

Reproducible Research Publication Workflow: A Canonical Workflow Framework and FAIR Digital Object Approach to Quality Research Output Postprint

Authors: Limor, Peer, Claudia, Biniossek, Dirk, Betz, Thu-Mai, Christian, Limor, Peer

Date: 2022-11-28T00:00:00+00:00

Abstract

In this paper we present the Reproducible Research Publication Workflow (RRPW) as an example of how generic canonical workflows can be applied to a specific context. The RRPW includes essential steps between submission and final publication of the manuscript and the research artefacts (i.e., data, code, etc.) that underlie the scholarly claims in the manuscript. A key aspect of the RRPW is the inclusion of artefact review and metadata creation as part of the publication workflow. The paper discusses a formalized technical structure around a set of canonical steps which helps codify and standardize the process for researchers, curators, and publishers. The proposed application of canonical workflows can help achieve the goals of improved transparency and reproducibility, increase FAIR compliance of all research artefacts at all steps, and facilitate better exchange of annotated and machine-readable metadata.

Full Text

Preamble

Reproducible Research Publication Workflow: A Canonical Workflow Framework and FAIR Digital Object Approach to Quality Research Output

Limor Peer^{1†}, Claudia Biniossek², Dirk Betz² & Thu-Mai Christian³

¹Institution for Social and Policy Studies, Yale University, Connecticut 06520, USA

²Center for Empirical Research in Economics and Behavioral Sciences (CEREB), University of Erfurt, Thüringen 99089, Germany

³Odum Institute for Research in Social Science, University of North Carolina System, North Carolina 27514-3916, USA

Keywords: FAIR research output; FAIR Digital Objects; Reproducible research; Research compendium; Compendium publication workflow; Data quality review

Citation: Peer, L., et al.: Reproducible research publication workflow: A canonical workflow framework and FAIR digital object approach to quality research output. *Data Intelligence* 4(2), 306-319 (2022). doi: 10.1162/dint_a_00133}

Received: September 4, 2021; **Revised:** February 4, 2022; **Accepted:** February 5, 2022

Corresponding author: Limor Peer (Email: limor.peer@yale.edu; ORCID: 0000-0002-3234-1593)

Abstract

In this paper we present the Reproducible Research Publication Workflow (RRPW) as an example of how generic canonical workflows can be applied to a specific context. The RRPW includes essential steps between submission and final publication of the manuscript and the research artefacts (i.e., data, code, etc.) that underlie the scholarly claims in the manuscript. A key aspect of the RRPW is the inclusion of artefact review and metadata creation as part of the publication workflow. The paper discusses a formalized technical structure around a set of canonical steps which helps codify and standardize the process for researchers, curators, and publishers. The proposed application of canonical workflows can help achieve the goals of improved transparency and reproducibility, increase FAIR compliance of all research artefacts at all steps, and facilitate better exchange of annotated and machine-readable metadata.

1. Introduction

In recent years, research funders and journals have increasingly expected researchers to share their data and code to encourage computational reproducibility and replicability [1, 2]. Open Science initiatives [3, 4] have further contributed to the expectation that research artefacts (i.e., data, code, etc.) be made available alongside the manuscript for review [5, 6]. A number of journals are implementing new policies that incorporate verification of computational reproducibility of results reported in manuscripts prior to and as a condition of publication, such as the *American Journal of Political Science* and the *American Economic Review* in the social sciences. These efforts are driven by the imperative to improve transparency and documentation of the processes that lead to

research results based on data, to verify and enable future reproducibility, to facilitate replication studies, and to increase the level of Findability, Accessibility, Interoperability, and Reusability, or “FAIRness” [7], at all steps. Additional goals include fostering transdisciplinary reuse and gaining new knowledge from existing data.

Promoting these goals requires redefining the evidence base that substantiates published data-based scholarly claims. While open data sharing practices are becoming more widespread, computational reproducibility sets even higher expectations for research transparency because it requires code sharing and execution. More specifically, it requires a set of digital files that supports the scholarly manuscript and includes the data, the code (software code of the conducted procedures, program files of the data analysis), and associated documents such as a codebook, README, and other supporting documentation and metadata. This set of files can be described as a research compendium [8] and represents a more complete scholarly record [9].

