Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern

What does data stewardship mean in physics?

Holger Israel[i]   Markus M. Becker[ii]

2023

Zitiervorschlag

Holger Israel, Markus M. Becker. 2023. What does data stewardship mean in physics?. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2023. DOI: https://doi.org/10.17192/bfdm.2023.1.8570.

Dieser Beitrag steht unter einer
Creative Commons Namensnennung 4.0 International Lizenz (CC BY 4.0).

[i]Leibniz Information Centre for Science and Technology, Hanover, Germany. ORCID: 0000-0002-3045-4412
[ii]Leibniz Institute for Plasma Science and Technology (INP), Greifswald, Germany. ORCID: 0000-0001-9324-3236

Abstract

In this article, we expand on the considerations on data stewardship in physics we have presented as a poster at the “Data Stewardship goes Germany” workshop held in Brunswick in October 2022.[1]

We start from the observation that despite the close links between research in physics and scientific computing as a tenet of research data management (RDM), currently, the research data produced by physicists are not as FAIR[2] (findable, accessible, interoperable, and reusable) as they could and should be. Physics research groups in Germany as of now do not feature explicitly designated data stewards.

Building on a survey on RDM in physics conducted among researchers in 2020, we lay out a clear case and a mission for more explicitly defined and acknowledged data stewardship in physics. We argue that because of the closeness between data stewardship and genuine research, adequate domain knowledge is indispensable: Data stewards in physics should ideally be trained physicists themselves!

Data stewards are going to face a heterogeneous research landscape in terms of group size and resources, defined by the pressure to “publish or perish”. We consider that the introduction of data stewardship presents an opportunity to the physics community to self-organize research support infrastructures where they are missing. Data stewards from the physics community would be ideally skilled to transform the existing data handling solutions into the RDM systems needed to achieve a future of FAIR data from physics. We envision them to contribute to scientific projects both as advisors and as active role models of good scientific practice and reproducibility.

1 Data Stewardship and Physics – a Perfect Match?

Data stewardship - in the broad sense of a role dedicated to data curation - and physics research seem like an obvious match. Historically intertwined with mathematics, statistics, and computer science, the term research data evoke examples from physics and astronomy, ranging from the determination of the orbit of Ceres in the early 1800s to today's experiments informed by numerical simulations. Although scientific computing forms one pillar of research data management (RDM), the physics community in Germany has not been among the early adopters of data stewardship so far. With this article, in which we explore its potential, we would like to draw attention to the topic and initiate a more concrete discourse on fostering data stewardship in physics.

[1] Israel Holger and Markus M. Becker.“What Does Data Stewardship Mean in Physics?” Poster at the “Data Stewardship goes Germany” workshop held in Brunswick (2022). https://doi.org/10.5281/zenodo.7147762.

[2] Wilkinson, Mark D. et al. “The FAIR Guiding Principles for Scientific Data Management and Stewardship”. *Scientific Data* 3, Nr. 1 (2016): 160018. https://doi.org/10.1038/sdata.2016.18.

(cc)(i)

Several concepts and definitions of data stewardship exist, and there are subtle differences between the responsibilities of data stewards at various institutions, as e.g. discussed by Hausen et al. (2020)[3]. To some extent, tasks attributed to data stewards overlap with those of several other "data X" job titles, such as data scientist or data manager but also with more generic IT support. Hence, while job specifications might differ in the details, all of the following can be considered relevant tasks of a data steward in physics research:

- Organizing short term data storage and archiving
- Digitization of analog data (e.g. from experimental setups, using an electronic lab notebook (ELN)
- Documentation and annotation of data, according to best practices
- Enabling (technical focus) and promoting (human focus) data sharing and reuse
- Planning for the data needs of current and upcoming research (e.g. writing and maintaining data management plans)
- Linking and enhancing data via the curation of databases
- (Preparatory) data analysis, especially screening of incoming data and data quality management

We have organized the tasks along a spectrum from the more support-oriented at the beginning to the more research-oriented towards the end. This list is neither meant to be exhaustive, nor as a check list of necessary criteria which a data steward has to fulfill.

The crucial point is that these RDM activities are among the core responsibilities of a research physicist. This becomes obvious when you compare physics to medicine (where there exists an established practice of outsourcing the data analysis part of doctoral theses) or even the humanities. In other words, researchers in physics are their own data scientists; and obviously competent ones, since many professional data scientists indeed have physics degrees. The same applies for similarly data-affine disciplines, in particular (applied) mathematics and computer sciences.

