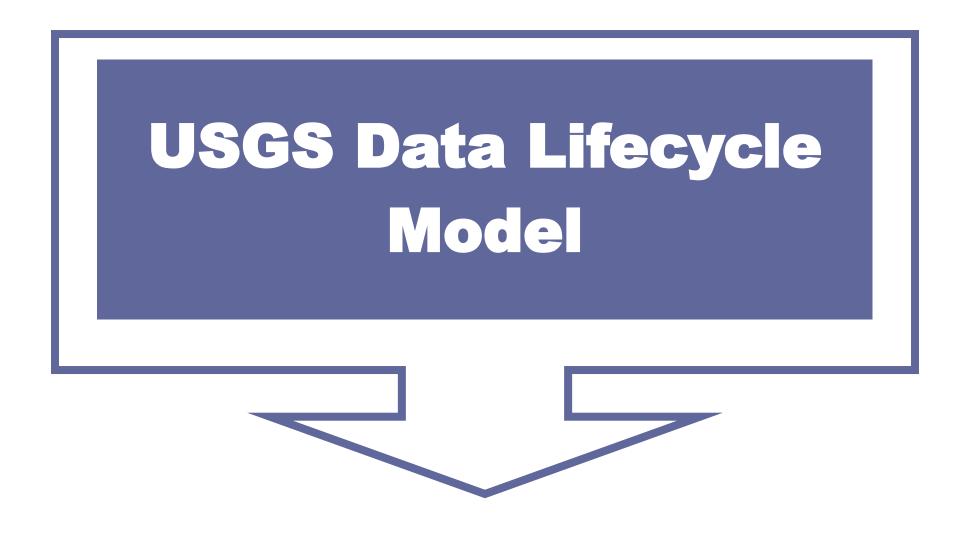


U.S. Department of Transportation Office of the Secretary of Transportation

Abstract

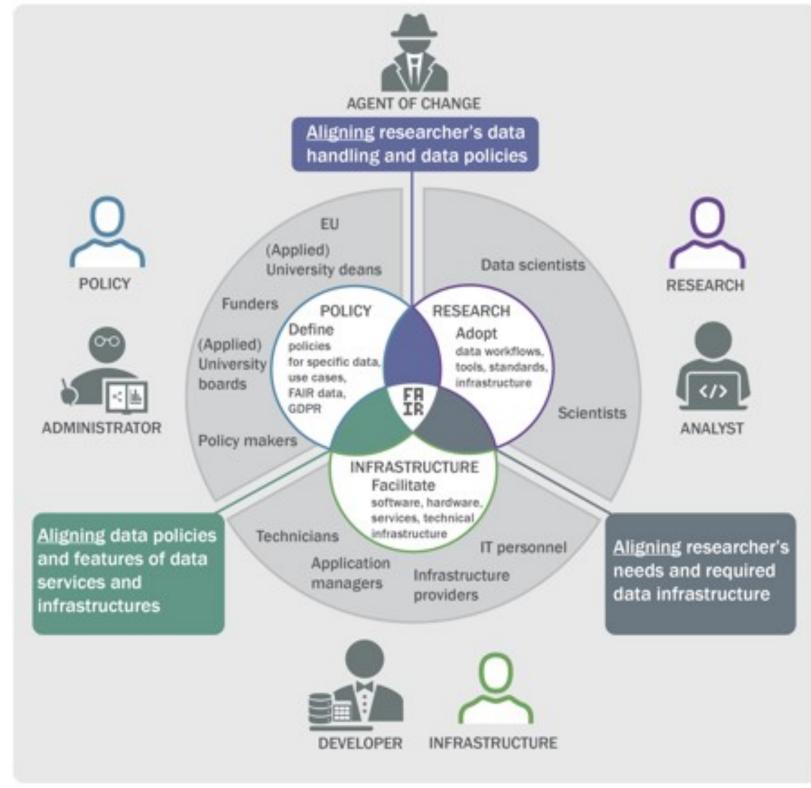
The practice of scientific research in the digital age has led to a significant increase in the amount of data generated and collected, resulting in a data deluge. In response, good practices for the management of research data -- from collection to long-term preservation and reuse -- have emerged both nationally and globally. Implementation of these practices are vital to meet growing demand and provide the skills necessary to care for research data throughout its entire lifecycle. This changing environment has further been enhanced by the global movement of open science. However, the fact that many of these practices are new or only recent additions to transportation research program workflows, have left many research program managers overwhelmed. However, the emergence of the new role of Data Steward is designed to help research managers and close the gaps within the Research Data Management (RDM) Lifecycle.