Publishing reproducible research requires an update to the manuscript publication workflow. While key components of the traditional scholarly publication workflow—manuscript submission, peer review, editorial decision, and publication—are well established, the review of associated files, documentation, data, and code as part of this workflow is a relatively new development [10], albeit one that we predict will become commonplace.

In this paper, we describe an updated manuscript review and publication workflow, the Reproducible Research Publication Workflow (RRPW). The workflow introduces procedures for quality review of the research artefacts necessary to reproduce the article’s findings (i.e., data, code, etc.), while also considering these artefacts as bound together into a unitary object for dissemination, interpretation, and reuse. We apply the Canonical Workflow Framework for Research (CWFR) approach to formalize the compulsory steps involved in the publication of reproducible research and identify recurring steps associated with artefact review. The CWFR approach requires codification of canonical steps and adherence to technical standards that support and enforce FAIR principles [11].

Following NASEM (2019), we define reproducibility as “obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis. This definition is synonymous with ‘computational reproducibility’.” We define replicability as “obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.”

We also harness the technological affordances of the FAIR Digital Object (FDO) structure [12] to establish the compendium itself as a robust digital object. The objective is to facilitate seamless reuse of high-quality research output in the future by: (a) improving transparency and reproducibility of the research process, (b) increasing FAIR compliance of all research artefacts at all steps, and (c) enabling the exchange of annotated and machine-readable metadata produced

by component parts and participating platforms. One overarching aim of the CWFR and RRPW is to highlight key automated steps throughout the research lifecycle to reduce the burden for all stakeholders along the way to producing published reproducible research.

2. Efforts to Ensure the Reproducibility of Published Research Results

We argue that the publication workflow should include a quality review of the full research compendium for the purpose of verifying computational reproducibility and enhancements to metadata in order to follow the FAIR principles. The workflow must be informed by researchers, data curators and archivists, publishers, and infrastructure experts.

A manuscript publication and compendium review process with the stated goal of computationally reproducing the claims reported in the associated manuscript requires new methods and processes. Christian et al. [10] describe a process consisting of six essential, or canonical, steps (Figure 1 [Figure 1: see original paper]): A manuscript submitted for peer review may lead to a conditional acceptance (1, 2) which triggers a request to submit the compendium containing data, code, software, and documentation (3). The materials undergo a curation and reproducibility verification process which, if successful (4), may lead to final acceptance of the article (5). This then leads to publication of the manuscript in appropriate journals and publication of data and software in designated data repositories (6). This process is used by several social science journals that publish quantitative research using statistical analysis methods. It is important to note that the generic nature of this process allows for parallel or inverse steps; for example, some high-ranking journals may require data and code submission before or at the same time as manuscript submission.

Peer et al. [13] describe in detail the activities performed as part of the curation and verification process. The goal of these activities is to ensure that digital artefacts supporting a scholarly claim meet quality standards for reproducibility and for sharing and archival preservation. The review activities are highly asynchronous and require interactions between humans (e.g., curators and authors) which should be supported by software where possible to increase efficiency. In particular, the data review and code review steps may present significant challenges [14]. Code review and reproducibility verification, for example, could lead to identification of errors which would then lead to new calculations and a new paper submission, increasing the overall review time.

One mechanism that can help implement steps in this typical workflow is the Yale Application for Research Data (YARD). YARD manages and tracks metadata production and other quality review activities that contribute to “FAIRer” research and exposes the review process to scrutiny [15, 16]. Figure 2 [Figure

2: see original paper] illustrates the actions of both machine and human actors (i.e., curators and authors) and the interactions between these two types of actors. Note that some curators' activities are performed "offline," meaning the curator uses manual or (semi-)automatic procedures that return "quality indicators" including formal specifications allowing machines to act. YARD and similar tools can ensure the compendium as a whole, and its component parts, hold necessary properties of a FAIR Digital Object that allow distributed actors to execute canonical steps.

3. The Reproducible Research Publication Workflow

The Reproducible Research Publication Workflow (RRPW) describes how the research compendium is processed during the manuscript submission, review, and publication workflow. The RRPW embeds quality review of the research compendium into the manuscript review and publication process, building on the process described in Christian et al. [10] and Peer et al. [13].

