

Increasing Public Access to the Results of USDOT-Funded Transportation Research

Docket No. DOT-OST-2023-0045

Document Number:2023-06373

<https://www.federalregister.gov/documents/2023/03/28/2023-06373/increasing-public-access-to-the-results-of-usdot-funded-transportation-research>

Comments from University of Minnesota Libraries (UL)

Authors:

Alicia Hofelich Mohr, PhD, Research Support Coordinator (LATIS, an integral UL partner)
Shanda Hunt, Public Health Librarian and Data Curation Specialist (UL)
Shannon Farrell, Interim Research Data Services Librarian (UL)
Allison Langham-Putrow, Scholarly Communications and Engineering Liaison Librarian (UL)

Submitted via email to public.access@dot.gov

1. How best to improve access to textual research outputs. A high percentage of DOT funded research results are delivered via technical reports, research briefs, manuals, technology transfer documents, and other grey literature, designed for immediate sharing and rapid implementation. The current DOT Public Access Plans allows researchers to distribute these outputs through the website or repository of their choice, and requires a copy be submitted to the DOT National Transportation Library digital repository for long-term preservation and public access. DOT seeks information on how to improve and streamline this submission process to improve timeliness; and, to avoid reinforcing inequities to access and submission, while not creating new ones.

If DOT seeks to streamline the submission process, we recommend prioritizing the submission to the National Transportation Library over other locations. Putting materials into one location and assigning one DOI will reduce burden on researchers who may be depositing work in multiple locations and make it easier for users of the research to find materials. This would align DOT with other federal funding bodies, such as NIH requiring submission to PubMed Central or NSF to their Public Access Repository.

3. How best to improve access to scholarly publications from DOT funded research. Section 3.a) of the 2022 OSTP memo calls on agencies to “update or develop new public access plans for ensuring, as appropriate and consistent with applicable law, that all peer-reviewed scholarly publications authored or co-authored by individuals or institutions resulting from federally funded research are made freely available and

publicly accessible by default in agency-designated repositories without any embargo or delay after publication.” DOT seeks information on:

i. How peer-reviewed scholarly publications should be made publicly accessible

Although researchers are allowed to use federal research grant funding to pay for article processing charges (APCs), we recommend de-emphasizing this route. DOT should include explicit language in its policy that states that APCs paid to publishers are not the preferred route for compliance with the updated policy.

Some publishers have been making public statements against the new public access policy requirements, suggesting that they must charge high APCs in order to maintain operations. However, we encourage DOT to provide pressure against this push by publishers, which has the potential to move the scholarly publishing system to one in which all authors must pay a fee to share their research. Such a system would prevent the many researchers who do not have funding from sharing their research, which would be detrimental to *all* researchers, including those funded by the DOT. There has been much discussion over the last few years about the negative effects of an "all must pay to publish" system. This was a major theme at the 2023 United Nations Open Science Conference (<https://www.un.org/en/library/OS23>). If the world aims to address the challenges facing our planet, (e.g., achieving the United Nations' Sustainable Development Goal), researchers from around the world must be able to share their research. DOT should establish policies that promote equitable publishing models and proactively avoid potential adverse outcomes.

"Diamond" open access is an equitable publishing model in which no one has to pay to read and no authors pay to publish. Support for diamond open access is growing, as evidenced by a recent statement of support for such equitable models from Deans at some of the most prestigious universities in the US (<https://libraries.mit.edu/news/libraries-support-3/34036/>) and researchers in the United Kingdom (<https://docs.google.com/document/d/1ZAIPDvECb5Zm1pqAf0l1f0sjcBqPbkPGMvGIhaCz6IM/e/dit#>). DOT should consider providing support, potentially working in concert with organizations in Europe that are exploring how to support diamond open access. Science Europe, cOAlition S, OPERAS, and the French National Research Agency jointly developed an Action Plan for Diamond Open Access with steps that DOT, in partnership with other federal funding bodies, could consider undertaking to support this open access model (<https://zenodo.org/record/6282403#.ZDhEvXbMI2w>).

