Position Paper

Perspectives and Potentials of Open Data for the Sports Sciences

The "What," the "Why," and the "How"

Melanie Krüger[1], Claudia Biniossek[2,3,4], Markus Stocker[2], and Dirk Betz[2,3,4]

[1]Institute of Sports Science, Faculty of Humanities, Leibniz University Hannover, Germany
[2]Knowledge Infrastructures Research Group, TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany
[3]Center for Empirical Research in Economics and Behavioral Sciences, University of Erfurt, Germany
[4]Bioninformatics Group, Institute for Computer Science, Albert-Ludwigs-University of Freiburg, Germany

Abstract: Open Science practices have become well established in recent years. In this position paper, we argue that Open Data in particular holds great potential for empirical research in sports science, and sport and exercise psychology in particular, since it fosters the reintegration of scientific knowledge as primary research data in subsequent research life cycles. On that account, the sports science community has to develop a unified position on research data management, which supports the implementation of Open Science practices and standards. To this end, in this article we first define Open Science and research data management (RDM) and describe them in the context of sports science. We then present examples of existing, relevant RDM solutions, with a particular focus on sport and exercise psychology and neighboring disciplines. Finally, we derive perspectives for the development of a sustainable RDM structure and present current developments within the German sports science community.

Keywords: Open Science, research data management, FAIR principles

Perspektiven und Potentiale von Open Data für die Sportwissenschaft: Das „Was", das „Warum" und das „Wie"

Zusammenfassung: Open Science-Praktiken haben sich in den letzten Jahren in vielen Wissenschaftsdisziplinen etabliert. In diesem Positionspapier argumentieren wir, dass insbesondere Open Data ein großes Potenzial für die empirische Forschung innerhalb der Sportwissenschaft birgt, da es die Reintegration von sportwissenschaftlichen Erkenntnissen als primäre Forschungsdaten in nachfolgende Forschungszyklen fördert. Dies erfordert innerhalb der Sportwissenschaft die Entwicklung einer gemeinsamen Position zum Forschungsdatenmanagement (FDM), welche die Implementierung von Open Science Praktiken ermöglicht. Zu diesem Zweck werden wir in diesem Artikel zunächst Open Science und Forschungsdatenmanagement definieren und im Kontext der Sportwissenschaft beschreiben. Anschließend werden Beispiele für bestehende FDM-Lösungen vorgestellt, unter besonderer Fokussierung auf Sportpsychologie und benachbarte Disziplinen. Abschließend werden Perspektiven für die Entwicklung einer nachhaltigen FDM-Struktur innerhalb der deutschen Sportwissenschaft, mit dem besonderen Fokus auf Sportpsychologie abgeleitet. Wir argumentieren, dass solch eine Struktur auf bereits etablierte FDM-Lösungen aufbauen müssen, um den spezifischen Herausforderungen der Sportwissenschaft als Querschnittswissenschaft Rechnung zu tragen.

Schlüsselwörter: Open Science, Forschungsdatenmanagement, FAIR Prinzipien

In the mid-2010s, the so-called replication crisis revealed that a larger number of scientific studies could not be replicated, attributable to various potential reasons, for example, methodological weaknesses such as low sample size, systemic reasons such as publications bias, but also "questionable research practices" (Renkewitz & Heene, 2019a). Baker (2015) showed that more than half of psychological studies could not be replicated. Since then, practical and theoretical approaches have emerged to overcome the replication crisis. In the context of sport and exercise psychology research, Raab and colleagues (2017) consider (voluntary) self-reflection as a feasible solution for the replication crisis. In the broader psychology community, Open Science practices have been discussed to overcome the replication crisis, for example, in an article collection in the *Zeitschrift für Psychologie* entitled "Open Science in Psychology. Progress and Yet Unsolved Problems," edited by Renkewitz and Heene in 2019. In this article collection, Crüwell and colleagues (2019) introduced the different elements of Open Science in their review article and concluded by stating:

> *Open Science practices are a collection of behaviors that improve the quality and value of psychological research*

> *and aim to accelerate knowledge acquisition in the sciences. One barrier that prevents psychological scientists from adopting Open Science practices is a lack of knowledge.* (Crüwell et al., 2019, p. 244)

