Big Data – Big Challenges

Catherine T. Lawson

University at Albany

APDU Annual Conference: A Sea Change for Public Data -- September 16 - 17, 2013

So much data – so little organization

- Where did this data come from?

- Who created this data?

- Under what conditions?

- Who cleaned this data?

- Who processed this data?

- How does this data relate to other data I have?

- **Why can't this data tell me about itself?**

The White House

Office of the Press Secretary

For Immediate Release
May 09, 2013

Executive Order -- Making Open and Machine Readable the New Default for Government Information

EXECUTIVE ORDER

- - - - - - -

MAKING OPEN AND MACHINE READABLE THE NEW DEFAULT
FOR GOVERNMENT INFORMATION

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. General Principles. Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation.

.........Sec. 2. Open Data Policy. ....(b) Agencies shall implement the requirements of the Open Data Policy and shall adhere to the deadlines for specific actions specified therein. When implementing the Open Data Policy, agencies shall incorporate a full analysis of privacy, confidentiality, and security risks into each stage of the information lifecycle to identify information that should not be released. These review processes should be overseen by the senior agency official for privacy. It is vital that agencies not release information if doing so would violate any law or policy, or jeopardize privacy, confidentiality, or national security......

BARACK OBAMA

APDU Annual Conference: A Sea Change for Public Data -- September 16 - 17, 2013

The Layers of Digital Services

"Customers"
American People   Employees

Govt Digital Services Private Digital Services	Presentation Layer
Systems, Processes, Management & Web APIs	Platform Layer
Open Data & Content (Information)	Information Layer

Transportation Data in the Age of BIG/OPEN DATA

- Operations Data
  - Archived Data Users Service (ADUS)
  - Travel Monitoring Analysis System (TMAS and TMAS 2.0)
  - General Transit Feed Specifications (GTFS and GTFS-R)
- Cloud-based Technologies
  - Using amassed data

ADUS – The ITS Add-On

- National ITS architecture
- "Show me the data" – no capacity to archive operations data
- FHWA/TRB/academia/private sector workshops
- Five-Year-Program Plan
- Only 3 variables
  - Latitude
  - Longitude
  - Time Stamp
- Issues of quality, completeness, and meta-data

TMAS 2.0

- Traffic Monitoring Guide (TMG) – April 2013 update
  - Standardization of data elements
  - Volume, Classification Counts, Weigh-in-Motion, and more
- TMAS Quality Checks
  - Rigorous approach to data stewardship
  - Systematic approach to cleaning and quality checking the data
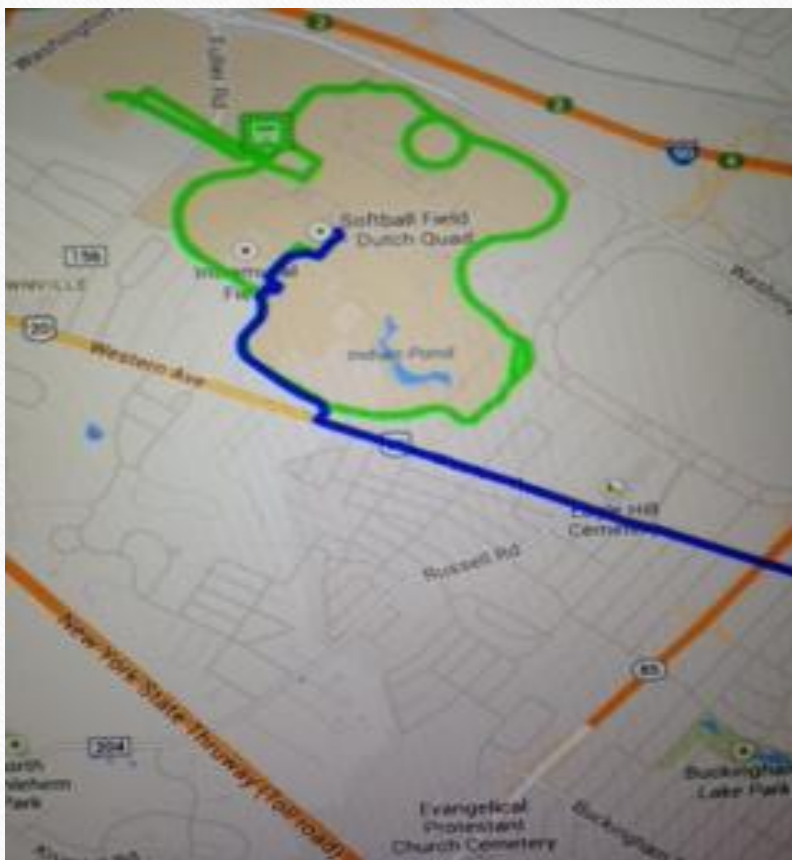
GTFS & GTFS-R

