers if they are relevant to the RDC community and data curator needs. All other content was either shared by the RDCs themselves, providing access to training material originally developed for their own staff, stems from other currently developed services within KonsortSWD, or is newly developed by the RDM Compas team in cases where the training needs could not be fulfilled with existing offers. Training materials can be easily submitted via the submission form on the RDM Compas website. After receiving a submission or after the RDM Compas staff research courses themselves, our team is responsible for evaluating the courses and materials. This means they assess the content's relevance for our target group, check whether its licence allows reuse on our platform, and decide into which step in the curation process the content fits best.

The submission form as well as the metadata for all other RDM Compas trainings follow the metadata schema for RDM training material developed by Biernacka et al.

(2020). Therefore, it facilitates easy integration and distribution of courses across various OER platforms, including other training services within NFDI.

Certification Course

One more step towards the establishment of professional and sustainable training for RDC staff is a certified course in data curation. We opted to integrate an advanced module, "Data curation and data management of sensitive data in research data centres (RDCs)", into the existing certificate course of the TH Köln, "Research data management", instead of developing a full, stand-alone course. The existing certificate course is aimed at staff at academic infrastructures, such as university libraries, data centres, and funding organisations, and covers all scientific disciplines. Our new module is particularly aimed at those wanting to work at research data centres, which deal with the curation of highly sensitive data from different social science fields. Participants on the course choosing this module will learn about the processes of data ingest and data publication, the legal and documentary requirements of data publications, and best practices in these areas (Zentrum für Bibliotheks- und Informationswissenschaftliche Weiterbildung (ZBIW) et al., 2024).

Implementing our module into an existing certification course has two advantages. First, we avoid developing a competing offer to a training course that is already established and widely accepted by the community. A stand-alone course would not have attracted many participants when a more comprehensive certification course exists that covers a broader range of participants' current or potential job tasks. Second, we can offer two of the four highest-rated certification methods among RDCs (see Figure 3), namely: a course developed by KonsortSWD and taught by KonsortSWD instructors, interpreted by RDC supervisors as the strongest indication that job candidates have acquired the necessary skills; and the ECTS credits that participants earn by completing the whole certification course, the fourth-highest-rated certification method among RDCs. Hence, our module sends a strong signal to RDC supervisors that employees and job candidates who complete both the module and the whole certification course possess the necessary skills for working at their RDC.

Similar to the other two components of RDM Compas, the Knowledge Base and the Training Center, the module's programme is based on the results of the survey conducted among the RDCs about their demands and expectations in relation to the skills and knowledge of job candidates. These demands were compared with commonly suggested competence matrices in the field of research data management (see above, Behrens & Kvetnaya, 2023) and with the module handbook and learning targets of the already implemented modules in the certification course. Those topics related to research data curation specifically at RDCs that were either not included in the learning curriculum or not sufficiently addressed were implemented in our module. To cover the main tasks at an RDC, a substantial part of the RDM Compas module's learning objectives focuses on data acquisition, ingest, processing, curation, and dissemination. Regarding data acquisition and ingest, these aspects include the basics of the OAIS model (Consultative Committee for Space Data Systems, 2012), plausibility checks, documentation, and enhancing data for reuse, as well as long-term preservation and accessibility. Considerations of research data dissemination, especially with a focus on sensitive data, include user management and user contracts, as well as access controls and output controls.

Conclusion and Future Plans

RDM Compas serves an important cause: the education and training of data curators and data stewards in the social sciences, with a focus on processing and publishing sensitive

social science data. The skills needed for this type of work are not typically taught at university. At the same time, these skills are more and more in demand, partly due to a push by governments to make government and administration data openly available (Kim et al., 2013).

The core challenge remains informing more university graduates and early PhD candidates about this new career path. Partly, this is achieved with various Bachelor's- and Master's-level study programmes, as well as certification courses and other types of training in RDM.[14] For a field, such as social science RDCs, that recruit mainly from their own discipline, we still need to address students and PhD graduates from this discipline and educate them in the relevant skills and about this potential career option. We aim to do so during the last year of KonsortSWD's first funding period.