Hence, in the above section, "physics" is to be understood in a broad sense and meant to encompass all research in the natural sciences that makes strong use of mathematical models and the outcomes of measurements in the form of numerical factual data. Next, detailing some subject-specific challenges, we are going to focus more on the research carried out in small labs working mostly with tabletop experiments, e.g. in universities' physics departments. Nevertheless, we hope that our concepts of data stewardship might find wider appeal.

[3] Hausen, Daniela et al. "Data Stewards an der RWTH Aachen University – Aufbau eines flexiblen Netzwerks".*Bausteine Forschungsdatenmanagement* 2 (2020). https://doi.org/10.17192/BFDM.2020.2.8278. See also references cited therein.

2 Lessons from the 2020 survey on RDM in physics

2.1 The survey

From March to May of 2020, the team that was at the time building a physics consortium to participate in the German National Research Data Infrastructure (*Nationale Forschungs-dateninfrastruktur*, NFDI) initiative, conducted an online survey among research physicists in Germany. The goal was to determine the status of their RDM and the resulting agenda for an NFDI consortium. The questions ranged over a wide array of topics, from modes of research and data categories to attitudes, practices and challenges with respect to the sharing and reuse of data. A total number of 488 replies were collected, resulting in 237 complete sets of answers, covering all but one of the German universities, which award doctorates in physics.[4] [5] A small majority of respondents self-identified as the principal investigator of a research group, with all major research fields in physics represented in the survey.

2.2 The state of RDM in physics in 2020

In physics, research data can take many forms, such as experimental data, simulations, theoretical calculations, and computational models. Good RDM practices ensure that these data are properly curated, preserved, and made available for reuse by other researchers. While sophisticated data structures, storage solutions and data processing procedures exist, one of the greatest challenges is the comprehensible labeling and description of the data. Without this, the subsequent use of data across persons or even groups is hardly possible. The widespread implementation of data stewards can provide a way to address this challenge.

Although it is not representative, the 2020 survey on RDM in physics revealed insights about RDM in the physics community that are relevant for developing a data stewardship strategy for physics. In the absence of domain-specific best practices for RDM, respondents referred to standards such as the DFG guidelines on Good Scientific Practice[6] and to general concepts of reproducibility and peer review when asked about criteria to assess data quality. Since prototypical research data consist of measurement results, not only does the measurement value need to be paired with the correct unit. Context information (what has been measured, how and why?) also needs to be transferred to a potential reuser.

[4] Israel, Holger, Esther Tobschall, and Frank Tristram. “Forschungsdaten FAIR verwalten”. *PhysikJournal* 20, Nr. 7 (2021): 35–38.

[5] Israel, Holger, Esther Tobschall, and Frank Tristram. “Dataset for the publication ‚Umfrage zum Forschungsdatenmanagement in der Physik”. *Physikalisch-Technische Bundesanstalt (PTB)* (2021). https://doi.org/10.7795/730.20210511

[6] Deutsche Forschungsgemeinschaft. “Guidelines for Safeguarding Good Research Practice. Code of Conduct”, 2022. https://doi.org/10.5281/ZENODO.6472827.

(cc)(i)

The main result is that documentation of research activities is not as seamlessly digitized as one might assume: A clear majority of respondents continues to rely on personal or instrument-related paper laboratory notebooks. The frequent use of log files from devices also poses interoperability questions. Interestingly, if electronic lab notebooks (ELN) are used, they often are one-of-a-kind systems the researchers have developed themselves.

Data loss or corruption due to conversion between different systems or from analog to digital thus presents a concern when in comes to keeping numerical factual data and context data together. For instance, context data can be rendered ambiguous due to inconsistent use of terminology. Community-wide standards can mitigate these issues.

In terms of terminology, physics largely lacks widely recognized controlled vocabularies for the description of phenomena and the documentation and annotation of research outputs, thus impeding the use of technologies based on the semantic enrichment and structuring of data. Asked if several broad categories of metadata were relevant in their field of research, more than half of the respondents stated the importance of administrative information (who did what and when), information on data generation, metadata on the test object, and information on quantities, units, and uncertainties.