DOT could support equitable publishing models by providing direct support to diamond open access publishers and educating DOT-funded researchers about these options. Additionally, DOT could adjust grant application assessments to reward publication in journals following equitable models over "big name" journals that often have high APCs (e.g., the average APC for Elsevier's Transportation Research journals is nearly \$3000).

Alternatively, DOT could build upon the current policy, which requires only the deposit of the final text of scholarly publications into the repository already designed for this purpose, the National Transportation Library, and specifically the Repository and Open Science Access

Portal (ROSA-P). For many journals, current default publisher policies are aligned with the existing federal public access policies, automatically granting authors the right to share their publications in a non-commercial repository 12 months after publication. Publishers may or may not respond to the updated plans resulting from the 2022 OSTP memo by granting authors the right to share articles immediately upon publication. To address this potential issue, DOT could consider developing language that entitles researchers the right to deposit their article in the DOT repository as an automatic condition of the grant award. For example, one of the methods to comply with Plan S open access requirements is the Rights Retention Strategy developed by cOAlition S (<https://www.coalition-s.org/>). This provides funded researchers with specific language to use to enable researchers to publish in the journal of their choice, without needing to pay an APC.

Finally, accessibility requires that researchers are able to find publicly available materials. To support findability, we recommend that the DOT repositories are optimized for indexing by online search engines, as other repositories are. The University of Minnesota manages AgEconSearch and actively works with platform developers to ensure that materials in the repository appear at the top of search results.

ii. How to maximize equitable reach of public access to peer-reviewed scholarly publications, including by providing free online access to peer-reviewed scholarly publications in formats that allow for machine-readability and enabling broad accessibility through assistive devices

PDFs are a common format for peer reviewed scholarly publications, yet are not as machine readable nor as accessible for all users and purposes, compared to other formats. Therefore, we suggest that DOT require publishers to share publications with complete Extensible Markup Language (XML) metadata. XML is more flexible and reusable than PDF or HTML, which will facilitate machine-readability to support findability, aggregation, and text mining. This requirement could be implemented in phases, starting with requiring that the abstract and citation data be provided in this format, as PubMed Central does, before implementing a requirement for the full text of publications.

DOT should also consider encouraging the use of Journal Article Tag Suite (JATS) XML. Although it is the industry standard, not all publishers are providing JATS XML. DOT could consider providing support for the development of free and open software that integrates JATS XML into the authoring and reviewing stage to enable smaller publishers to comply with a new requirement.

However, if text material is for some reason not "born digital", we recommend that it be shared in PDF form, with optical character recognition applied to allow for text recognition.

iii. The circumstances or prerequisites needed to make the publications freely and publicly available by default, including any use and re-use rights, and which restrictions, including attribution, may apply.

Researchers may be subject to various requirements from funders when assigning reuse rights to their work. For example, US federal government employees' works do not have a copyright, while Plan S (www.coalition-s.org/plan_s_principles/) requires all publications resulting from funded work to be published open access with a CC-BY (Creative Commons Attribution license), which only applies to works which have a copyright. The current federal public access policies do not require any particular license, such as specific Creative Commons licenses. However, DOT should be clear in the updated policy as to whether a specific license will now be required. And allowing a range of licenses including CC0 (the Creative Commons Public Domain Dedication, which is compatible with works that don't have a copyright), would reduce conflict for authors.

Regardless of specific license selected or recommended, we encourage DOT to disambiguate reuse guidelines by making license information about articles and data both human and machine readable.

4. How best to improve access to datasets. The 2015 DOT Public Access Plan required all data underlying research conclusions be made publicly accessible, while protecting sensitive personal, business, and security information. Further, the Plan required research proposals include a data management plan (DMP) that, among other things, detailed where datasets would be preserved. However, DOT allowed researchers a choice of where to preserve the data: in an institutional or third-party domain-specific or generalist repository; with DOT; or, to self-distribute data when requested by the public. Further the 2015 DOT plan allowed researchers to include reasonable preservation costs in their research proposal. Going forward, the updated plan will continue to mandate research data must be shared while protecting sensitive information. However, in order meet the requirements of the 2022 OSTP memo and to better ensure long-term preservation of data of interest to DOT, the broader research community, and the public, researchers must preserve data in an institutional or third-party repository or with DOT, but self-distribution will no longer be allowed. DOT will continue to encourage researchers to plan for, and budget for, long-term data preservation as part of the research proposal process. DOT seeks information on how to best facilitate, support, and fund long-term data preservation and sharing.

