

Data Curation & Transparent Federal Statistics

Some Suggestions

Leighton L Christiansen

 <http://orcid.org/0000-0002-0543-4268>

Data Curator, National Transportation Library,
Bureau of Transportation Statistics,
OST-R , US Department of Transportation
leighton.christiansen@dot.gov
ntldatacurator@dot.gov

Jesse Long

 <https://orcid.org/0000-0002-4962-1380>

Data Curation & Data Management Fellow,
National Transportation Library,
Bureau of Transportation Statistics,
OST-R , US Department of Transportation
jesse.long.ctr@dot.gov

Overview

About BTS & NTL

About Data Curation

NCSES Charge Review

Data Curation for Transparent Statistics: Suggestions

Conclusions

Questions

About BTS

Founded in 1991

Preeminent source of statistics, and statistical datasets, on:

Commercial Aviation,

Multimodal Freight Activity, and,

Transportation Economics,

Provides context to decision makers and the public for understanding transportation statistics

BTS Director is, by law, the senior advisor to the Secretary of Transportation on data and statistics

<https://www.bts.gov/>

About NTL

NTL is an **open access** digital repository of transportation information

All collection materials are in the **public domain**, available for reuse **without restriction**

NTL is one of five national libraries

NTL is the only national library within a Principal Federal Statistical Agency

NTL **provides access** to:

Digital collections

Data services

Reference services

Knowledge networking

NTL's Guiding Mandates

**Transportation
Equity Act for
the 21st Century
(TEA-21) 1998**

Established NTL
to provide national
and international
access to
transportation
information

**Moving Ahead
for Progress in
the 21st Century
(MAP-21) 2012**

Expanded NTL
role as a central
clearinghouse for
transportation
research
publications and
data

**US DOT Public
Access Plan
2016**

Requires NTL
host repository for
research and
datasets; **provide**
searchable DMP
collection, and,
assign persistent
identifiers

**Foundations for
Evidence-Based
Policymaking
Act 2018**

Codifies efforts to
ensure public
access to
federally-funded
research reports
and datasets

About Us

Leighton:

MLIS, CAS Data Curation (UIUC) 2012

Library Director and Data Governance
Committee (Iowa DOT) 2012 – 2016

NTL Data Curator, May 2016

Public Access Implementation

BTS Data Curation

Data.gov listings for BTS

Jesse:

MLIS, 2019

NTL Data Management and Data Curation
Fellow, June 2019

Preservation of Legacy BTS data

Lifecycle data management for airline data

About Data Curation: Reactive Actions

Reactive

Curation & Preservation

Repository Ingest

Access & Reuse

Preservation/Mitigation

Format Migration

Disposition

About Data Curation: Proactive Actions

Reactive

Curation & Preservation

Repository Ingest

Access & Reuse

Preservation/Mitigation

Format Migration

Disposition

Proactive

Creation & Collection

Standard Workflows: *File Naming*

Data Management & Training: *DMPs*

Robust Documentation: *Readme & Codes*

Controlled Vocabularies: *Data Dictionaries*

Metadata Standards: *Choose & Publicize*

Persistent Identification: *DOI, ORCID, ROR*

Preservation Planning: *Repository & Backups*

Benefits of Data Curation

Protects Unique Data from
Loss

Improves Data Search &
Retrieval

Enables Reuse

Facilitates Longitudinal
and/or Meta Analyses

Avoids Duplication of Effort
& Spending

Increases Verifiability

Opens New Lines of
Scientific Discovery

Satisfies Public Access &
Open Government & Legal
Requirements

Data Curation: Definitions

Data Management:

deliberate planning, creation, storage, access and preservation of data produced from a given investigation^{1, 2}

Data Curation

enables data discovery and retrieval, maintains data quality, adds value, and provides for re-use over time³

Data Science

drawing useful conclusions from large and diverse data sets through exploration, prediction, and inference⁴