Meanwhile, Open Science has become well established with a substantial number of scientists having acquired related knowledge and funding bodies, publishers as well as academic institutions standardly demanding the implementation of Open Science practices by researchers (see, e.g., German Research Foundation, 2022). In this context, Open Data constitutes a particularly sensitive aspect of Open Science practices. In this position paper we argue that for researchers in the field of sport and exercise psychology, and (sports) science communities in general, to develop a stance on and later to adopt and implement Open Data practices or alternative approaches (e.g., the FAIR Data Principles, which include restricted access to data as well as secure data centers), it requires careful (self-)reflection on what is appropriate and required for the individual research project and data (c.f. Raab et al., 2017). Importantly, we do not argue for all data to be openly published, but to be aware of the future potential of Open Data practices for collaborative and inter- as well as transdisciplinary research. Further, on a conceptual and technical level, a seamless and sustainable research data management infrastructure based on distributed systems and encompassing the entire research data cycle, that is, obtaining informed consent of participants, data collection, analysis, and data publication, is a critical precondition for the implementation of Open Data practices. We will further argue that Open Data is a topic of great relevance for the sports science community, as its interdisciplinary nature and methodological diversity make it particularly susceptible to the challenges but also the potentials of Open Data. This is particularly evident in the field of sport and exercise psychology: When studying the motivation for physical activity across the lifespan, for example, the documentation of the research instruments used for capturing the hypothetical construct of motivation is particularly important, as cognitive, affective, and conative dimensions are often weakly correlated and data can only be re-assessed and reused based on accurate documentation. The problem of data anonymization is also highly relevant in sport and exercise psychological research, since it is often concerned with special populations, for example, minors or elite athletes.

On that account, we first define Open Science in general and Open Data in particular, distinguish them from the FAIR Data Principles and research data management (RDM), and describe them in the context of sport and exercise psychology. We then specify current developments within the German sports science community and derive a perspective on Open Data for the sports sciences.

The "What" – Open Science and Research Data Management

The term "Open Science" has become a buzzword, often used synonymously with data archiving and the accessibility of primary research data. However, standard specifications, for example, by the German Research Foundation (2022; DFG) or UNESCO (2021), define the term Open Science in a much broader way (UNESCO, 2021):

> *Open Science is defined as an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community. [...]. It builds on the following key pillars: open scientific knowledge, Open Science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems*. (UNESCO, 2021, p. 7)

In this paper, we focus specifically on the aspect of open research data, commonly termed "Open Data," as part of the pillar of open scientific knowledge (see Figure 1). Importantly, while Open Science and FAIR principles overlap to a large extent, they are not fully congruent, conceptually. For research data to be Open Data means not necessarily adhering to the FAIR data principles (Wilkinson et al., 2016), namely, *findability*, *accessibility*, *interoperability*, and *reusability*. Strictly speaking, Open Data grants access to, but not necessarily permissive reuse of, research data. Conversely, data accessibility does not imply Open Data per se, or in other words, it does not mean that data *not* openly accessible to the broader public due to necessary embargoes or due to data protection issues are not accessible in terms of the FAIR data principles.

This delimitation becomes a relevant aspect to consider in particular in the field of sport and exercise psychology and broader sports science communities, when thinking, for instance, about the small population under study (e.g., sports talents and elite athletes), the sensitivity of data obtained (e.g., acquired in the context of mental coaching), and the various potential end users of data (e.g., for

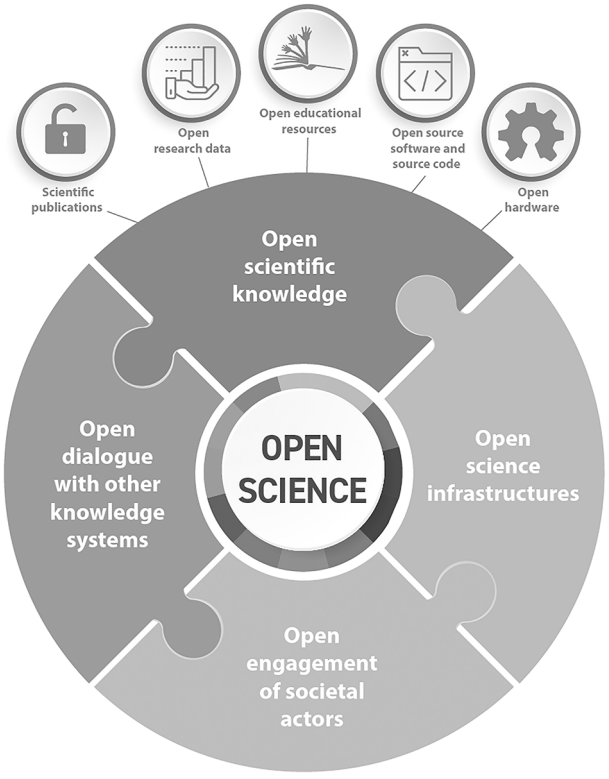

Figure 1. UNESCO recommendation on Open Science (Source: UNESCO, 2021, p. 11 under a CC BY-SA 3.0 IGO license).

educational and health policy development or talent scouting). For example, while the reuse of motor skill data, recorded through the German Motor Test 6–18 (GMT 6–18; Bös et al., 2009), for larger cohort studies might clearly be of strong scientific interest, the impact is ultimately gained from deriving recommendations for actions in, for example, school education or for talent development in sports.

Critically, depending on who reuses data, FAIR data might require different levels of data curation. In this context, anonymization, persistent identifiers (PIDs) for participants[1], as well as secure data centers are only a few possibilities for ensuring the privacy rights of the participants but still making the data FAIR. Thus, while the *FAIRification* of data can be challenging, it is an essential part of good scientific practice and scientific progress (Wilkinson et al., 2016) and should be accounted for when managing research data.