Further plans to accomplish up to the end of the first funding period are to promote RDM Compas and evaluate its usage, as well as to evaluate the advanced module "Data curation and data management of sensitive data in research data centres (RDCs)" in the RDM certificate course. This can only be an initial evaluation, as a longer time frame will be required to observe and evaluate any potential benefit for the hiring of data curators.

KonsortSWD is currently preparing a proposal for continued funding. Plans for the following funding period in terms of RDM skills development include expanding the scope of RDM Compas beyond data curators to all types of staff in professional social science RDM, including data stewards and persons consulting research projects in higher education institutions. We also plan to strengthen the role of RDM staff by offering coaching.

Data Availability Statement

Due to the small sample size and the potentially revealing information about employers and their opinion of job candidates, supporting data is not available in a data repository. Instead, we have listed all relevant frequencies and statistics in the supplement.

Acknowledgements

We thank Kathrin Behrens, Katarina Blask, Ruben Brück, Tatiana Kvetnaya, and Maria Ricci for their tremendous efforts put into RDM Compas.

RDM Compas is part of KonsortSWD, which is funded by the German Research Foundation (DFG) as part of the German National Research Data Infrastructure (NFDI), Project Number 442494171.

References

Adena, M., Aßmann, C., Bambey, D., Blask, K., Blätte, A., Bosnjak, M., Buck, D., Bug, M., Busch, A., Fräßdorf, A., Fuß, D., Goebel, J., Hollstein, B., Jungbauer–Gans, M., Kern, D., Klas, C.-P., Kretzer, S., Liebig, S., Mayer–Ahuja, N., … Zindler, S. (2020). *Consortium for the Social, Behavioural, Educational, and Economic Sciences (KonsortSWD)*. doi:10.5281/ZENODO.4446457

[14] See for example: https://forschungsdaten.info/praxis-kompakt/fdm-professionalisierung/

Behrens, K., & Kvetnaya, T. (2023). *A Competence Matrix for Data Curation*. IASSIST 2023, Philadelphia, PA. doi:10.5281/ZENODO.8013777

Biernacka, K., Danker, S. A., Engelhardt, C., Helbig, K., Hendriks, S., Jacob, J., Jagusch, G., Lanza, G., Leone, C., Meier, K., Neumann, J., Odebrecht, C., Peters, K., Rehwald, S., Rex, J., Senft, M., Strauch, A., Thiemann, K., Trautwein–Bruns, U., … Ziedorn, F. (2020). *Metadatenschema für Schulungsmaterialien zum Thema Forschungsdatenmanagement* (Version 1.0). Zenodo. doi:10.5281/ZENODO.3760398

Blask, K., Förster, A., Lemaire, M., & Minn, G. (2018). *Forschungsprozessspezifische Kompetenzmatrix für die Einführung des Forschungsdatenmanagements (FDM)*. Universität Trier. doi:10.25353/UBTR-2751-5387-20XX

Blätte, A., Fräßdorf, A., Heuer, J.–O., Hoffstätter, U., Leonhardt, C., Menze, L., Miller, B., Mozygemba, K., Pattloch, D., Rakers, J., Reineke, S., Runge, T., Schlücker, F., Schmidt, T., Wenzig, K., & Wolf, C. (2022). Eine Dateninfrastruktur für die Gesellschaftswissenschaften. Unterstützung in der Arbeit mit Forschungsdaten durch KonsortSWD. *Zeitschrift Für Bibliothekswesen Und Bibliographie* 69(1–2), 48–58. doi:10.3196/1864295020691276

Böhmer, J. (2019). *Data Stewardship in the Netherlands 2019*. GO TRAIN Workshop, Hamburg, Germany. doi:10.5281/ZENODO.3552065

Consultative Committee for Space Data Systems. (2012). *Reference model for an Open Archival Information System (OAIS)*. Retrieved from https://public.ccsds.org/Pubs/650x0m2s.pdf

Corti, L., Van den Eynden, V., Bishop, L., & Woolard, M. (2019). *Managing and sharing research data: A guide to good practice* (2nd ed.). Los Angeles, CA: SAGE Publications.

