

Trans/Hack

Opening and Preserving Transportation Data Hackathon

Leighton Christiansen, Iowa DOT
Kenda K. Levine, UC Berkeley
Mary Moulton, National Transportation Library
Amanda J. Wilson, National Transportation Library



Welcome

We want to thank our sponsor:



National Transportation Library



Agenda

Section 1: Introduction

- Opening and Preserving Transportation Data (45 min)
- Break (30 min)

Section 2: Hackathon

- Small group/Large group procedural hacking (90 min)
- Conclusions (20-30 min)



Who We Are

Leighton Christiansen, Iowa DOT

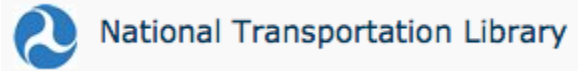
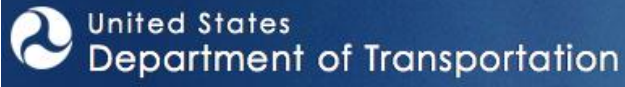
Kenda K. Levine, UC Berkeley

Mary Moulton, National Transportation Library

Amanda J. Wilson, National Transportation Library



Who Are We?



TRB Committee (B0002)



TRANS
DM



all images used under claim of educational fair use

#TransHack

http://libraryconnectivity.org/datamgt/index.php/Main_Page



Navigation

- [Main page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

Toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Page information](#)

http://libraryconnectivity.org/datamgt/index.php/Main_Page

Log in

Search

Main Page

Greetings and welcome to the Transportation Data Management wiki. This wiki is intended to help transportation agencies, researchers, information professionals, and librarians understand data management rules and issues presented by the 2013 memo from the Office of Science and Technology Policy.

The Transportation Data Management wiki, or Trans_DM wiki, is edited by a volunteer group of transportation librarians and information professionals. If you have questions, or would like to become a contributor to this wiki, please contact our corresponding editor, Leighton Christiansen [leighton.christiansen\[at\]dot.iowa.gov](mailto:leighton.christiansen[at]dot.iowa.gov). For more information about Trans_DM wiki, please see the page [Data Management:About](#).

Please click on a link below to be taken to pages of interest.

[OSTP memo "Increasing Access to the Results of Federally Funded Scientific Research"](#)

[Glossary of Terms](#)

[List of Data Management resources](#)

[List of Data Management events and webinars](#)

[Handout](#)

[IDCC Workshop Brainstorming](#)

[State Laws Relating to Open Data](#)

[Data Jokes](#)

[Videos about Data Management](#)

Follow Transportation Data on Twitter: [@Trans_DM](#)



What is Transportation?

Transportation Research Thesaurus

A coordinated system made up of multimodal services serving a common purpose, the movement of people and goods.

(Source: AASHTO Glossary) <http://trt.trb.org>



Aviation



Highway



Marine
Transportation



Motor
Carriers



Pedestrians
and
Bicyclists



Pipelines



Public
Transportation



Rail

all images from [http://www.trb.org/](http://www.trb.org) used under claim of educational fair use

#TransHack



What is Transportation Data?

Data Types -

- Traffic (volume, flow, speed)
- Vehicle (emissions, fuel consumption, sales)
- Logistics (routes, commodities)
- People (census, choice, modes)
- Land-Use



What is Transportation Data?

Data Types -

- Sensor data (large databases, exported to CSV)
- Video and image data
- Survey data (CSV)
- Hand counts
- GIS



What is Transportation Data?

Research Projects

- [The SHRP2 Naturalistic Driving Study.](#) Thousands of hours of video and sensor data.
- [PeMS from Caltrans.](#) Traffic data for CA state highways from sensors.



What is Transportation Data?