In 2021, Boccali and colleagues (2021) defined RDM as:

> *An essential element of the research process. It is an ongoing activity throughout the data lifecycle, including the active organisation and maintenance of data during the research process and suitable archiving for future reuse. It requires researchers and their host institutions to consider, from the start of a research process, how to collect, reuse, store, and curate their data.* (Boccali et al., 2021, p. 6)

In other words, RDM refers not only to the storage of primary research data at completion of a study but also, for example, to the handling of raw and processed data, its sharing with partners at different research sites, its export to and import into data processing and analysis software, as well as its graphic representation.

In sum, Open Science, FAIRification of data, and RDM have become increasingly important areas of consideration and discussion in the scientific community (for sport and exercise psychology, see, e.g., Schönbrodt & Scheel, 2017). Reasons to implement Open Science and FAIR strategies are the striving of the community for increased transparency and credibility, as a reaction to the replication crisis, as well as for reuse potential. This is paralleled by the increasing mandatory requirements of funding institutions, journals, and universities. In addition to the underpinning policies (e.g., DFG, Leibniz Association[2]), the DEAL negotiations[3] for open access, and the National Research Data Infrastructure[4] definitely feature in this context. In sports science, existing statements or guidelines regarding Open Science, RDM, and FAIRification of research data originate from the initiative of individuals (see, e.g., Raab et al., 2017) or parent disciplines (e.g., joint position paper in the field of educational sciences and didactics[5]). Further, different sports science communities, for example, sport and exercise psychology and sports pedagogy, vary strongly with regard to research methods and tools used (e.g., interviews, physiological measurements, surveys, tests), consequently resulting in very different preconditions, requirements, and needs for an effective RDM. Nevertheless, the heterogeneity between individual researchers and research groups should not hinder the development of a joint position in sports science, as they are all concerned with managing research

1 https://www.toolpool-gesundheitsforschung.de/produkte/pid-generator
2 https://www.leibniz-gemeinschaft.de/en/research/open-science-and-digitalisation/
3 https://deal-konsortium.de/
4 https://nfdi.de
5 https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2020.07_Kontrollierte_%C3%96ffnung.pdf

data and have to find suitable and sustainable solutions to meet the developing standards of Open Data.

The "Why" – Potentials and Challenges of Open Data

Researchers hold different, potentially complementary, roles and responsibilities when it comes to RDM in general and Open Data in particular: On the one hand, there is the role of the "data generator," who addresses a specific research question or pursues a scientific mission in the context of which data are collected and need to be managed. At this early stage in the research data cycle, the researcher takes personal responsibility for good scientific practice and the implementation of adequate RDM standards depending on the particular context in which the data are collected. Further, there is the role of the "data provider," who makes primary research data available to a particular target audience, which varies depending on the type of research conducted and may encompass the scientific community but also the broader public. The extent to which data is provided may also be influenced by the requirements of the specific funding agency, the researchers' own institution, the academic publisher, and importantly, also, the researchers' own position on Open Data. Data provision can be realized by publishing the data in raw or processed form and with or without access restrictions. However, to date, aggregated datasets are typically published in processed, textual, or graphical form in research publications. Importantly, when psychological and behavioral data is involved, such as in sport and exercise psychology and sports science in general, data providers have to ensure participants' privacy rights. In these circumstances, data curation is mandatory before the data can be published. When adopting a wider perspective, it is of importance to note that in, for example, representative population surveys such as the Socio-Economic Panel (SOEP)[6], that is, for large datasets with very high potential for reuse, data curation is implemented by specialized staff using existing research data infrastructures, resulting in comprehensively documented and citable data publications with their own digital object identifier (DOI). However, in sports science, this role is most commonly filled by the researchers themselves, leading to potential concerns and barriers in adopting Open Data practices (see Schönbrodt & Scheel, 2017).

Then again, there is the role of the "data user," who reuses and cites shared and published research data, in the context of their own research questions and projects. Data sharing, data publication, and data reuse are not principally new developments. With the upswing of the Open Data movement, however, the last two roles have become more explicit, with regard to the steps to be taken by the researchers as well as with regard to the practices and standards demanded by other stakeholders, such as publishers and funding bodies. While this might seem to come with new challenges and workloads for the individual researcher in the short run, as will be discussed in the following, we argue that it also holds great potential, in particular for researchers within the sports science community, in the longer run.

Challenges

In the literature, fundamental challenges for the adoption of Open Data practices have been described by various experts from different backgrounds: On the one hand, librarians and data stewards, pointing to the gold standard of RDM in surveys and randomized controlled trials (RCTs), see challenges in the creation of incentives and the establishment of a culture of data sharing (e. g., Betancort, Cabrera et al., 2020). On the other hand, data scientists, focusing on state-of-the-art technical solutions, distributed systems, and automatization, highlight the challenge of reducing the actual workload for researchers throughout the RDM process (see Betz et al., 2022, Jeffery et al., 2021).