It is important to note that the RRPW is adjacent to the preceding data generating and data analytic workflows, the parallel data management workflow (with substantial intersections), and the subsequent transdisciplinary data reuse workflows. While the RRPW aims to achieve publication of reproducible research, it is not solely the purview of publishers. Indeed, the RRPW relies on authors and allied professionals, such as data curators and stewards, to contribute metadata from the beginning of the research lifecycle in order to achieve metadata-rich published materials. In comparison to other workflows, the RRPW has special features. Unlike the typical data-generating workflow, the RRPW involves additional actors aside from the researcher who performed the data collection and analysis presented in the article. These external actors may include individuals from the researcher's own institution, the journal, or the data archive, who each execute assigned RRPW tasks. This is largely dependent on funder requirements, the data management plan (DMP), quality assurance procedures, and the curation and archiving strategy.

3.1 RRPW and FDOs

Ideally, the research compendium not only contains the requisite files but also encapsulates all other elements that would qualify it as a FAIR Digital Object (FDO). FDOs are machine-actionable units that bundle the data with all components necessary for identifying, rendering, interpreting, and accessing the data [12]. They are represented by a bitstream, referenced and identified by a persistent ID, and have properties described by metadata. The FDO model for research specifies the properties of these data packages that satisfy both FAIR principles and the needs and expectations of the scholarly community. As an application of the CWFR framework, the RRPW is designed to output research compendia as FDOs [17, 18] by establishing the necessary activities to ensure

that the research compendium and its component parts are identified by universally unique, persistent, and resolvable identifiers, that they are associated with comprehensive metadata, and that they are stewarded by a trusted repository for long periods of time.

The RRPW establishes standardized processes for subsequent data sharing and reuse workflows, thus increasing efficiency and thereby reducing the cost of labor-intensive data curation and verification procedures. In doing so, the RRPW can help streamline these tasks (which are often decentralized, with workflow steps distributed across various actors working in their own institutions) and the interactions and/or dependencies among them. Moreover, it generates a portable FDO, which enables external actors to engage with the canonical workflow as required while ensuring that the research community has access to the FDO.

The RRPW also helps anticipate the subsequent data reuse workflow and address any special processing needs for data intended for transdisciplinary reuse (e.g., digitization of cultural heritage in archaeology). Here, additional workflow steps are necessary to capture comprehensive contextual information in a Project-Metadata-Digital Object, or MD-DO after data generating. The metadata schema must also make the digital objects findable and accessible for researchers from all target disciplines. If transdisciplinary reuse is intended, these criteria must be incorporated into the quality assessment and reflected in the appraisal system and batches. Above all, the involvement of other disciplines raises the question of overarching standards and rules.

To the extent that journals are using repositories committed to long-term preservation and archiving in accordance with industry standards, e.g., ISO 14721: The Reference Model for an Open Archival Information System [19], they are already making use of standard actions that can have different implementations. For example, journals may have different criteria around file formats, metadata, the use of licensed statistical software, repositories for storing the artefacts, methods of linking to these repositories, and different workflow management protocols, including whether to involve third parties. A CWFR framework—which unifies standard actions and workflow technologies, as well as facilitates FAIR principles—is useful for the manuscript publication and artefact review workflow.

3.2 RRPW Canonical Components

We use the same diagram style employed in other cases included in this journal issue to introduce the transition to FDOs (Figure 3 [Figure 3: see original paper]). We indicate human responses by pale blue boxes and use the term “package” (instead of “file” and “config file”) which is defined by the MD-DO and refers to all relevant documents including those added by curators. To simplify, we do not indicate all possible feedback loops where, for example, a data curation step leads to a request for authors to improve metadata. It should be noted, however, that researchers do not have to start from scratch

in certain cases but could refer to the appropriate MD-DO object and adapt it or its components. Some updates are so simple they can be carried out in the “perform updates” action (see below).