Data Stewardship --a built-in approach to data management, curation, preservation, sharing, and reuse --offers many practices to improve research coordination, collaboration, and technology transfer. If data is indeed to be considered an asset each research project needs a specialized expert to serve within the research team to ensure that RDM strategies and practices are effectively implemented into the research lifecycle. By operating within the research team, a Data Steward can be solely focused on the project's data management, and a constant resource for the researchers. To maximize its effectiveness, RDM needs to be hands-on and consistent throughout the lifecycle, it doesn't just happen at one stage. Robust Research Data Management starts before a project begins and continues for as long as the data is preserved. Embedding trained Data Stewards into research programs and projects can increase team efficiency and save research resources by lifting the data documentation and preservation burden from researchers, allowing them to focus on data collection and analysis.



There are several definitions of the data lifecycle. To paraphrase many of them, the data lifecycle is "All the phases of data's existence from planning to collection, through preservation, to reuse and potential destruction" Additionally, there are several models of the data lifecycle, below is the USGS Data Lifecycle. This is one of the simpler versions which NTL views as more beneficial, because the very first action is planning.

https://www.usgs.gov/data-management/data-lifecycle



Frederike Schmitz. (2020). The roles of data stewards in the data stewardship landscape

The above image demonstrates how interconnected the role and work of a data steward to all aspects of the research data lifecycle. It is this interconnectedness that makes data stewards the bridge in research data management, because they are integrated into a research project and can ensure that data management best practices are maintain from inception and planning to completion and preservation. A data steward can be seen as a single solution to multiple data management and curation issues that we are currently faced with today when it comes to the proper long-term care of research data. Having a n expert and dedicated individual is both he sustainable and logical solution for research institution who value data as an asset and are committed to treating it as such.



Authors

Jesse Ann Long https://orcid.org/0000-0002-4962-1380 Data Management and Data Curation Fellow National Transportation Library Jesse.long.ctr@dot.gov

Peyton Tvrdy https://orcid.org/0000-0002-9720-4725 Data Management and Data Curation Fellow, National Transportation Library Peyton.Tvrdy.ctr@dot.gov

Leighton L Christiansen https://orcid.org/0000-0002-0543-4268 Data Curator, National Transportation Library leighton.christiansen@dot.gov

Recommended Citation

Long, Jesse A. Christiansen, Leighton L. Tvrdy, Peyton. "Data Stewardship: The Bridge to Optimize Research Data Management (RDM) across the Data Lifecycle." Transportation Research Board 103th Annual Meeting. Washington, D.C., USA. <u>https://doi.org/10.21949/1529902</u>

Data Stewardship has been developed in accordance with many already established nation and global practices around the research data lifecycle, increasing its effectiveness and ability to seamlessly integrate in established workflows to enhance research data.

Outlined, in the chart to the left, are some of these key standards that are highly relevant to NTL, but this is by no means an exhaustive list.



Data Stewardship: The Bridge to Optimize Research Data Management (RDM) across the Data Lifecycle

Defining the Role of Data Stewards

"Data Stewards are [...] making connections and opening channels of communication between researchers, policy makers, software developers, and infrastructure providers (inside or outside the institution) so that the necessary elements that enable researchers to successfully implement RDM can be put in place. The role of data stewards as promoters of FAIR data principles is essential to help researchers and institutions transition to modern RDM practices."