Frequently, only small, selected groups of people are allowed to reuse research data: Members of the authors' own research group (70% affirmation among respondents) and, upon request, external partners (44% affirmation) are often granted access. While in 24 % of the cases, the public can use the data after registration, open access without registration is being practiced by 17% of the sampled physicists (see Figure 1).

Asked about their RDM-related challenges, respondents to the survey state complexity in data structures and formats (69% approval), the large number of tools and methods (61% approval), complexity of documentation (59% approval), and confusion about underdeveloped metadata standards (50% approval) as the main obstacles to data reuse. Backwards compatibility of software (e.g. in devices which often use proprietary data formats), and difficulties to assess the quality and usefulness of data are frequently cited as well. While the surveyed community considers good scientific practice and reproducibility vital, there seems to be no real consensus on how to safeguard data quality besides carefully designing statistical tests and analyses.

In summary, the survey finds evidence for a *data curation gap,* i.e. that the daily practice of data handling in many cases does not live up to the standards of RDM best practices. The results suggest a qualitative and quantitative lack of documentation of physical and virtual experiments, compromising the FAIRness of the resulting data.

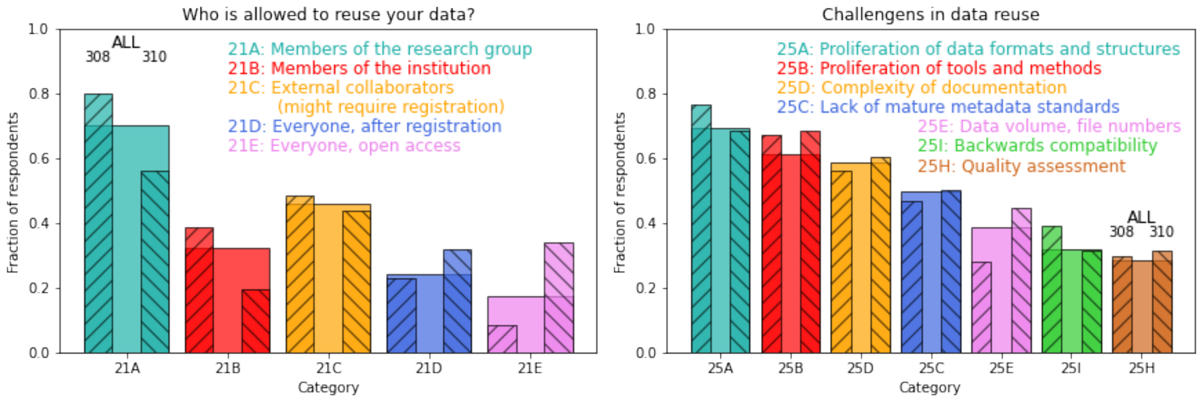

Figure 1: Share of most frequent answers to a selection of question from the 2020 survey on RDM in physics. For each answer option ("Category"), approval is shown for: (1) ALL respondents, irrespective of their subject area (coloured bars); (2) respondents self-identifying as researchers in DFG subject area 308 (atoms, molecules, optics, plasmas; hatched bars: //); (3) respondents self-identifying as researchers in DFG subject area 310 (statistical physics, soft matter, biological physics, nonlinear dynamics; hatched bars: \\). Respondents could select multiple categories.

2.3 Why physics needs data stewards

We conclude from the survey that research data management in physics needs improvement. While physicists, obviously, successfully handle and analyze their research data, the potential offered by databases, knowledge graphs, and semantic technologies remains under-explored, due to the fundamental issues highlighted by the survey. Moreover, there is still considerable cultural change necessary to build a culture of data sharing.

We emphasize that this is not the researchers' fault. The survey also shows that physicists are generally open-minded to the possibility offered by FAIR data and structured RDM. Subsequent, more detailed discussions with various scientists in the field that we contacted in the context of the NFDI corroborate this picture on the level on anecdotal evidence. However, our respondents also, and rightfully, maintain a critical attitude, requesting that RDM must tangibly improve their research instead of eating into their tight time budgets.

We hypothesize that a mixture of several systemic effects likely cause the deficits in research data management and the surrounding infrastructures uncovered by the survey.