Research Projects

- [Bike Count Data Clearinghouse](#) from UCLA Luskin Center.
- [Capital Bikeshare Data](#) includes realtime system data and surveys.
- [GTFS](#) transit data for routes and schedules



Big Bytes

Kilobyte	1,000
Megabyt	1,000,000
e	1,000,000,000
Gigabyt	1,000,000,000,000
e	1,000,000,000,000,000
Terabyt	1,000,000,000,000,000,000
e	1,000,000,000,000,000,000,000



White House OSTP Memo

- February 22, 2013 – Office of Science & Technology Policy issues memorandum entitled “Increasing Access to the Results of Federally Funded Scientific Research.”
- Memorandum addresses new requirements for both intramural and extramural publications and digital data sets resulting from federally-funded scientific research.
- Expand on Data.gov, Open Government requirements already in place
- RITA assigned responsibility for preparing response.



USDOT Response

Publications

- Most publications arising from DOT-funded research are already made available to the public (including intramural reports); principal DOT shortfall is availability of peer-reviewed, “scholarly” publications (i.e., journal articles)
 - Per MAP-21, NTL is now the federal repository for all Department research information and a clearinghouse for all Government transportation research information
 - Plan requires submission of accepted, final manuscript to NTL
- §Internal reports will be available through, and preferably stored by, NTL
- §To protect publishers, all manuscripts will be embargoed for a period not less than 12 months (prefer 18 months)
- §Will require minimal upgrade of NTL software – open source software is preferred meaning minimal additional costs
- §All publications, data sets and authors will have unique permanent identifiers for correlation of articles with authors and relevant underlying data



USDOT Response

Data

- DOT makes some, but not all data available
- excludes that which has confidentiality, privacy, proprietary, IP, security, and other exemptions and protections
- Intramural
 - DOT's Data Release Policy is the governing document for research data management
 - Management will remain with each OA
 - Will require minor updates to OA capabilities, including permanent identifiers
- Extramural
 - Will require submission and DOT approval of data management plan from all awardees
 - Awardees will determine repository for depositing data; MUST be accessible by NTL
 - Data will be included in DOT Enterprise Data Inventory



USDOT Response

Research Project Records

- Goal: to enable linkage of publications and underlying data via the NTL, to enable public access through a simple unique identifier
- DOT will require the submission of project records by DOT funding recipients (intramural and extramural) that fully describe project activities
- Project records will be required to be submitted to the Transportation Research Board's Research in Progress (RiP) database
- Summary descriptions of any documented project outputs and outcomes resulting from research implementation will be required to be submitted to the USDOT Research Hub
- Project records will be required to be updated over the duration of the project
- DOT will develop step-by-step guidance for the submission process



Data Available from USDOT

NTL Data Catalog <http://ntlsearch.bts.gov/repository/ntlc/btsdd/index.shtm>

- Statistical data sets
- Geospatial data
- Sensor data
- Administrative (e.g. bridge inventories)
- Naturalistic study data
- Simulations and models



US DOT Data Challenge

Data Innovation www.transportation.gov/datachallenge

- Safety: address concerns
- Transportation access: how transportation connects people to jobs, school, housing, community resources
- Traffic management and congestion: understand and reduce traffic and congestion



Roadblocks

Administrative Issues

- Many funding sources = many headaches
- Many funding sources = many terms of deliverables
- Lack of coordination for tracking compliance



Roadblocks

Legal Issues

- Privacy and NDA issues
 - Human subject testing
 - Industry secrets
- Rights and ownership
- Ability to license, re-use, and buy/sell data?



Roadblocks

Open Issues

- Interoperability
- Data formats
- Metadata schema
- Data sets are scattered by funder (if at all)
- When is “final” dataset made available?



Roadblocks

Funding Issues

- Unfunded Mandates
- Long-term vs.Short-term funding
- Recharge model and pricing
- Cost of infrastructure and overhead



Break

30 minutes or so



HACKATHON!!!!!!!!!!!!!!!!!!!!!!

- 15:30 – Introductions
- 15:40 Trans Data Hackathon – small group work
- 15:40 -- Groups report back
- 16:10 Trans Data Hackathon -- Small group work/shuffle groups?
- 16:10 – Groups report back
- 16:25 Conclusions and thank you
- 16:25 –
- 16:55