Hasani-Mavriqi, I., Reichmann, S., Gruber, A., Jean-Quartier, C., Schranzhofer, H., & Rey Mazón, M. (2022). Data Stewardship in the Making (1.0), page 9. https://doi.org/10.3217/p9fvw-rke48

Data stewards play a crucial role in managing and maintaining an organization's data assets. Their responsibilities generally revolve around:

1.Data Quality Management

2.Data Governance

3. Data Classification and Metadata Management

4.Data Security and Privacy

5.Data Lifecycle Management

6.Collaboration with Data Users

7.Data Documentation and Communication

8.Data Training and Support

9.Continuous Improvement

Data Stewardship Leads to Proactive Care of Data

Data stewardship ensure proactiveness, thus removing the gaps that come with being reactive on dataset management and curation. Having a data steward integrated within the RDM lifecycle ensures data management strategies are implemented at the time of data creation which leads to through documentation and data preservation for years or decades in the future. Approaching data management, curation and preservation for legacy or already existing datasets, is often harder, and suffers from incomplete knowledge or information due to limited documentation. However, data stewards can prevent this from happening to present and future research.

tions when it comes to data management and data curation.

Reactive

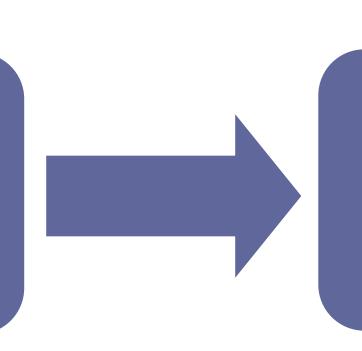
- Curation & Preservation
- Repository Ingest
- Access & Reuse Preservation/Mitigatior
- Format Migration
- Disposition



identified in Denmark and the Netherlands. Zenodo.

https://doi.org/10.5281/zenodo.4321265

Process



Data Stewardship aligns with National and Global Standards

FAIR Principles

Data stewardship plays a pivotal role in realizing the FAIR principles by ensuring that data is well-managed, documented, and accessible. Through meticulous metadata management, standardization of formats, and adherence to data governance, data stewards enhance the findability of datasets, promote interoperability among diverse systems, guarantee accessibility while upholding security measures, and improve data quality, thereby maximizing its potential for reuse across various domains and applications

FAIR Principles: https://doi.org/10.1038/sdata.2016.18

Data stewardship contributes sig Plans (DMPs) by outlining strated managing and maintaining data t

Data Managen

ards ensure that DMPs include of documentation, metadata standa rity measures, and compliance guarantees that DMPs are practi practices, enabling researchers share, and preserve data effectiv ty and longevity of research outp

NTL's Libguide on Research Da transportation.libguides.com/res

How does Data Stewardship fits within the Research Data Management (RDM) Lifecycle ?

The below chart helps highlight the differences between reactive and proactive ac-

Implementing Data Stewardship Strategies in the Research Process

There are many benefits to implementing data stewardship strategies into the research process, some key highlights are shown below:

- Embedding trained Data Stewards into research programs and projects can increase team efficiency and save research resources by lifting the data documentation and preservation burden from researchers, allowing them to focus on data collection and analysis.
- The most likely person to use your data, in the future, is you. So, make it easier on you in the future by implementing data management best practices.
- Having good practices in place will prevent time consuming legacy data rescue in the future throughout the

At the National Transportation Library we have recently assessed our current alignment with data stewardshi practices and where we would like to go in the future:

Proactive

- Creation & Collection Standard Workflows: File
- Naming Data Management & Train-
- ing: DMPs Robust Documentation: Re-
- adme & Codes

Where We Are Now Where We Want to Go • Embedded Data Stewards for each office Partnership with BTS offices, serving as a constant resource for their data manor research project to maximize effectiveness and ensure that RDM is hands-on agement needs. and consistent throughout the lifecycle, it

- Currently the work is more reactive vs. proactive
- doesn't just happen at one stage.
- Robust Research Data Management starts before a project begins and continues for as long as the data is preserved.

Analyze

Preserve

nent Plans (DMPs)	Persistent Identifiers	USDOT Publi
significantly to Data Management ategies and guidelines for effectively ta throughout its lifecycle. Data stew- e comprehensive approaches to data ndards, data quality assurance, secu- e with regulations. Their involvement actical, realistic, and aligned with best rs and organizations to handle, ctively, thereby supporting the integri- utputs.	Data stewards play a crucial role in implementing and maintaining PIDs, ensuring their proper assignment to datasets, enabling persis- tent and unique identification of data resources. By overseeing the governance and management of PIDs, data stewards uphold the reli- ability and longevity of data references, facilitating data discovery, ac- cess, and citation across various platforms for decades into the fu- ture, thereby enhancing the traceability and integrity of datasets.	Data stewardship aligns with the Unite tion's (DOT) Public Access Plan by er- ing and accessibility requirements. Da- mentation of practices that facilitate p data, ensuring adherence to the plan' formats, and metadata. Thus playing ment to transparency and open data p stewards support the DOT's Public Ac- reuse, and promoting broader visibility search data.
Data Management: <u>https://</u> esearchdatamanagement	NTL's LibGuide on Persistent Identifiers: <u>https://</u> <u>transportation.libguides.com/persistent_identifiers</u>	US DOT's Public Access Plan: <u>https://</u>