In addition, the following challenges apply to the sports science communities: *First*, sports science is an interdisciplinary science, commonly also referred to as a "cross-sectional science" (Krüger, 2022) and is composed of heterogeneous communities, which can show greater similarity in their research practices and standards to their neighboring parent disciplines, for example, sport and exercise psychology versus general psychology, than to other sports science communities, for instance, sport and exercise psychology versus sports history. This challenge can be further exemplified when considering decision-making research. In sports science, research on motor decision-making is central to movement scientists and sports psychologists and focuses on questions such as: Do humans take motor costs into account when choosing between potential actions? Do elite athletes attend to different stimuli than novices do when deciding between competing motor actions? Do joint goals and social inter-

[6] https://www.diw.de/en/diw_01.c.615551.en/research_infrastructure__socio-economic_panel__soep.html

action in sports teams affect motor decision-making? Decision-making has also been the focus of various other research disciplines, such as economics, political sciences, sociology, and communication sciences. Depending on whether a study is to be published in a psychology, economics, or sports science journal, it has to comply with the respective guidelines (e.g., ethics certificate, preregistration) and RDM policies (e.g., data archiving in a certified data repository) of the target journal. This circumstance makes it difficult to establish efficient, comprehensive, and tailored RDM structures, including metadata standards, tools, etc. that equally serve the needs of the different communities.

Second, individual sports science communities are relatively small, often lacking the resources to establish such structures on their own. This leads to significant downsides for researchers, not having the capacity and/or support for sustained curation of their data. Thus, researchers in these fields are strongly affected by demands imposed by funding bodies as well as being dependent on the existing standards of their parent disciplines, of university infrastructure and cross-disciplinary transparency, as well as on the access to established tools and infrastructures to be adopted throughout the research (data management) process. However, there is a risk that the RDM needs of the individual sports science communities will not be adequately addressed. Metadata standards, for example, developed within and for psychology may not adequately describe sport and exercise psychological data, making these data less likely to be found and reused. In day-to-day research practice, this might result in additional or increased barriers of individual researchers to comply with Open Data demands.

Consequently, depending on the focus of the particular research question at hand, sports science researchers in general, and sport and exercise psychologists in particular, have to flexibly adapt to various RDM standards, practices, and tools existing in different parent disciplines in the course of their studies. Usually, solutions tailored for the sports science community cannot be used because they do not exist. This might represent a particular challenge and additional load for junior scientists, who are commonly the "data generators" on-site, who have to account for these varying demands under working conditions often characterized by workplace mobility (with ever-varying standards at different research institutions and within different projects and research groups), and performance pressure due to fixed-term contracts.

Potential

Besides the particular challenges, the prospects of Open Data and sustainable RDM structures hold great potential in particular for sport and exercise psychological research and neighboring sports science communities: The recording and assessment of human behavior is a costly and elaborate process, in particular when considering, for example, the population of elite athletes. This makes the obtained data particularly valuable for use and reuse. Further, collaborative projects already hold an important position in the sports sciences[7,8,9,10,11]. In the general scientific community, they form a quality measure for funding institutions[12] as well as for universities[13,14]. Sports science can play a pioneering role in this ever-growing sector because of its long-standing expertise in collaborative projects once seamless RDM infrastructure exists.

The "How" – Current State of Research Data Management

During the past few years, different conceptual frameworks, technical solutions, and initiatives have evolved targeting the increasing demands and needs for a sustainable RDM and supporting Open Data and FAIRification efforts. A few of these frameworks will be described in the following. As this description can clearly not be comprehensive in the context of this perspective article, we will focus on initiatives and developments of particular relevance for sport and exercise psychology and neighboring sports science communities.

[7] https://www.kiju-sport.nrw/forschung/projektuebersicht/
[8] https://www.sport.fau.de/das-institut/forschung/bewegung-und-gesundheit/forschungsprojekte/verbund/
[9] https://www.inprove.info/
[10] http://lauflabor.ifs-tud.de/doku.php?id=projects:projects_whitebox
[11] https://www.healthtech.kit.edu/59.php
[12] https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2021/04/3534_bekanntmachung.html
[13] https://www.rhein-main-universitaeten.de/forschung
[14] https://www.lmu.de/de/forschung/forschungsprojekte/index.html; https://epub.sub.uni-hamburg.de/epub/volltexte/2019/87302/pdf/HWWI_Research_Paper_181.pdf

