

Keeping Maps Safe and Accessible: Digitization and Preservation at the Library of Congress



September 27, 2017



Agenda

- **Jacqueline Nolan**, *Cartographer*
- **Colleen Cahill**, *Digital Conversion Coordinator*
 - Geography and Map Division Scan Lab
 - Physical item preservation and storage
 - Digital data storage and file arrangement
 - Metadata
- **Tim St. Onge**, *Cartographer*
 - Digital data strategies of the Geography and Map Division
 - Challenges in geospatial data preservation
 - The “Geospatial Hosting Environment” project



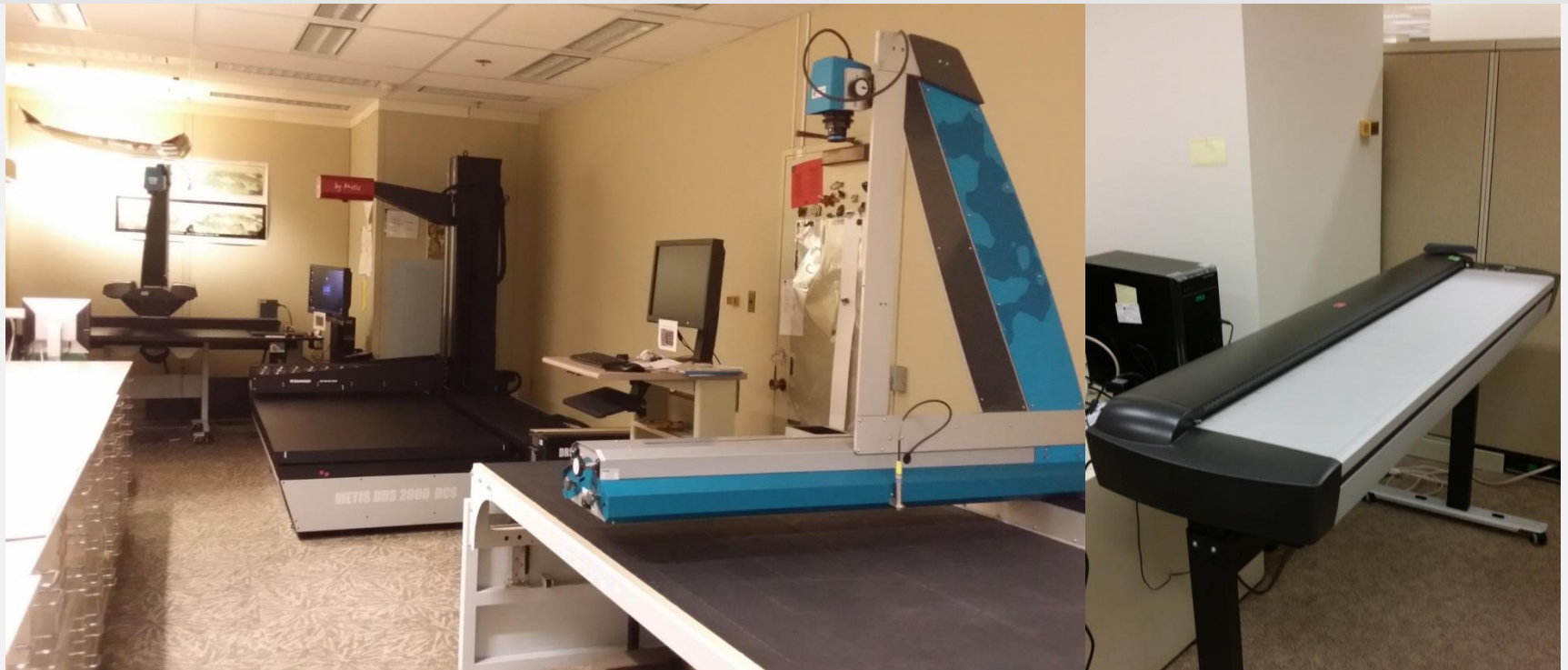
It Began in 1995...



- Scan for web presentation, not preservation
- Conservation review before scanning
- Digitization means less handling of maps
- Some items would not be scanned



Moving Forward to Today



- New scanners for various materials
- Have over 70,000 online maps, atlases, & globes
- Preservation is an acknowledged factor in scanning



Physical Preservation



- Conservation staff review materials
 - Provide stabilization and supports
 - Train all scanning staff in materials handling
- Scan in Mylar where needed
- Rehousing of items before or after scanning