Curdt, C., Dierkes, J., Helbig, K., Lindstädt, B., Ludwig, J., Neumann, J., & Parmaksiz, U. (2021). Data Stewardship im Forschungsdatenmanagement—Rollen, Aufgabenprofile, Einsatzgebiete. *Bausteine Forschungsdatenmanagement* 3, 70–81. doi:10.17192/BFDM.2021.3.8347

Cyra, M., & Fingerhuth, M. (2022). FDM-Landesinitiativen und NFDI. In V. Heuveline & N. Bisheh (Eds.), *E-Science-Tage 2021: Share Your Research Data*. doi:10.11588/HEIBOOKS.979.C13736

Davidson, J. (2015). Fostering Open Science Practice through Recognising and Rewarding Research Data Management and Curation Skills. In C. Bitso & R. Raju (Eds.), *LIS education and research in a dynamic information landscape: Proceedings of the Library and Information Studies Centre 75 years commemorative conference*.

Day, S., Lehmann, A., & Woessner, E. (2023). *Datenkompetenz in NFDI – Eine Projektvorstellung*. E-Science-Tage, Heidelberg. doi:10.5281/ZENODO.7741359

Engelhardt, C., Biernacka, K., Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., Downes, S., Erdmann, C., Garbuglia, F., Germer, K., Helbig, K., Hellström, M., Hettne, K., Hibbert, D., Jetten, M., Karimova, Y., Kryger Hansen, K., Kuusniemi, M. E., Letizia, V., … Zhou, B. (2022). *D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions* (Version V1.2.1). Zenodo. doi:10.5281/ZENODO.6674301

Federer, L. (2018). Defining data librarianship: A survey of competencies, skills, and training. *Journal of the Medical Library Association* 106(3), 294–303. doi:10.5195/jmla.2018.306

Greci, C., Welpton, R., & Woods, C. (2018). *The Safe Data Access Professionals (SDAP) Competency Framework*. Retrieved from https://securedatagroup.org/wp-content/uploads/2018/07/sdap_competency_framework-01_00.pdf

Heuer, J.–O., Kretzer, S., Mozygemba, K., Huber, E., & Hollstein, B. (2020). *Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung: Eine Handreichung für Forschende zur Erstellung eines Studienreports*. doi:10.26092/ELIB/166

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

Humphrey, C. (2006). *E-Science and the life cycle of research*. University of Alberta, Canada. Retrieved from https://era.library.ualberta.ca/items/3334684b-fa6a-4c9d-a74b-559fecd42f9f/view/79b064d6-7b51-4d18-8e4e-3d42b9faa81f/Lifecycle-science060308.pdf

Johnston, L. R., Curty, R., Braxton, S. M., Carlson, J., Hadley, H., Lafferty–Hess, S., Luong, H., Petters, J. L., & Kozlowski, W. A. (2024). Understanding the value of curation: A survey of US data repository curation practices and perceptions. *PLOS ONE* 19(6). doi:10.1371/journal.pone.0301171

Kelly, K., Marlino, M., Mayernik, M. S., Allard, S., Tenopir, C., Palmer, C. L., & Varvel Jr., V. E. (2013). Model Development for Scientific Data Curation Education. *International Journal of Digital Curation* 8(1), 255–264. doi:10.2218/ijdc.v8i1.258

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

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

Lehmann, A., Day, S., & Zinke, W. (2023). *Workshopbericht zum Community Workshop: Kartierung der Datenkompetenzlandschaft*. Zenodo. doi:10.5281/ZENODO.7920833