Current Conceptual and Technical Developments

Different perspectives come into play when considering the conceptual and technical aspects of RDM. Researchers demand explicit guidelines on Open Science standards and practices, together with simple and automated RDM processes. These user demands, in turn, are being met by new developments on the data science side, of which we can mention only a small selection here: On a conceptual level, the Canonical Workflow Framework for Research (CWFR) emerged. This concept assumes that every research project ultimately follows defined workflow steps. FAIR Digital Objects[15] (FDO) connect individual workflow steps in the research process as a "glue." An enriched metadata layer in FDOs allows individual workflow steps to be machine-actionable and enables further operations such as licenses, access categories, anonymization level, etc. (see Figure 2). For instance, the technical implementation of this framework would allow data from data collection tools such as PsychoPy (Peirce et al., 2019) to be published directly into data repositories. To achieve this, merging disparate disciplinary metadata standards, licenses, and access classes will be required. While FDOs are currently still in the conceptual phase[10], Research Object Crates (RO-Crates) have already successfully taken the first steps toward implementing seamless RDM across the research lifecycle (Soiland-Reyes et al., 2022), including machine-actionable research data management plans (maDMP). RO-Crates, which are considered to be a type of FDO, pack research data with their metadata using, for example, schema.org annotations. RO-Crate was created by research communities to provide other researchers with a lightweight tool to make their research data FAIR.

A further development is the Open Research Knowledge Graph (ORKG, Auer et. al., 2020; Stocker et al., 2023). As a digital infrastructure for the production, curation, publishing, and reuse of machine-actionable scientific knowledge, it applies the FAIR principles to the scientific knowledge published in articles in order to enable the efficient reintegration of scientific knowledge as primary research data in research life cycles. For such Research Knowledge Graphs, the technical challenges primarily consist in the efficient production of machine-actionable scientific knowledge, quality assurance, and service usability.

Last, publishers demand technical solutions, focusing on securing quality and efficiency of RDM. However, the specific perspective might vary: Top-ranked journals review manuscripts and data separately and focus on technical solutions for workflows and data quality review processes (Peer et al., 2022). Other journals do not separate data from manuscripts and instead focus on the technical accessibility of datasets to reviewers and automated quality measurement.

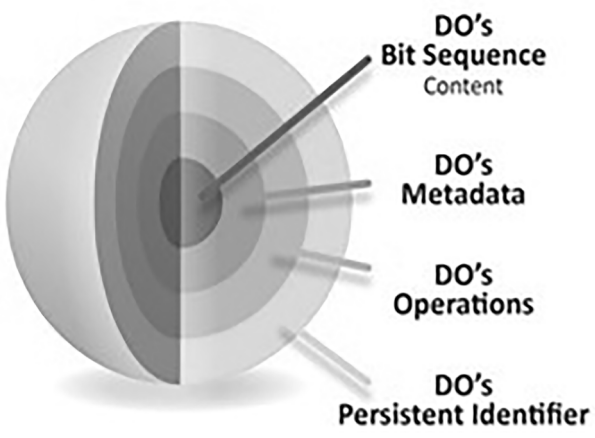

Figure 2. FAIR Digital Object, adapted from Jeffery et al. (2021).

Developments in the dvs

As reflected by the current and previous special issue on the topic (Renkewitz & Heene, 2019b), Open Science has become a topic of continuous, if not increasing, interest and relevance in the German sport and exercise psychology and sports science community. Open Data is a domain with a particularly fast growth in awareness and action in this context.

First, the sports science community has been prominently involved in a National Research Data Infrastructure (NFDI) consortia initiative since 2021/2022: "METHODS – National Research Data Infrastructure for Empirical Research on Human Behavior in Sports Science, Economics, Social Sciences, Psychology, and Medical Informatics." The NFDI is a funding scheme initiated by the German Research Foundation (DFG) in 2018, aiming to support the development of an infrastructure for sustainable RDM across research disciplines. The initiative gained support from central stakeholders in the sports science community, including the German Society of Sport Science (dvs) and its sections sport psychology (asp) and sport economics, the Federal Institute of Sport

[15] https://fairdo.org/

Science (BISp), as well as MO|RE data[16] and further individual sport and exercise psychologists and sports scientists. While the consortium initiative itself did not receive a funding recommendation, the potential and need particularly for sports science was emphasized by the reviewing board.

Second, in the follow-up of the application process, the participating and supporting sports science researchers, groups, and institutions formed an interest group to build on the consortium initiatives work. As an outcome of this ongoing exchange, the Executive Board of the dvs decided to establish an ad hoc committee "Research Data Management" in October 2022, to work on the topic and to develop a concept for the German sports science community.

Last, the increasing awareness for and relevance of this topic is also reflected, for example, by a pre-conference workshop and a subsequent session of oral presentations on RDM during the last biannual meeting of the dvs section "Sportmotorik," in September 2022. In this context, the MO|RE data project was presented as a flagship project for a discipline-specific data repository for motor skill data.