- Transit data community effort
- Open Data/Open Source strategies for transit scheduling
- Designed data structure to minimize complexity and maximize functionality
  - Meta-data
  - "Published" to the web
  - Meeting the needs of the traveling public
- Archive-able

APDU Annual Conference: A Sea Change for Public Data -- September 16 - 17, 2013

Cloud-based Technologies

- Application Processing Interface (API)
  - Census

- Moving transportation forward with "amassed" data
  - "Publishing" TMAS
  - Integrating GTFS & GTFS-R into planning
  - Integrating weather data with transportation data

- Appending lifecycle meta-data to each record – now that's REALLY BIG DATA!

Upload GTFS File | Sign In
All Agencies | Recent Updates
New Agencies
CTS (8 hours ago ago), MiWay (1 week ago ago), TransBunbury (1 week ago ago), Non-public entity (2 weeks ago ago), University of Wisconsin (2 weeks ago ago) (**View All 689 Agencies**)
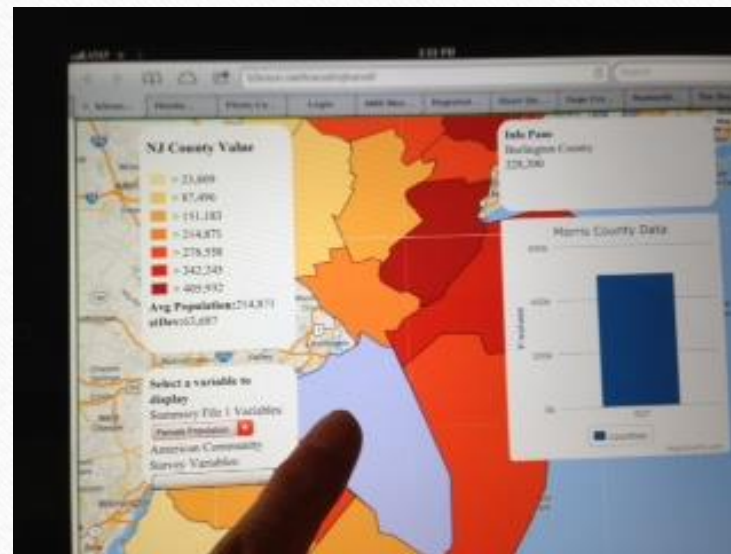
Recent Data Updates
 RSS of All Updates
Upload By meroje.f 8 hours ago ago   pertains to CTS
GTFS File
cts_20130915_1535.zip 4784692 more info »
Upload By wmata-archiver Yesterday ago   pertains to DC Circulator MET

Census
API
requires
computer
science
talent

Conclusion……….. We are just at the beginning

Govt Digital Services	Presentation	GTFS & GTFS-R
Private Digital Services	Layer	
Systems, Processes,		
Management &	Platform	CENSUS API
Web APIs	Layer	
Open Data &		
Content	Information	TMAS
(Information)	Layer	ADUS