Linked Processes

DM is a
**Necessary
Element** of DC

DC **Enables** DS

Data Management \in Data Curation

Data Curation \Rightarrow Data Science

Data Curation Dependencies Model

Data Management \in Data Curation \Rightarrow Data Science

DM \in DC \Rightarrow DS

Data Curation & the Data Lifecycle

Data Curation

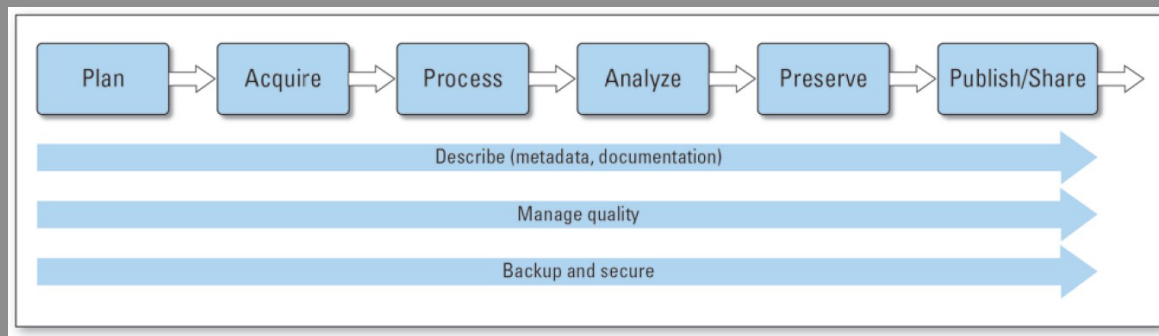
Enables data discovery and retrieval, maintains data quality, adds value, and provides for re-use over time³

Data Lifecycle

All the phase of data's existence from planning to collection, through preservation, to reuse and potential destruction

USGS Data Lifecycle Model

Plan FIRST!!
Collect second
Curation steps
throughout



NCSES Charge Review

***From
Emilda B. Rivers
May 21, 2019***

1. Best practices to foster transparency and reproducibility
2. Guidance, standards, and tools for documenting and archiving
3. Approaches to minimize cost
4. Feasible implementation steps – low hanging fruit

Data Curation for Transparent Statistics: Three Main Suggestions

**Data
Management
&
Sharing
Plans**

**Plan for
FAIR & to
Share**

**Embed
Data
Curators &
Curation
Practices**

Suggestion 1: Data Management [& Sharing] Plans

Explicit documentation of
knowledge

Sets project standards

Plan for data capture

Links to policies

Living document: review
and update

Potential DMP Sections

Project Title and Information

Data Description

Roles & Responsibilities

Standards Used

Access Policies

Sensitive Data Policies

Sharing Policies

Archiving and Preservation Plans

Applicable laws and policies

Suggestion 2: Plan for **FAIR**₇ and to Share

Findable
Accessible
Interoperable
Reusable

Sharing Data

Last step of USGS Data Lifecycle: Publish/Share
Sharing: Culture Change that affects decisions
Encourages new discovery & efficiencies
Consistent with developing U.S. policy and law

<https://www.force11.org/group/fairgroup/fairprinciples>

Suggestion 3: Embed Data Curators & Curation Practices

Necessary skills other team members may not possess

Fresh eyes for workflows and implicit knowledge

Assume preservation and sharing

Improve team efficiency around sharing and preservation

Lifecycle view of data

End of lifecycle planning

NCSES Charge Challenge

JISC Report:
FAIR in Practice⁸

Tools are needed,
remain elusive

While there is “[s]trong support for growing the body of tools and resources available that reduced the burden of data management,” there is also a “[l]ack of good tooling to support metadata capture at data generation.”



Conclusions & Suggestions Review

- Data curation enables data science
- Data Curation lifecycle view defaults to transparency
- Data management and sharing planning is ***THE*** first step
- FAIR data principles apply to metadata, data, and paradata
- Plan for sharing; create a sharing culture
- Embed data curators and curation practices into projects from the start for best results and most transparent statistics

References

1. University Library, Texas A&M University. “Data Management Defined - Research Data Management - Guides at Texas A&M University.” Research Data Management, October 1, 2013. <http://guides.library.tamu.edu/DataManagement>
2. Briney, Kristin. 2015. Data management for researchers: organize, maintain and share your data for research success. <http://www.pelagicpublishing.com/data-management-for-researchers.html>
3. Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. “Specialization in Data Curation,” 2013. http://www.lis.illinois.edu/academics/programs/specializations/data_curation
4. Definition based on Ani Adhikari and John DeNero, “The Foundations of Data Science” <http://www.inferentialthinking.com/index.html> “What is Data Science” <http://www.inferentialthinking.com/chapter1/what-is-data-science.html>
5. Digital Curation Centre. Data Curation Lifecycle Model. <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
6. Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013, The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., <http://dx.doi.org/10.3133/ofr20131265>
7. FORCE11. “The FAIR Data Principles.” 2016. <https://www.force11.org/group/fairgroup/fairprinciples>
8. Allen, Robert, & Hartland, David. (2018, May 21). FAIR in practice - Jisc report on the Findable Accessible Interoperable and Reuseable Data Principles (Version 1). Zenodo. <http://doi.org/10.5281/zenodo.1245568>

Thank you!

Questions?

Leighton L Christiansen

 <http://orcid.org/0000-0002-0543-4268>
Data Curator, National Transportation Library,
Bureau of Transportation Statistics,
OST-R , US Department of Transportation
leighton.christiansen@dot.gov
ntldatacurator@dot.gov

Jesse Long

 <https://orcid.org/0000-0002-4962-1380>
Data Curation & Data Management Fellow,
National Transportation Library,
Bureau of Transportation Statistics,
OST-R , US Department of Transportation
jesse.long.ctr@dot.gov