The process comprises canonical steps and associated actions, potentially performed by an editorial system. The MD-DOs, capturing the state at each step and replacing the “configuration” file in RRPW, are first-class citizens on the Internet, are themselves FAIR, and not hidden in some tool-driven database. Actions can be amended with specific library packages that deal with special requirements dependent on institution and community. Of course, it could be a simple and temporary modification first to maintain the configuration file and to create in addition an MD-DO. Crucial for “FAIRness” is that all references made in the MD-DOs are based on Handles that are well-supported. This suggestion extends to the reviewer-author interaction, which can also be collected in an FDO that is cumulatively growing and referred to by the MD-DO.

The following are the various steps in this canonical workflow, indicated in Figure 3 by the seven blue arrows, and associated actions, represented below in bullet points.

Step 1: Start CWFR project. A researcher initiates a reproducible research publication project by submitting a paper to an editorial system, resulting in a first MD-DO after having prepared all needed information (i.e., the research compendium), including metadata (assuming journal policy and guidelines supporting FAIR are established). - Researchers submit paper to an editorial system
- System assigns an ID number to the manuscript

Step 2: Request paper review. The paper component of the package is sent to reviewers and their comments are collected. After some iterations, the review process may result in approval. The reviewers’ comments are aggregated in a Review FDO which is also referred to via the MD-DO. - Editorial system initiates peer review process for manuscript (editor selects reviewers, including reviewer comments)

Step 3: Ingest package. The user uploads all information via an ingest front-end of a curation tool or curation-enabled platform into a temporary store (workspace), where the package is established as a record (i.e., an abstract representation of the package, its components, and their relationship). The uploaded package is given initial archival processing treatment to enable review of the package as an FDO, and the MD-DO is enhanced to include the reference to the package FDO. In addition, the usual information is added such as a time stamp, etc. (note: might be sequential to or part of Step 1): - Establish the record in the system; assign an ID number to the package as a unique record (persistent identifier, or PID, including the version) - Create metadata for record
- Assign an ID number to the files as unique objects (PID) within the record (versions) - Establish relationship among the files (e.g., via metadata) - Create metadata for each file/component (e.g., file type) including checksum - Link all relevant files to manuscript to define the record (establish a compendium via

packaging system/software or container) - Identify unstructured documentation (e.g., README, codebook, data dictionary)

Step 4: Request data and code review. An action to initiate a data and code review is taken. - Editorial system initiates quality review process for research compendium (editor selects reviewers [could be peer or specialized], including reviewer comments)

Step 5: Perform curation. Action is taken to verify the computational reproducibility of data and code, including some iterations as indicated by the green arrow. This may result in updates to the package, such as new metadata, files, or versions of the manuscript (note that the audit trail of curation actions and updates should be linked with the FDO to document the process). This action will eventually lead to final acceptance of the whole submitted package. - Enhance metadata

Step 6: Publish all. Upon approval, the newly created package—which is associated preferably with a final FDO including all information and references needed for reuse—is deposited into a trustworthy repository. - Push the package to a repository via API (capture actions) - Publish manuscript on journal website - Issue a bi-directional link manuscript to research compendium

Step 7: End CWFR project. After some checks the project may be closed. - Apply long-term preservation policy (if any)

Recall that the concept of FDO always includes an associated PID (e.g., Handle or DOI) and metadata describing the nature of the FDO's bitstream. In addition, to ensure compliance with data protection laws and still be able to automate the reproducible research publication workflow to the greatest extent, we argue that it is indispensable to tag the FDO with additional information. For example, the Weblicht software [20] might be used to define an error standard.

3.3 Extensions of the Generic RRPW

As indicated above, the RRPW may vary by community. Here we mention four cases that nevertheless demonstrate the generality of the chosen approach.

Case A: Access to data. The volume of data may be so high, or access may be otherwise restricted, that copying them is not recommended or practical. In this case the MD-DO would include a reference to externally stored data; i.e., the package would include some metadata but not the data. The workflow would proceed as usual with two exceptions: (1) The data review would need access to the remotely stored data and ability to execute tests on them, and (2) the publishing step would need to include a check of repository quality, i.e., determine whether it is a trustworthy one (i.e., certified as CoreTrustSeal) that supports “FAIRness.” In case of a negative result, a warning needs to be sent to the researcher and repository team to take remedial action per repository policies. Finally, this step would result in publishing the FDO that may include references to external repositories.