Perspectives of Open Data for Sports Science

When aiming at fulfilling the potentials of Open Data under the boundary conditions that apply to sport and exercise psychological and broader sports science research, a seamless and sustainable RDM infrastructure has to account for a range of different requirements: *First*, it requires a strict community and researcher orientation, with the guiding principle of respecting the communities' specificities. Thus, RDM solutions have to be strengthened and/or developed, which account for the specificity of the data, for example, with regard to data curation and annotation and data protection needs. Here, existing RDM solutions in sport and exercise psychology and sports science (e.g., data repositories such as MO|RE data) have to be identified, and, where required, interoperability with solutions of neighboring disciplines should be examined. In the field of medical informatics, for example, research data management practices have developed, including international IT, terminology and metadata standards[17], advanced anonymization techniques (PID for patients)[18], data access classification, trust systems, embargoes, and secure data centers to comply with data protection regulations. Further, various NFDI initiatives have been established in neighboring disciplines and are currently opening up for an integration of further communities. This includes KonsortSWD[19], with the Leibniz Institute for Psychology (ZPID) as participant, BERD@NFDI[20], and also NFDI4Health[21]. In addition, the EOSC[22] has to be mentioned as a European-wide platform supporting Open Data and Open Science efforts. However, to which extent these services are able to account for the specific needs of individual sports science and sport and exercise psychology communities remains to be worked out.

Second, any infrastructure should allow individual researchers to comply with the guidelines imposed by other stakeholders, for example, home institutions, funding bodies, or publishers. This requirement could be fulfilled by generic solutions, for instance, generic data repositories provided by the home institutions themselves or third-party providers (e.g., OSF[23], Zenodo[24]). However, those generic solutions might hinder the FAIRification of specific datasets. Simultaneously, and *third*, RDM solutions should account for varying Open Science and Open Data demands and standards between sports science disciplines, and with it facilitating and creating transparency about them.

Fourth, a sustainable RDM infrastructure should limit the temporal and financial load of individual researchers throughout the research process. Depending on the research project and the resources available, RDM can be a time-consuming process for the individual researcher since, to date, data documentation, for example, has to be done manually at each workflow step (so-called data-wrangling). Consequently, seamless RDM supporting Open Data requires not only a repository, where research data can be archived long term, but also, for example, a

[16] https://www.ifss.kit.edu/more/index.php
[17] https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative
[18] https://www.toolpool-gesundheitsforschung.de/produkte/pid-generator
[19] https://www.konsortswd.de/
[20] https://www.berd-nfdi.de/
[21] https://www.nfdi4health.de/
[22] https://eosc.eu/
[23] https://osf.io/
[24] https://zenodo.org/

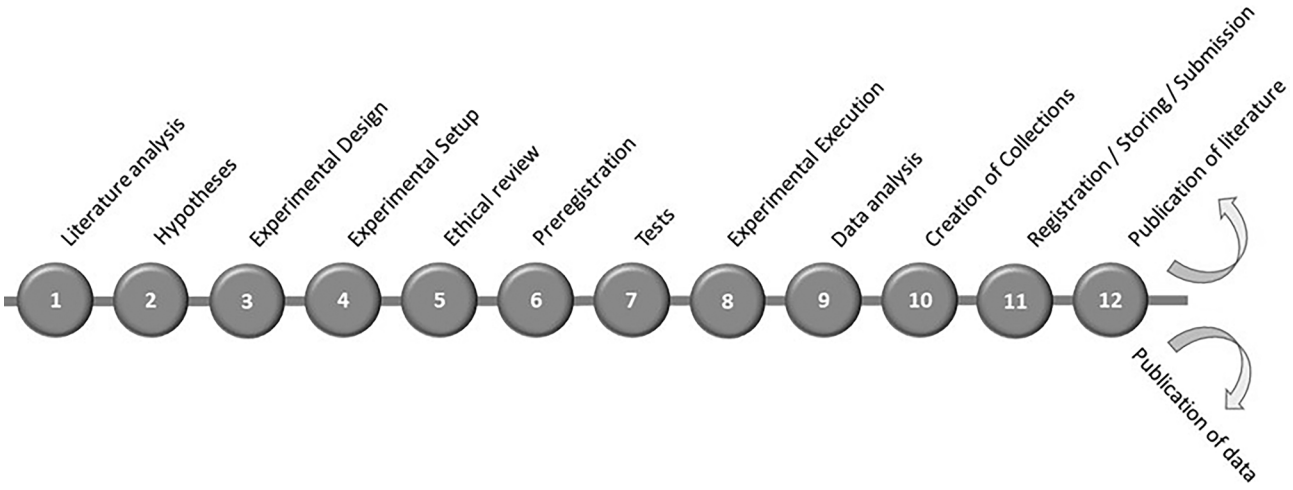

Figure 3. Research workflow, adapted from Jeffery et al. (2021).

technical infrastructure underlying the whole workflow, shared (meta-data) standards for data description, and importantly, built-in interfaces between data handling and management tools at each step along the workflow (see Figure 3). Effective RDM includes internal data exchange during the duration of, for example, collaborative sport and exercise psychology projects, and may also support data sharing after project finalization. Beyond that, Open Data refers to the reintegration of scientific knowledge as primary research data in subsequent research life cycles. Both could be achieved through an approach of data FAIRification "at birth," that is, from the beginning of the research data cycle, building on established tools already in use by the researchers. Besides these technical aspects, to acknowledge Open Data efforts and the accompanying temporal and financial investments, FAIR data publications should receive scientific recognition. *Last*, ethical and legal aspects, for example, with regard to data protection, have to be taken into account.