(cc)(i)

- In contrast to subdomains such as particle physics where research takes place in large international collaborations, and hence support infrastructures are organized on at least the European level, for physics as a whole, coordinated research infrastructure efforts are largely absent in Germany.
- Especially the more experimental sub-disciplines, e.g. optics, are characterized by small laboratories consisting of an individual university professor and a few of their students. In case they are not tethered to larger, collaborative projects, they tend to have rather limited resources. In particular, these groups strive to use person power in a way that maximizes research output.
- The persistence of a "publish quickly or perish" science system that rewards short-term thinking stands in the way of developing the present ad-hoc-solutions into FAIRer and more sustainable research infrastructures. The egregious lack of permanent, non-professorship positions leaves academia in a state of constant and worsening brain drain.

Additionally, we observe that research workflows in many branches of physics tend to be more variable than in, e.g. biomedical research or chemistry, where they more often include frequently repeated steps carried out with standardized, off-the-shelf equipment.[7] In contrast, experiments in low-temperature plasma physics provide a good example of this type of versatile workflow (see Figure 2)[8]. Different elements of the setup can be labelled as "source", "medium", "target", or "diagnostic". But what is an element of the measurement equipment in today's experiment might become the object under test tomorrow, and vice versa. Moreover, combinations of and comparisons between data generated by different methods (e.g. experimental vs. numerical), and often in different formats are common.

This has repercussions for the use of ELN, and in turn, the tasks of data stewards. In general, existing ELN systems do not meet the needs of physicists, especially in terms of flexibility of data models. ELN are usually geared towards experiments with simpler data models and standardized protocols, where an efficiency gain from using ELN is obvious.

Thus, as demonstrated by the survey, in many cases physicists have developed ingenious do-it-yourself solutions for RDM. These are often created out of necessity and tailored to their own needs, so that subsequent use by others is not foreseen and thus made more difficult. Data stewardship in physics means, in our view, collecting, harmonizing, and FAIRifying these isolated RDM systems in a coordinated approach. The issues we just described are especially acute in the small research groups mentioned before. Nonetheless, with ever-increasing data volumes, even small research teams

[7] As evidenced by the publication of whole books devoted to "methods and protocols", e.g.: Ziegler, Slava, and Herbert Waldmann, e.d. Systems Chemical Biology: Methods and Protocols. Bd. 1888. Methods in Molecular Biology. New York, NY: Springer New York, 2019. https://doi.org/10.1007/978-1-4939-8891-4.

[8] Plathe, Nick et al.. "Methods and Tools for Data-Driven Science in Applied Plasma Physics", 2021. https://doi.org/10.5281/ZENODO.5579012.

CC BY

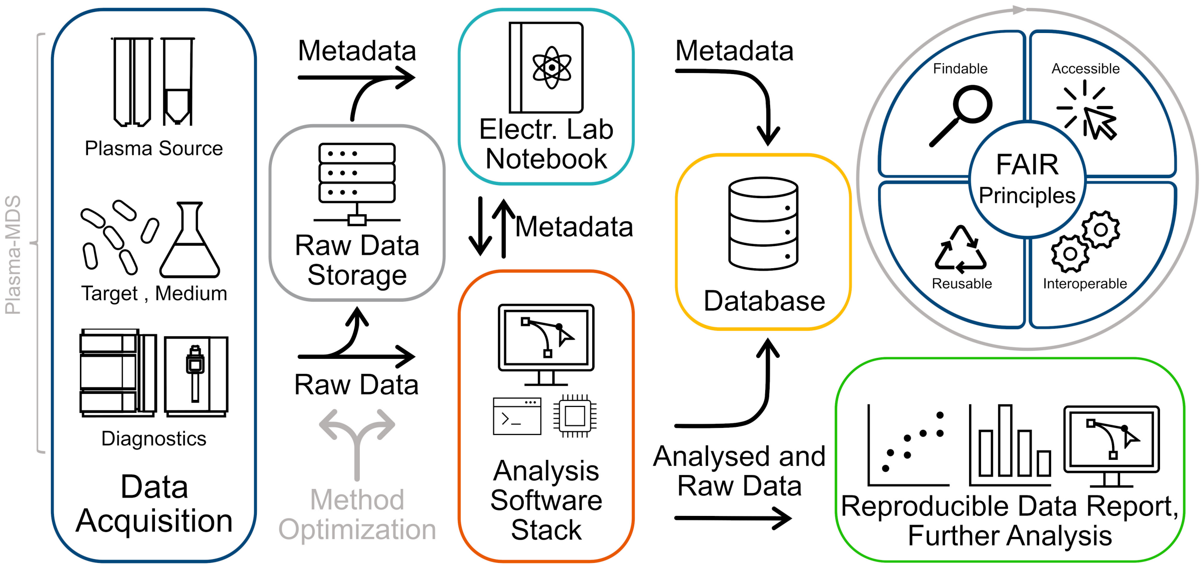

Figure 2: Example of a FAIR research workflow in plasma physics. A data steward in the field will need to be familiar with all of the elements shown here, from the equipment used to acquire data to the analysis software stack and databases. Flexible data models are needed to reflect the high degree of variability in the data acquisition in all of the elements further downstream.