Case B: Multiple datasets. For example, in the social sciences, data on large-scale surveys such as a census or nationwide representative study are generated by multiple research groups in independent projects. These datasets are assembled into collections, meaning that a data file and its metadata descriptions are replaced by a collection DO and additional files. However, this does not change the canonical workflow. The collection DO resolves into its components to track the individual components of the previous canonical workflow.

Case C: Standard software. In many research areas, such as materials science, a few large and well-known software packages are used by the community to generate simulation results data. Such software packages undergo extensive review by their creators. The code review component would be simplified to a formal check of whether the software package is mentioned in some list of community-accepted software. In other research areas, computations are performed by individually created scripts or programs that require a separate review process. We note that evaluating the reproducibility of program code must account for variability in some parameters; for example, the same source code of a machine learning program might lead to different sets of outputs and weights depending on the machine [21].

Case D: Research with human subjects. In research with human subjects (e.g., clinical trials), data may more often be subject to restrictions due to ethical concerns rather than size constraints. More complex or additional steps must be taken for ethical reasons or because of privacy laws, some of which may be difficult to automate (e.g., ethics review, curation). However, this does not fundamentally change the basic structure of the workflow.

4. Discussion

In this paper we present the Reproducible Research Publication Workflow. A key aspect of the RRPW is the inclusion of artefact review and metadata creation as integral to the publication workflow. As noted by Velterop and Schultes [22], “the most important role of publishers and preprint platforms is to ensure that detailed, domain-specific, and machine-actionable metadata are provided with all publications...[and] all ‘research objects’ that they publish.” The RRPW does not place the onus on publishers, or any single actor, alone. On the contrary, object and process metadata is accumulated and updated along the way in collaboration with authors and allied professionals such as data curators, stewards, and research software engineers, leading to metadata-rich and therefore more reusable published materials. Indeed, we encourage such professionals to bring their expertise to bear on the CWFR.

In the context of quantitative research that uses statistical analysis methods, the RRPW can be considered a canonical workflow following the tradition of the CWFR. First, it includes recurring patterns with common practices “frequently resulting in fragmented and potentially irreproducible sequences that

mix manual and machine-based steps” [23]. These actions are carried out by researchers, publishers, and curators, each of whom can have different implementations dependent on their role. Second, there is potential for integrating these actions into a CWFR by developing operations that follow import and export standards and that can be put into libraries and then reused when needed. Third, the RRPW supports FAIR principles by, for example, ensuring the assignment of PIDs and the creation of rich comprehensive metadata including provenance information.

This CWFR approach to publishing reproducible research has several advantages. First, the RRPW is a canonical, generic workflow that constitutes a standard promoting data and code reuse through effective quality control and metadata. By elevating digital objects to a higher level of FAIR compliance, the RRPW contributes to research reproducibility and transparency. The RRPW facilitates accounting of all digital objects, whether they are FAIR or not.

Second, the RRPW has the potential to make a large number of digital objects visible and FAIR and represents a step toward the ability to handle a growing volume of digital objects using protocols and standards. Canonical components based on routine standards and protocols can enable more automated and scaled publication of reproducible research and contribute to the development of the Global Interoperable Data Space (with DOIP type of interactions) [12]. Moreover, the RRPW has the potential not only to accommodate multiple technologies but to inspire improved interoperability between those tools to better fit the standard workflow.

Third, the workflow can be adapted at any point subject to the overarching goals of producing a digital object that is permanent and freely accessed and includes rich annotated and machine-readable metadata so it can be interpreted, understood, and used by the Designated Community without having to resort to special resources not widely available. For example, the paper illustrates how the RRPW operates in the context of quantitative social science research, and we encourage other communities (e.g., qualitative methods, hermeneutics research) to consider relevant component libraries. Importantly, a core component of the RRPW is recognizing and codifying the complex relationships between scholarly findings and supporting artefacts, yet it is flexible enough to accommodate future representations of the findings (i.e., other than the traditional form of publication via PDF) and artefacts of various kinds.

Fourth, researchers interested in using published research outputs will benefit from automated procedures for accessing these outputs, especially if they have small budgets. Researchers participating in the RRPW are encouraged to produce FAIR research output at the onset of a research project, thus reducing the burden of creating FAIR Digital Objects at the end of the research lifecycle.