Physical Storage - Maps



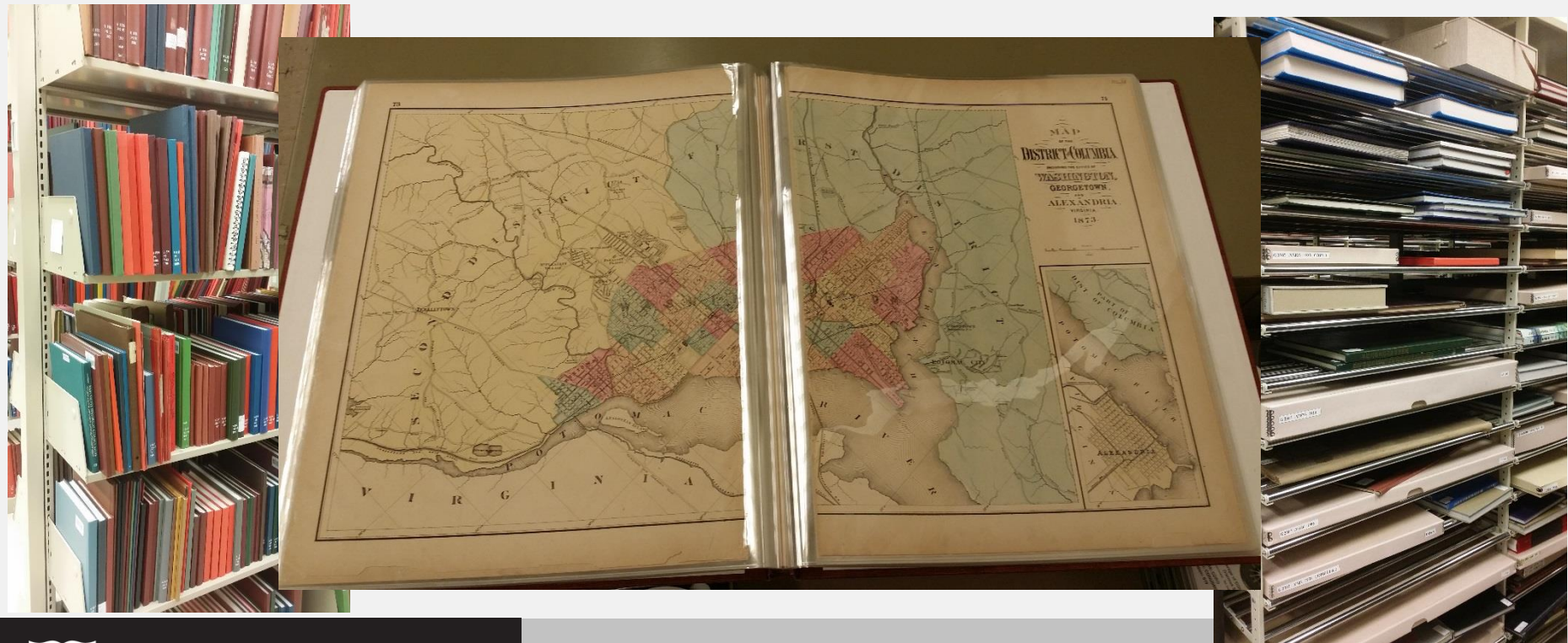
- Maps are stored in map cases
 - Each single map is in an individual, labeled folder
 - Acid-free folders are used and rehoused as needed
 - Sheets are placed in Mylar, resources are limited
 - Set maps may be in folders, but not individually, as room and weight are issues



Physical Storage - Atlases



- Atlases are stored on shelves
 - Smaller atlases are stored upright
 - Folio atlases are stored lying flat
 - Some atlases are rebound in Mylar



Physical Storage - Globes



- Globes are stored in boxes where possible
 - Keeps light out





- Content Transfer System
 - Three storage areas:
 - Long term tape storage
 - Master
 - Service
 - Checksums of all files
 - Reports on:
 - Number of files
 - Space usage
 - Bagit - <https://www.loc.gov/preservation/digital/>
 -

Digital Data - Digitization



- Digitized Library Collections
 - TIFF format, 24 bit, RGB, uncompressed
 - Keep both the unprocessed TIFF and processed TIFF
 - Processed TIFF
 - Rotated
 - Cropped
 - LAB lightness channel adjusted for web optimization
 - Unprocessed TIFF is useful for two reasons
 - Printing – Processed TIFF is lightened in sometimes to large to print
 - Rebuild Processed TIFF if any issues arise
 - Processed TIFFs are master files
 - JPEG2000 for service files



Digital Data – Born Digital



- Born digital from various sources
 - Copyright deposit
 - Government deposit
 - Purchase
 - Donation
 - Download
- Many formats
 - Raster
 - Vector
 - Databases
 - Text documents



File arrangement



- Digitized items arranged in LC Classification system
 - Utilize part of the G Schedule
 - Arranged digital images in similar way to their analog counterparts