From our perspective, this can be best implemented by adopting a method-oriented approach: The research workflow, commonly shared by researchers empirically investigating human behavior in sport and exercise psychology and neighboring disciplines, can be typically described as a track of steps, as depicted in Figure 3. This workflow can serve as a kind of generic foundation (cf. CWFR) on which different (community) specific solutions can be built. This requires a conceptual openness for different standards and practices, as well as technical interoperability (see "Current Conceptual and Technical Developments"). In sport and exercise psychology, for example, this approach has already started to be taken up through the adoption of the researcher-driven PsychDS[25] standard for the description of scientific datasets, which uses metadata schemas from Schema.org, an open, community-driven process.

Building on the shared workflow to integrate already established yet disciplinary RDM solutions has the potential to comprehensively approach the aforementioned challenges: *First*, for sports science as a cross-sectional science, a variety of RDM solutions supporting Open Data practices will likely be required at each step along the workflow to fulfill the specific demands of the different sports science communities, thus no "one-solution-fits-all'' structure would be successful and sustainable. Importantly, individual sports science communities are already using tools and services of neighboring disciplines, which fit their particular demands. However, knowledge about these suitable solutions is not systematically available to the broader sports science community yet. Consequently, rather than developing new RDM structures, a distributed systems approach, integrating successful solutions along the different steps of the joint workflow, holds more potential for long-term use and, thus, sustainability. For this reason, the first step must be to identify the status quo regarding RDM knowledge, practices, and standards in different sports science communities. Based on this, generic as well as community-specific needs for efficient and sustainable RDM considering the entire research data cycle need to be identified.

The previous action would also pick up the *third* challenge of varying Open Data and RDM demands between studies in sports science communities. A marketplace providing RDM solutions along the research

[25] https://github.com/psych-ds/psych-DS

workflow with built-in interfaces enabling the interoperability of different tools would allow researchers to seamlessly adapt to these varying demands. It would also account for a more general challenge of adopting and implementing Open Science practices, namely, the costs (time and money) coming with it. Building on established solutions that are already in use by (sports) science communities and integrating them would reduce these accompanying costs.

Overcoming the challenges of RDM is particularly relevant to the practices and training of early career researchers. Depending on the aim of a study, the study population, and various other factors, completely different approaches to RDM might have to be chosen. This requires a researcher-oriented overview of the tools and standards that exist and are suitable for the particular demands, but also encouragement of early-career researchers to find their own position and to question it and others' positions critically, since divergences in the practices and standards of, for example, department heads and early-career researchers might exist. Future curricula at institutes of higher education should teach these diverse approaches to RDM, show how to implement different RDM practices, and encourage young researchers to find and strengthen their own position. As sport and exercise psychologists are often responsible for methodological training in many institutions and have particular expertise in RDM in collaborative research projects, they will play an important role in both training and implementation of RDM standards now and in the future.

Last but not least, the aforementioned aspects require the development of an RDM position, based on a concept that also identifies options for securing financial, technical, and user sustainability of the pursued RDM structure. This will be the focus of the newly established dvs ad hoc committee "Research Data Management".

Conclusion

In sum, Open Data holds great potential for the reintegration of sport and exercise psychological and sports scientific knowledge as primary research data in subsequent research life cycles of (inter- and trans-) disciplinary research projects. To this end, the sports science community has to develop a joint position on RDM, which acknowledges the heterogeneous boundary conditions, interests, and needs of the different sports science communities, for example, the particular sensitivity of data of elite athletes, to support the implementation of Open Science practices and standards. We argue that building on methodological workflows, shared within the sports science community as well as with neighboring disciplines, based on which existing RDM solutions and tools are then integrated, will make it possible to account for the specific challenges sports science faces as a cross-sectional science. In this context, the establishment of the dvs ad-hoc committee "Research Data Management" represents a recent development in taking up these challenges.

References

Auer, S., Oelen, A., Haris, M., Stocker, M., D'Souza, M., Farfar, K. E., Vogt, L., Prinz, M., Vitalis Wiens, & Jaradeh, M. Y. (2020). Improving access to scientific literature with knowledge graphs. *Bibliothek Forschung und Praxis,44*(3), 516 – 529. https://doi.org/10.1515/bfp-2020-2042

Baker, M. (2015). Over half of psychology studies fail reproducibility test. *Nature*, *27*, 1 – 3. https://doi.org/10.1038/nature.2015.18248

Betancort Cabrera, N., Bongartz, E. C., Dörrenbächer, N., Goebel, J., Kaluza, H., & Siegers, P. (2020). *White paper on implementing the FAIR principles for data in the social, behavioural, and economic sciences (No. 274).* RatSWD Working Paper. https://doi.org/10.17620/02671.60