We applaud the change in no longer allowing self-distribution of data and mandating sharing in repositories. We suggest that DOT specifically recommend that utilized repositories meet the White House's "Desirable Characteristics of Data Repositories for Federally Funded Research" (<https://doi.org/10.5479/10088/113528>). As we saw with the recent release of the NIH Data Management and Sharing Policy, forecasting the cost for data management and long-term sharing is difficult and will require distribution of specific guidance for researchers. See examples here: Forecasting Costs for Preserving, Archiving, and Promoting Access to Biomedical Data (<https://www.nationalacademies.org/our-work/forecasting-costs-for-preserving-archiving-and-promoting-access-to-biomedical-data>) from National Academies of Science, Engineering, and Medicine; and COGR Review of the Final NIH Policy for Data Management and Sharing

Budgeting and Costing

(<https://www.cogr.edu/cogr-readiness-guide-chapter-4-budgeting-and-costing>).

5. How to implement evolving ethical frameworks to DOT-funded research. A percentage of transportation research involves the direct study of human subjects as they interact with the transportation infrastructure and operations. Transportation researchers have a long history of protecting human subjects under academic Institutional Review Board (IRB) and similar ethical guidelines. With the increase in volume of digital data collected about people and populations during research execution, some collectively in identified public settings and some oriented to observation of individuals requiring their knowledge and consent, the global movement towards open science and data sharing has developed new ethical frameworks. One example of these is the “CARE Principles for Indigenous Data Governance” << <https://www.gida-global.org/care> >>, created to allow Indigenous People to assert greater control over the use of Indigenous data and knowledge. DOT seeks information on how to ensure DOT supported research is engaged with and implements these evolving ethical frameworks.

We applaud DOT's proactive consideration of ethical issues when sharing data. We recommend that DOT provide principal investigators (PIs) guidance on writing Informed Consent and participant agreements that align with the data sharing policy, such as the guides released by NIH on Informed Consent for Secondary Research on Data and Biospecimens (<https://osp.od.nih.gov/wp-content/uploads/Informed-Consent-Resource-for-Secondary-Research-with-Data-and-Biospecimens.pdf>). If data from human participants will be shared via DOT's repository, we also strongly encourage providing options to allow for restricted access to data that can not be sufficiently de-identified to allow for public release. Review of data and consent language is advised before accepting data collected from human participants into an open access repository. Even data from studies that do not meet federal requirements for "Human Subjects Research" review by an institution's Institutional Review Board (IRB; for example, data collected for program evaluation or for administrative purposes) may present harm to respondents if distributed more widely than appropriate.

6. How to best improve access to other types of research outputs. The 2015 DOT Public Access Plan focused on making text-based research outputs and digital datasets accessible to the public. But transportation research is not confined to only these two types of outputs. More and more research outputs include software, code, simulations, visualizations, and others yet to come. With the need to update our Public Access Plan, DOT is interested in having supported researchers share all research outputs with the public, where practicable and within legal parameters. DOT seeks information on the projected types of research outputs, the level of effort and expense in sharing them, as well as ethical and legal concerns with sharing other types of research outputs.

All of the research outputs mentioned in this question (including software, code, simulations, etc.) could be viewed by many as digital data—and thus covered under the current public access plan. One option is to encourage PIs to submit these and other source files alongside the data in the main data repository, or closely tie to the data through metadata references to

other repositories designed for code or software (such as GitHub). However, DOT will need to consider their stance on preservation—while software and code rely on dependencies that should be captured, preservation of the computational environment (e.g., through a container such as Docker) may or may not be necessary. Similarly, this may be true with web and dashboard based visualizations—as long as the raw data is shared, others can recreate the product, even if the original presentation of the data is not archived.