[g3824a](#)
[g3824b](#)
[g3824c](#)
[g3824d](#)
[g3824e](#)
[g3824f](#)
[g3824g](#)
[g3824h](#)
[g3824i](#)
[g3824j](#)
[g3824k](#)
[g3824l](#)
[g3824m](#)
[g3824n](#)
[g3824o](#)
[g3824p](#)
[g3824r](#)

[cw0358500Z.tif](#)
[cw0358500a1X.tif](#)
[cw0358500a2X.tif](#)
[cw0358500a3X.tif](#)
[cw0358500a4X.tif](#)
[la000783.tif](#)
[la000783X.tif](#)
[la002074.tif](#)
[la002074X.tif](#)
[la002122.tif](#)
[la002122X.tif](#)
[la002168.tif](#)
[la002168X.tif](#)
[pm008020.tif](#)
[pm008240.tif](#)
[pm008250.tif](#)
[pm008260.tif](#)
[pm008261.tif](#)



“Bird’s eye view of Philadelphia” by Theodore R. Davis and Harper & Brothers, 1872. Geography and Map Division, Library of Congress. [LCCN: 75694509](#)

File Arrangement – Born Digital



- Arranged by the Library of Congress Catalog Number (LCCN)
 - Folders for files are named by the LCCN
 - Provides user with a way to find the bibliographic metadata quickly
 - Notes and tags in the Content Transfer system also provide some information
 - These items are put in Bagit

2014588476
2014594164
2014594165
2014594172
2015593313
2015593314
2015593315
2015593316
2015593317



Types of Metadata



- Bibliographic metadata
 - Library of Congress uses MARC format with RDA rules
- Administrative metadata
 - Digitization
 - Who scanned it
 - When was it scanned
 - What scanner
 - Why was it scanned
 - Copyright or donor issues
 - Born digital
 - Acquisition source
 - Date acquired
 - Any restrictions



More Metadata



- Tracking metadata
 - Where is it in the processing workflow
 - Is it ready for review
 - Is it live online
 - Notes
- Preservation metadata
 - PREMIS - <https://www.loc.gov/standards/premis/>



Agenda

- **Jacqueline Nolan**, *Cartographer*
- **Colleen Cahill**, *Digital Conversion Coordinator*
 - Geography and Map Division Scan Lab
 - Physical item preservation and storage
 - Digital data storage and file arrangement
 - Metadata
- **Tim St. Onge**, *Cartographer*
 - Digital data strategies of the Geography and Map Division
 - Challenges in Geospatial Data Preservation
 - The Geospatial Hosting Environment



Born-Digital Geospatial Data Sources



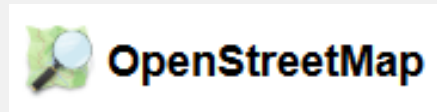
- Same wide variety of data sources as analog
 - Federal/State/Local governments
 - Commercial vendors/copyrighted data
 - Foreign governments, international sources
 - Nonprofit and non-governmental organizations