Finally, and crucially, the RRPW can help address real-world challenges by promoting interoperability for heterogeneous data via transdisciplinary exchange of DOs. For example, the Virus Outbreak Data Network (VODAN) is committed

to making SARS-CoV-2 virus data FAIR to enable harnessing “machine-learning and future AI approaches to discover meaningful patterns in epidemic outbreak” [24]. Other examples from environmental and life science are discussed in Harjes et al. [25].

We note that it might be challenging in some cases to carry out review of data and software due to their specific nature and the specialized knowledge required. For studies with smaller budgets, for example, it might be beneficial to reduce the number of curation steps. However, we maintain that the proposed application of the CWFR will help achieve the goals of transparency and reproducibility, increased FAIR compliance of all research artefacts at all steps, and the exchange of annotated and machine-readable metadata.

Acknowledgements

The authors thank Peter Wittenburg for early contributions to this paper, Vicky Rampin for critically reading the manuscript and suggesting improvements, and two reviewers for their thoughtful comments. Limor Peer and Thu-Mai Christian acknowledge funding from the Institute of Museum and Library Services (RE-36-19-0081-19).

Author Contributions

All authors have made valuable and meaningful contributions to the manuscript. L. Peer (limor.peer@yale.edu) led the writing of the paper; C. Biniossek (claudia.biniossek@methods4nfdi.org), D. Betz (dirk.betz@methods4nfdi.org), and T.-M. Christian (thumai@email.unc.edu) contributed ideas, text, and comments in the production and review of the manuscript.

References

- [1] Munafò, M.R., et al.: A manifesto for reproducible science. *Nature Human Behaviour* 1(1), Article number 0021 (2017)
- [2] National Academies of Sciences, Engineering, and Medicine (NASEM): Reproducibility and replicability in science. The National Academies Press, Washington, DC (2019)
- [3] UNESCO Recommendation on Open Science. Available at: <https://en.unesco.org/science-sustainable-future/open-science/recommendation>. Accessed 16 December 2021
- [4] Nosek, B.A., et al.: Promoting an open research culture. *Science* 348(6242), 1422–1425 (2015)
- [5] Crosas, M., et al.: Data policies of highly-ranked social science journals. arXiv preprint SocArXiv: 10.31235/osf.io/9h7ay (2018)

- [6] Willis, C., Stodden, V.: Trust but verify: How to leverage policies, workflows, and infrastructure to ensure computational reproducibility in publication. *Harvard Data Science Review* 2(4) (2020). Available at: <https://doi.org/10.1162/99608f92.25982dcf>. Accessed 4 February 2022
- [7] Wilkinson, M.D., et al.: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(1), Article No. 1 (2016)
- [8] Gentleman, R., Temple Lang, D.: Statistical analyses and reproducible research. *Journal of Computational and Graphical Statistics* 16(1), 1-23 (2007)
- [9] Claerbout, J.F., Karrenbach, M.: Electronic documents give reproducible research a new meaning. *SEG Technical Program Expanded Abstracts* 1992, 601-604 (1992). Available at: <https://doi.org/10.1190/1.1822162>. Accessed 4 February 2022
- [10] Christian, T.-M. L., et al.: Operationalizing the replication standard. *International Journal of Digital Curation* 13(1), Article No. 1 (2018)
- [11] The CWFR Group: Canonical workflow framework for research. Position paper, version 2 (2020). Available at: <https://osf.io/9e3vc/>. Accessed 5 February 2022
- [12] De Smedt, K., Koureas, D., Wittenburg, P.: FAIR digital objects for science: From data pieces to actionable knowledge units. *Publications* 8(2), Article No. 2 (2020)
- [13] Peer, L., Green, A., Stephenson, E.: Committing to data quality review. *International Journal of Digital Curation* 9(1), Article No. 1 (2014)
- [14] Peer, L., et al.: Challenges of curating for reproducible and FAIR research output. Available at: <https://zenodo.org/record/5094155#.Yf6qnt9OmUl>. Accessed 5 February 2022
- [15] Peer, L., Wykstra, S.: New curation software: Step-by-step preparation of social science data and code for publication and preservation. *IASSIST Quarterly* 39(4), Article No. 4 (2016)
- [16] Peer, L., Dull, J.: YARD: A tool for curating research outputs. *Data Science Journal* 19(1), Article No. 1 (2020)
- [17] Group of European Data Experts in RDA (GEDE-RDA), Paris-FDO-workshop, 2020. Available at: <https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects/Paris-FDO-workshop>. Accessed 5 February 2022
- [18] Jeffery, J., et al.: Not ready for convergence in data infrastructures. *Data Intelligence* 3(1), 116-135 (2021)
- [19] Reference model for an open archival information system (OAIS). Available at: <http://www.oais.info/>. Accessed 5 February 2022
- [20] Hinrichs, E., Hinrichs, M., Zastrow, T.: WebLicht: Web-based LRT services for German. In: *Proceedings of the ACL 2010 System Demonstrations*, pp. 25-29 (2010)
- [21] Chassanoff, A., Altman, M.: Curation as 'Interoperability with the Future': Preserving scholarly research software in academic libraries. *Journal of the Association for Information Science and Technology* 71(3), 325-337 (2020)
- [22] Velterop, J., Schultes, E.: An academic publishers' GO FAIR implementation network (APIN). *Information Services & Use* 40(4), 333-341 (2020)