Transportation Research Board 103rd Annual Meeting Washington, D.C., January 7-11, 2024 **Poster: P24-20524**

DCN's CURATE(D) Steps and Data Packages

To enhance NTL's approach to data stewardship practices and ensure each dataset that is submitted to the digital repository ROSA P receives consistent and documented treatment we have recently adapted the CURATE(D) steps created by the Data Curation Network (https://datacurationnetwork.org/outputs/workflows/). I customized and implemented the CURATE(D) steps per NTL's workflow, policies, systems, and expectations. The benefits of the CU-RATE(D) workflow is long-term knowledge management and thorough documentation of curation actions.

The CURATED Workflow is a standardized set of steps and checklists to ensure all datasets receive consistent and documented treatment.

- C: Check files/code and read documentation (risk mitigation, file inventory, appraisal/selection) U: Understand the data (or try to), if not... (run files/code, QA/QC issues, readmes)
- R: Request missing information or changes (tracking provenance of any changes and why)
- A: Augment metadata for findability (DOIs, metadata standards, discoverability)
- T: Transform file formats for reuse (data preservation, conversion tools, data visualization) E: Evaluate for FAIRness (transparent usage licenses, responsibility standards, metrics for tracking use)
- D: Document all curation activities throughout the process

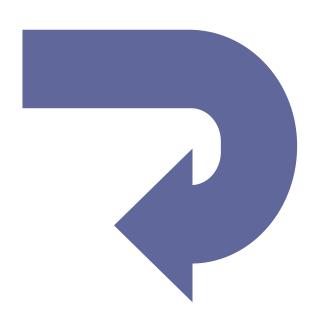
The result of successful data stewardship across the research data management lifecycle is a robust data package. These packages serve as organized containers, containing detailed information about the dataset's origin, structure, and context, fostering improved data discoverability, understanding, and usability By employing data packages, data stewards ensure better documentation, quality control, and compliance with standards, streamlining data management practices throughout its lifecycle. The structured nature of data packages facilitates easier sharing, reproducibility, and collaboration while reinforcing data stewardship's role in maintaining data integrity, accessibility, and longterm usability. To the left you can see an image of the elements that NTL requires or suggests in order to have a robust data package.

NTL Dataset Data Package Elements

- \Rightarrow .csv or other open format) Readme.txt
- \Rightarrow Includes Data Dictionary ⇒Notes standards used ⇒Defining Zero, Null, and Unknown
- \Rightarrow FAQs and other notes 3) Metadata file in Project Open
- 4) Data Management Plan (DMP) 5) Code or scripts used in data analysis 6) Supporting files, tables, etc.

(Bold = Required; Italics = Optional, or Required if

Publish/Share



ic Access Plan

nited States Department of Transportaensuring compliance with its data shar-Data stewards would oversee the implepublic access to DOT-funded research an's guidelines regarding data availability ng a vital role in enabling DOT's commita practices. Through their efforts, data Access Plan objectives, fostering data ility and utility of transportation-related re-

s://doi.org/10.21949/1503647

Conclusion

In conclusion, the surge in data generated within the realm of scientific research has necessitated robust and strategic data management practices. This paradigm shift, catalyzed by the digital age and the advent of open science, has introduced new challenges and opportunities. The integration of Data Stewards into research teams is instrumental in ensuring the successful implementation RDM strategies across the entirety of a project's lifecycle. By shouldering the responsibilities of data managemen and preservation, Data Stewards allow researchers to focus on core aspects of their work, such as data collection and analysis. This embedded approach to RDM not only enhances team efficiency but also ensures the longevity and accessibility of research data, positioning it as a valuable asset for current and future scientific endeavors.



Scan the QR code above to download and save this poster for future reference.