Betz, D., Biniossek, C., Blanchi, C., Henninger, F., Lauer, T., Wieder, P., Wittenburg, P, & Zünkeler, M. (2022). Canonical workflow for experimental research. *Data Intelligence, 4*(2), 155 – 172. https://doi.org/10.1162/dint_a_00123

Boccali, T., Sølsnes, A. E., Thorley, M., Winkler-Nees, S., & Timmermann, M. (2021). *Practical guide to sustainable research data.* https://doi.org/10.5281/zenodo.4769703

Bös, K., Schlenker, L., Büsch, D., Lämmle, L., Müller, H., Oberger, J., Seidel, I., & Tittlbach, S. (2009). *Deutscher Motorik Test 6 – 18: (DMT 6 – 18)* [German Motor Test 6 – 18: (GMT6 – 18)] (Vol. 186). Czwalina

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., Orben, A., Parsons, S., & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. *Zeitschrift für Psychologie*, *227*(4), 237-248. https://doi.org/10.1027/2151-2604/a000387

German Research Foundation. (2022). *Open Science als Teil der Wissenschaftskultur. Positionierung der Deutschen Forschungsgemeinschaft* [Open science as part of research culture. Positioning of the German Research Foundation]. Zenodo. https://doi.org/10.5281/zenodo.7193838

Jeffery, K., Wittenburg, P., Lannom, L., Strawn, G., Biniossek, C., Betz, D., & Blanchi, C. (2021). Not ready for convergence in data infrastructures. *Data Intelligence, 3*(1), 116 – 135. https://doi.org/10.1162/dint_a_00084

Krüger, M. (2022). Sportwissenschaft: Zur Geschichte einer Querschnittswissenschaft [Sports science: On the history of a cross-sectional science]. In A. Güllich, & M. Krüger (Eds.), *Grundlagen von Sport und Sportwissenschaft* [Fundamentals of sport and sports science] (pp. 39 – 57). Springer Spektrum.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195 – 203. https://doi.org/10.3758/s13428-018-01193-y

Peer, L., Biniossek, C., Betz, D., & Christian, T. M. (2022). Reproducible research publication workflow: A canonical workflow framework and fair digital object approach to quality research output. *Data Intelligence, 4*(2), 306–319. https://doi.org/10.1162/dint_a_00133

Raab, M., Hoffmann, S., Laborde, S., & Lobinger, B. (2017). Von Vertrauenskrisen und Handlungsnormen zu Selbstreflexion und persönlichem Handeln [From confidence crises and action norms to self-reflection and individual actions]. *Zeitschrift für Sportpsychologie, 24*(3), 111–119. https://doi.org/10.1026/1612-5010/a000199

Renkewitz, F., & Heene, M. (2019a). The replication crisis and Open Science in psychology. *Zeitschrift für Psychologie*, *227*(4), 233–236. https://doi.org/10.1027/2151-2604/a000389

Renkewitz, F., & Heene, M. (2019b). *Topical issue: Open Science in psychology: Progress and yet unsolved problems.* Hogrefe Publishing.

Schönbrodt, F. D., & Scheel, A. (2017). FAQ zu Open Data und Open Science in der Sportpsychologie [FAQ on Open Data and Open Science in sports psychology]. *Zeitschrift für Sportpsychologie*, *24*(4), 133–176. https://doi.org/10.1026/1612-5010/a000217

Stocker, M., Oelen, A., Jaradeh, M. Y., Haris, M., Oghli, O. A., Heidari, G., Hussein, H., Lorenz, A.-L., Kabenamualu, S., Farfar, K. E., Prinz, M., Karras, O., D'Souza, J., Vogt, L., & Auer, S. (2023). FAIR scientific information with the open research knowledge graph. In B. Magagna (Ed.), *FAIR Connect* (Vol. 1, Issue 1, pp. 19–21). IOS Press. https://doi.org/10.3233/fc-221513

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., RO-Crate Community, Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, *5*(2), 97–138. https://doi.org/10.3233/DS-210053

UNESCO. (2021). *UNESCO recommendation on Open Science.* https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L., 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), 1–9. https://doi.org/10.1038/sdata.2016.18

Conflict of interest

The authors declare no conflict of interest.

Author contributions

All authors contributed equally to the process of drafting and revising the manuscript.

Funding

Open access publication enabled by Technische Informationsbibliothek (TIB).

ORCID

Melanie Krüger
https://orcid.org/0000-0001-5344-5060
Claudia Biniossek
https://orcid.org/0000-0002-2202-7875
Markus Stocker
https://orcid.org/0000-0001-5492-3212
Dirk Betz
https://orcid.org/0000-0002-6411-4758

Melanie Krüger
Institute of Sports Science
Leibniz University Hannover
Am Moritzwinkel 6
30167 Hannover
Germany
melanie.krueger@sportwiss.uni-hannover.de