Ludwig, J., & Enke, H. (Eds.). (2013). *Leitfaden zum Forschungsdaten-Management: Handreichungen aus dem WissGrid-Projekt*. Hülsbusch. Retrieved from https://www.forschungsdaten.org/images/b/b0/Leitfaden_Data-Management-WissGrid.pdf

Mannheimer, S. (2024). Comparison of Issues and Data Curation Strategies. In S. Mannheimer, *Scaling Up: How Data Curation Can Help Address Key Issues in Qualitative Data Reuse and Big Social Research* (pp. 73–83). doi:10.1007/978-3-031-49222-8_5

Mertzen, D., Neuroth, H., Schneemann, C., & Woywod, K. (2023). Zertifikatskurs „Forschungsdatenmanagement für Studierende“ der Landesinitiative für Forschungsdatenmanagement in Brandenburg. In *Bildung-Aktuelle Entwicklungen und Perspektiven (an Hochschulen) im Bereich Data Science* (pp. 95–105). doi:10.18420/INF2023_08

Newbold, E., Wang, Y., Lehtsalu, L., Zhou, B., & McCutcheon, V. (2024). *RDA Professionalising Data Stewardship—What Does a Career Track for Data Stewards Look Like?* doi:10.15497/RDA00102

Pampel, H., Bertelmann, R., & Hobohm, H.–C. (2010). *Data Librarianship—Rollen, Aufgaben, Kompetenzen* (RatSWD Working Paper, 144). doi:10.2139/ssrn.1639856

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. (2023). *Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards* (Version 2). Zenodo. doi:10.5281/ZENODO.8010617

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). doi:10.7191/jeslib.2017.1096

Schüller, K. (2020). *Future Skills: A Framework for Data Literacy - Competence Framework and Research Report*. Working Paper No. 53 – Future Skills: A Framework for Data Literacy. https://doi.org/10.5281/ZENODO.3946067

Schultes, E. (2020). *Challenging areas for the Early Career Researcher in Data Stewardship*. International FAIR Convergence Symposium 2020, virtual.

Tang, R., & Hu, Z. (2019). Providing Research Data Management (RDM) Services in Libraries: Preparedness, Roles, Challenges, and Training for RDM Practice. *Data and Information Management* 3(2), 84–101. doi:10.2478/dim-2019-0009

Teal, T. K., Cranston, K. A., Lapp, H., White, E., Wilson, G., Ram, K., & Pawlik, A. (2015). Data Carpentry: Workshops to Increase Data Literacy for Researchers. *International Journal of Digital Curation* 10(1), 135–143. doi:10.2218/ijdc.v10i1.351

Whyte, A., Green, D., Avanço, K., Di Giorgio, S., Gingold, A., Horton, L., Koteska, B., Kyprianou, K., Prnjat, O., Rauste, P., Schirru, L., Sowinski, C., Torres Ramos, G., van Leersum, N., & Lazzeri, E. (2024). *Minimum Viable Skillsets—Catalogue of Open Science Career Profiles* (Version 1.0.0). Zenodo. doi:10.5281/ZENODO.11469300

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. 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(1), 160018. doi:10.1038/sdata.2016.18

Wilson, J. A. J., & Jeffreys, P. (2013). Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford. *International Journal of Digital Curation* 8(2), 235–246. doi:10.2218/ijdc.v8i2.287

Zentrum für Bibliotheks- und Informationswissenschaftliche Weiterbildung (ZBIW), ZB MED - Informationszentrum für Lebenswissenschaften, fdm.nrw - Landesinitiative für Forschungsdatenmanagement, & Technische Hochschule Köln. (2024). *Zertifikatskurs Forschungsdatenmanagement—Modulhandbuch*. Retrieved from https://www.th-koeln.de/mam/downloads/deutsch/weiterbildung/zbiw/angebote/zbiw_modulhandbuch_zk_fdm_2024-25.pdf