- [23] Wittenburg, P., et al.: Canonical workflows to make data FAIR. *Data Intelligence* 4(2), 286-305 (2022)
- [24] Mons, B.: The VODAN IN: Support of a FAIR-based infrastructure for COVID-19. *European Journal Human Genetics* 28(6), Article No. 6 (2020)
- [25] Harjes, J., et al.: FAIR digital objects in environmental and life sciences should comprise workflow operation design data and method information for repeatability of study setups and reproducibility of results. *Database* 2020, Article ID baaa059 (2020)

Author Biography

Limor Peer, Ph.D., is Associate Director for Research and Strategic Initiatives at the Institution for Social and Policy Studies (ISPS), Yale University. Limor works on research transparency and reproducibility and is especially interested in the connection between generating and preserving scientific knowledge. Limor created the ISPS Data Archive, a digital repository for research produced by scholars affiliated with ISPS with a focus on experimental design and methods. She led the project to develop YARD, the Yale Application for Research Data, a workflow tool for reviewing and enhancing research outputs. Limor is co-founder of the CURE (Curation for Reproducibility) Consortium of social science data archives. ORCID: 0000-0002-3234-1593

Claudia Biniossek is a Senior Researcher at TIB - Leibniz Information Centre for Science and Technology Hannover, and co-spokesperson of the METHODS NFDI-consortium project. She is affiliated with the Center for Empirical Research in Economics and Behavioral Sciences (CEREB) at the University of Erfurt. Her research interests lie in bringing together theories, methods, and data infrastructures in a transdisciplinary approach to improve research on human behavior. Together with Dirk Betz, she developed and established the repositories for behavioral economics (x-econ.org) and experimental social sciences (x-science.org). ORCID: 0000-0002-2202-7875

Dirk Betz is a Senior Researcher at TIB -Leibniz Information Centre for Science and Technology Hannover, affiliated with the Center for Empirical Research in Economics and Behavioral Sciences (CEREB) at the University of Erfurt, and a co-spokesperson of the METHODS NFDI-consortium project. Together with Claudia Biniossek, he developed the repositories x-econ and x-science for data from experimental economics and social sciences. The purpose is to test the basic principles of human decision-making in the field of experimental data coming from economics, sociology, political sciences, psychology, and neuroscience to bring together theories, methods, and data infrastructures within a transdisciplinary approach. ORCID: 0000-0002-6411-4758

Thu-Mai Christian is the Assistant Director for Archives at the Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill. She is responsible for the day-to-day operations of the Odum Insti-

tute Data Archive as well as establishing and enforcing policies in accordance with archival standards and best practice. Thu-Mai received her M.S. in Information Science with a concentration in Archives and Records Management and is currently pursuing a Ph.D. in Information and Library Science from the University of North Carolina. ORCID: 0000-0002-3658-9692

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.