Depending on where researchers deposit their data, there may be restrictions on the type of data ingested. Therefore, if researchers are depositing parts of their complete dataset in many separate repositories, there must be a clear link between the various datasets (for example, code could be deposited in GitHub, and the associated spreadsheet could be deposited in another repository as long as the two are linked together via metadata).

As an example, the University of Minnesota Libraries operates the Data Repository for the University of Minnesota (DRUM), which accepts deposits of code and simulations. DRUM ensures that digital materials are findable and downloadable. Each dataset has a README file that documents the who/what/where/why/when of the dataset, including a description of the variables, what publications are associated with it, and if and where there are other related datasets. DRUM also stores associated publications and datasets in relevant Dublin Core standard metadata, which is machine readable and can be passed to DOI level metadata. We recommend this type of documentation be shared with datasets.

Further, careful consideration should be made as to how to bundle datasets, as it is ill-advised to generate DOIs for each file (we have seen cases where datasets have hundreds of DOIs), which makes downloading and tracking of a complete dataset difficult.

Finally, allowing submission of very large files (e.g., exceeding 3GB) may restrict access to users without fast internet access. With our data repository, DRUM, we ask that submissions keep file sizes to 3GB and total datasets to 50GB. Although there are exceptions, file size and dataset size should be a consideration if providing access to the general public is the desired outcome.

7. How to implement persistent identifiers (PIDs) for people; research documents and outputs; and, research entities. The 2015 DOT Public Access Plan called for persistent identification of research outputs and researchers. The 2022 OSTP memo, section 4.b) requires all federally funded researchers to have a personal persistent identifier as defined in NSPM-33 Implementation Guidance << <https://www.whitehouse.gov/wp-content/uploads/2022/01/010422-NSPM-33-Implementation-Guidance.pdf> >> section 2. Further, the OSTP memo section 4.c) requires persistent identification of research and development awards, such as research grants and contracts. Finally, DOT is interested in being able to uniquely and persistently identify research entities, to enable analysis of outputs and research relationships. DOT seeks suggestions on improving the use of persistent identifiers and their metadata, including adoption use cases from institutions.

Persistent identifiers are essential for tracking researchers, publications, and data. ORCID is the primary PID for researchers. DOT, and other federal funding bodies, currently require the use of ORCID iDs as the persistent identifier for people. We recommend continuing this requirement.

We recommend DOT carefully consider how persistent identifiers will be used across research documents and data. As stated above, DOT should provide guidance on how to bundle all of the outputs together to ensure that the associated metadata links all pieces of a project together. There are various persistent identifiers available that serve the same purpose.

For research institutions, we recommend DOT explore requiring the use of Research Organization Registry (ROR) identifiers. ROR is the only PID for research organizations that is open (with a CC0 license). We would prioritize ROR over other identifiers, like Ringgold, because it is openly available. Additionally, it is already integrated into metadata sources like Crossref and DataCite and supported in ORCID.

8. How to improve research project lifecycle management. The 2015 DOT Public Access Plan commits DOT to sharing research project information through a publicly accessible database. DOT seeks suggestions on improving our research project management tools and practices, and welcomes institutional use case examples.

We recommend that within DOT's data management plan requirements there is a section on active research data management (i.e., during data collection). Many data management and sharing plans focus entirely on the preservation and sharing of the final datasets, but gloss over how data is managed during the project. Adding in consideration around active data management (including a plan for storage, backup, version control, etc.) ensures that the data will be well organized and readily accessible for sharing at a project's completion. It helps researchers if funding agencies guide them to existing resources at their institutions or within their disciplines. It is common in academia that the institution's library provides data management training (libraries also manage and curate institutional data repositories). At the University of Minnesota, this group - called Research Data Services (<https://www.lib.umn.edu/services/data>) - provides training to students, staff, faculty, research teams, and labs, broadly as well as customized.