are facing ever more complex RDM tasks. They, too, advance into a “big data” regime in which “mere” data handling and explorative data analysis become more and more entangled. It is therefore crucial that data management systems reach out to these small groups, and it can be observed that more and more research communities are facing up to these challenges and trying to establish community standards[9] [10] [11] and common tools[12] [13] [14].

[9] Franke, Steffen, et al.. “Plasma-MDS, a Metadata Schema for Plasma Science with Examples from Plasma Technology”. *Scientific Data* 7, Nr. 1 (2020): 439. https://doi.org/10.1038/s41597-020-00771-0.

[10] Alves, Luís L. et al.. “Foundations of plasma standards”. *Plasma Sources Science and Technology* 32, Nr. 2 (2023): 023001. https://doi.org/10.1088/1361-6595/acb810.

[11] Poeschel, Franz, et al. “Transitioning from File-Based HPC Workflows to Streaming Data Pipelines with OpenPMD and ADIOS2”. *Driving Scientific and Engineering Discoveries Through the Integration of Experiment, Big Data, and Modeling and Simulation*, Jeffrey Nichols et al.(eds.), 1512:99–118. Communications in Computer and Information Science. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-96498-6_6.

[12] Chaerony Siffa, Ihda, Jan Schäfer, and Markus M. Becker. “Adamant: A JSON Schema-Based Metadata Editor for Research Data Management Workflows”. *F1000Research* 11 (2022): 475. https://doi.org/10.12688/f1000research.110875.2.

[13] Fitschen, Timm et al.. “CaosDB—Research Data Management for Complex, Changing, and Automated Research Workflows”. *Data* 4, Nr. 2 (2019): 83. https://doi.org/10.3390/data4020083.

[14] Stocker, Markus et al. “FAIR scientific information with the Open Research Knowledge Graph”. Barbara Magagna (ed.). *FAIR Connect* 1, Nr. 1 (2023-01-11): 19–21. https://doi.org/10.3233/FC-221513.

(cc)(i)

We conclude that there is a pronounced need for data stewards and a long, but fascinating agenda for them. Data curation in the sense of quality assurance (upholding reproducibility) and data aggregation into high-level databases by enhancing data with metadata and documenting (implicit) connections need to be developed to attain "FAIR data in physics".

3 Data stewards as advisers and reproducibility champions

In the remaining part of the article, we would like to present our vision what a widespread adoption of data stewardship in physics could ideally achieve. In short, data stewards have the potential to serve as role models of reproducibility and good scientific practice. If the physics community were to agree on this goal, data stewardship could provide the condensation nucleus around which a better and more sustainable research culture could emerge.

3.1 Why physics data stewards should be physicists

In the introduction, we have argued that, in physics, many RDM tasks attributed to data stewards and bona fide research activities are deeply intertwined. Given the complex and variable research workflows (see Figure reffig:example), we observe that many of the pressing RDM issues, e.g. modelling such workflows for implementation in ELN, involve the construction of domain models (ontology engineering) of physics. The Plasma-MDS metadata schema for plasma science describing all elements of data acquisition in Figure reffig:example (plasma source, target, medium, and diagnostics) represents the result of a domain modelling effort by the plasma community.[15] Hence, data stewards working towards these goals require the depth of domain knowledge that comes with a physics degree, in order to accurately represent the rationales behind the research workflows in their domain models, and in turn, make a meaningful impact in terms of improving processes.

This is most evident for the most research-oriented item on the list in Section 1 of this article: data quality management in the sense of assessing the plausibility of raw data as well as of data products to be shared for reuse. Reusability entails that the dataset not only has to meet fit-for-purpose criteria for the science case it has been created for, but also for at least some likely alternative use cases. Expertise and ideally some experience in the subject matter are required to evaluate the usefulness of data in such a way.

[15]Franke, Steffen et al. See footnote 9.

CC BY

The screening of measurements or computation results for monitoring if their outcomes are plausible (and if they are not, discerning between simple malfunctions and more interesting anomalies) means a more cursory analysis than the complete set of steps necessary for publication-grade results. However, it usually involves a relevant number of measurement devices and analysis codes. Familiarity with the underlying physical processes has to underpin every assessment of data quality. Consequently, data stewardship blends into data science, even though these are generally two distinct concepts.

Domain knowledge is crucial, but it does not suffice: In order to improve research practices in terms of the FAIR principles, a person needs to combine sound methodological and statistical expertise with a broad knowledge in RDM, library and information science, and scientific computing. Acknowledging that broad array of qualifications, this person should not be considered a lesser researcher than anyone fulfilling a more traditional role in a research project. Consequently, they also ought to be co-authors on relevant publications building on their work as data stewards. Since the formal role of data steward does not yet exist, we are not aware of any explicit coauthorships for contributing via data stewardship. However, we consider it likely that this is already an implicit practice. Naturally, publication practices need to adapt as well, and allow for more flexibility in content (e.g. data or software publications) as well as form (e.g. living / versioned documents or executable publications).

Because of the frequent use of, often customized, research software at many of the steps along the research data lifecycle (indeed all of them for pure simulation studies), an effective physics data stewards is going to be a *research software engineer*[16] at the same time. As such, most of the observations and conclusions brought forward by Anzt et al. also apply to data stewards in physics. Most notably, that despite living a core function of the scientific method, research data specialists currently not only lack public recognition and, in turn, funding. Even within academia, and physics in our case, both roles do not receive the appreciation they deserve, given their importance for a sustainable development of science.

We stress the importance of domain knowledge not as an attempt to impose an artificial barrier or even hierarchy between physics and other research fields. In contrast, we are very aware of what can be called the fuzzy edges of physics. Rather, our motivation is to caution against modes of “data stewardship” that are detached from day-to-day research, be it as providers of an outsourced services or as a bureaucratic controlling entity. We expect that neither approach would yield the desired results in terms of FAIR data and improved research quality, but lead to frustrations on all sides.[17]

[16] Anzt, Hartwig et al. “An Environment for Sustainable Research Software in Germany and beyond: Current State, Open Challenges, and Call for Action”. *F1000Research* 9 (2021): 295. https://doi.org/10.12688/f1000research.23224.2.

[17] See also Awre, Chris et al. “Research Data Management as a “Wicked Problem””. *Library Review* 64, Nr. 4/5 (2015): 356–71. https://doi.org/10.1108/LR-04-2015-0043.

cc by

This view is backed up by the replies of the respondents of our survey to the final, free-form question asking for “further suggestions or ideas for the work” of the then-planned NFDI consortium. The overall thrust emphasizes the need for consensus building and self-organization within the physics community for the acceptance of any RDM initiative, and in particular for the researchers’ motivation to contribute to it. Our respondents ask for a well-planned and thorough strategy based on an appraisal of their individual science cases and are wary of the prospect of having to cope with “one-size-fits-all” solutions. While we are concerned with physics, we consider it likely that researchers in other fields hold similar views, and would agree to corresponding statements about data stewardship and the relevant domain knowledge for their respective subjects.

3.2 Bringing about cultural change

Which role for data stewards and which route towards FAIRer research practices can we support instead? Our vision for the emerging role of data stewards in physics is that they take on an advisory role in research projects. Data stewards should contribute to the scientific value of research outputs by championing FAIRness and reproducibility over the whole data life cycle, beginning at the planning phase.

The planners of such a data stewardship scheme have to decide whether the data steward role should coincide with one of the main persons responsible for the data analysis, as insinuated by the list of tasks in section 1. This is the current default in which no explicit data steward role is being assigned, and conforms with the tradition of the universal scholar who (equally) masters every aspect of their research and who, driven by furthering “their own” research, often works on small, self-contained projects. There is nothing inherently “wrong” or “un-FAIR” about this mode of science.

However, the majority of physics studies nowadays is not being conducted in this mode, but by research groups with division of labour already in place. This more typical setting could benefit more from a data steward in the sense of a FAIRness and reproducibility champion distinct from the person(s) analyzing the data (often a master or PhD student). Such an advisory data steward would function to some degree as an internal reviewer of the study. However, in sufficiently large teams where this is possible, we recommend keeping these roles distinct such that the person conducting an internal review can focus on the studies’ main scientific impetus while the data steward commits to the more data-centric perspectives of the FAIR principles. In the same vein, the data steward’s role should not coincide with the principal investigator’s role as project manager and supervisor of students.

Independence from other core responsibilities allows our data steward to advocate for the interests of the wider community of potential data reusers. Furthermore, one data steward associated with a small institute or department would be advising several

(cc)(i)

groups and studies in parallel and must not favour one project over another. Returning to the free-form suggestions in the 2020 survey, the issue of allocating any additional RDM personnel was brought up, too. One respondent, for instance, poignantly voiced concerns about what they described as an increase in compulsory activities that are not compensated by matching funding, and thus effectively funneling personnel away from research. Pointing to both acceptance by the community and effectiveness of the scheme, they appeal for a lightweight approach which makes it easy for institutions to participate without having to invest significant time up-front. We concur the latter would disadvantage scientists at less resourceful institutions. Instead, we envision a community-driven negotiation process to design data stewardship schemes that combine a low entrance barrier for interested parties with an effective uptake of best practices in RDM. In addition to funding issues, the exact responsibilities of data stewards need to be deliberated as well as how their efforts should be distributed between various projects and groups.

Data stewards at different institutions should be well connected not only amongst each other but also with NFDI consortia and other relevant organizations, e.g. learned and professional societies such as Deutsche Physikalische Gesellschaft (DPG). We suggest a round table or “network of data stewards” as a forum to facilitate the exchange of information between RDM professionals, i.e. experts at libraries or computing centers, and research practitioners. An analogous network of research software engineers has been suggested[18] to address the challenges scientists face regarding software sustainability.

Because they are rooted in the physics community, our proposed data stewards should be in a good position to hear, relay, and respond to the needs and demands of the research communities of which they are an integral part. In doing so, they will be well-placed to serve as role models of good scientific practice and thus instigate and promote a cultural change towards a spirit of data sharing.

4 Conclusion

Explicitly organized data stewardship in physics is still in the very beginning but it should be acknowledged that the discourse on improving physics RDM is taking place in the context of a challenging environment. Numerous stakeholders are involved in a field of science with a patchy landscape of existing research-supporting infrastructures.[19] Researchers distributed in many small groups and adapted to an academic

[18] Anzt, Hartwig, et al. See footnote 16.

[19] And this is a simple case. For the challenges of data stewardship for person-related data, see: Wendelborn, Christian, Michael Anger, and Christoph Schickhardt. “What Is Data Stewardship? Towards a Comprehensive Understanding”. Journal of Biomedical Informatics 140 (2023): 104337. https://doi.org/10.1016/j.jbi.2023.104337.

CC BY

system demanding fast-paced publication output would have to cut out time and self-organize in a collaborative process aimed at reworking many aspects of their research practices on the fly.

The 2020 survey on RDM in physics has shown that making data FAIR needs to start at the foundational level of terminology, file formats and, most importantly, awareness. As much as acknowledging this agenda is a prerequisite for the successful introduction of data stewards to physics, the converse holds true as well: By explicitly designating data stewardship roles, attention and researcher's fresh ideas can be directed towards basic RDM challenges. Hence, these two things have to go hand in hand.

Our survey also provides evidence for the skill and creativity physicists already devote to solving problems that fit the job description of a data steward because these tasks are integral to their research. Hence we assert that trained physicists make the best-suited data stewards for physics. Not only can physicists bring forth the domain expertise necessary to model physics research workflows. Data stewards from the physics community moreover help to represent the communities best interests.

Since the purpose of data stewardship is to realize the standards of good scientific practice, we consider it essential to have a well-informed discourse involving all status groups and the entire breadth of the physics community, followed by a well-thought and deliberate implementation. We hope that presenting our vision of data stewards as reproducibility champions contributing their advice to scientific studies will help to initiate such a process.

Acknowledgments

The authors express their thankfulness to the organizers of the "Data Stewardship Goes Germany" workshop for giving us the opportunity to present a poster in what has been an inspiring and welcoming workshop. We thank Esther Tobschall, Diana Slawig, Anna-Lena Lorenz, Giacomo Lanza, Joachim Meier, Johanna Hickmann, and Lena Steinmann for helpful discussions. We thank Nick Plathe for permission to use his workflow schematic. M. Becker acknowledges funding by the Federal Ministry of Education and Research (BMBF), grant 16QK03A.