Armed Conflict Location & Event Data Project



Ordnance Survey



New Collection Policy Statement: Digital Geospatial Materials



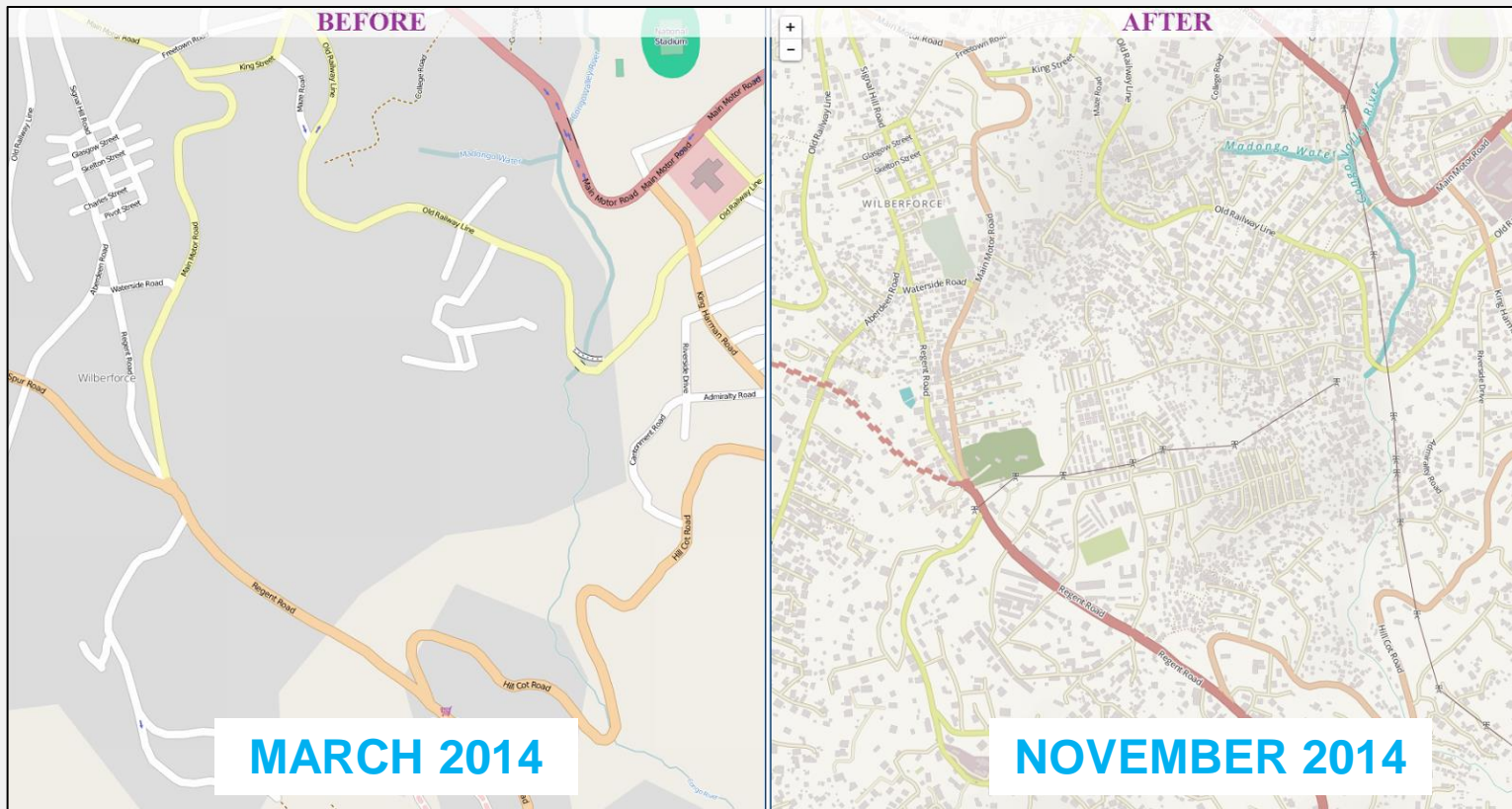
- “Analog” and digital data used to be covered in one CPS
- New CPS written in 2016 addresses unique opportunities and challenges of digital data
 - <https://www.loc.gov/acq/devpol/geospatial.pdf>
- Objectives
 - Briefly define digital geospatial data types
 - Define limiting criteria (storage space, licensing, etc.)
 - Broadly define scope of digital acquisitions



Challenges in Geospatial Data Preservation



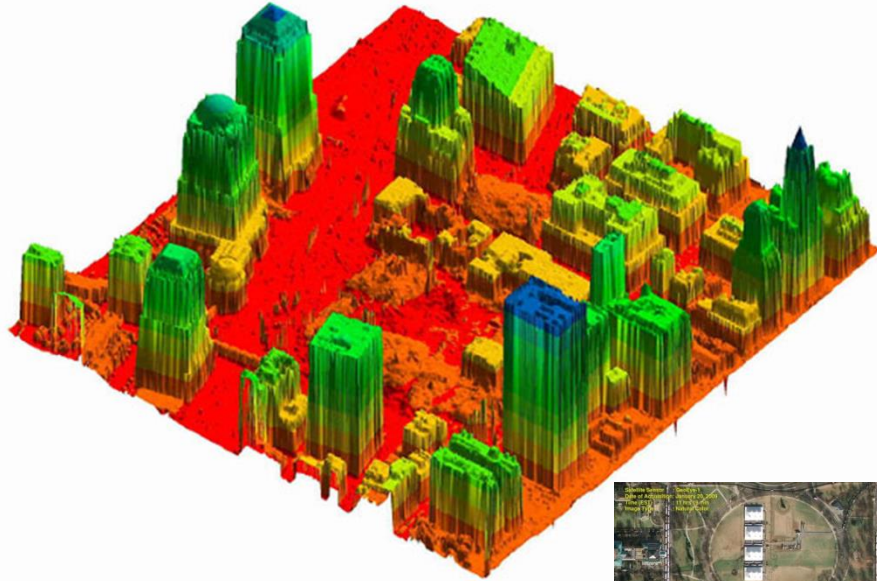
- How do we preserve geospatial data forever?
- How do we curate when the scale of data production is so massive?



Open Street Map of Freetown, Sierra Leone, before and after outbreak of Ebola in March 2014.



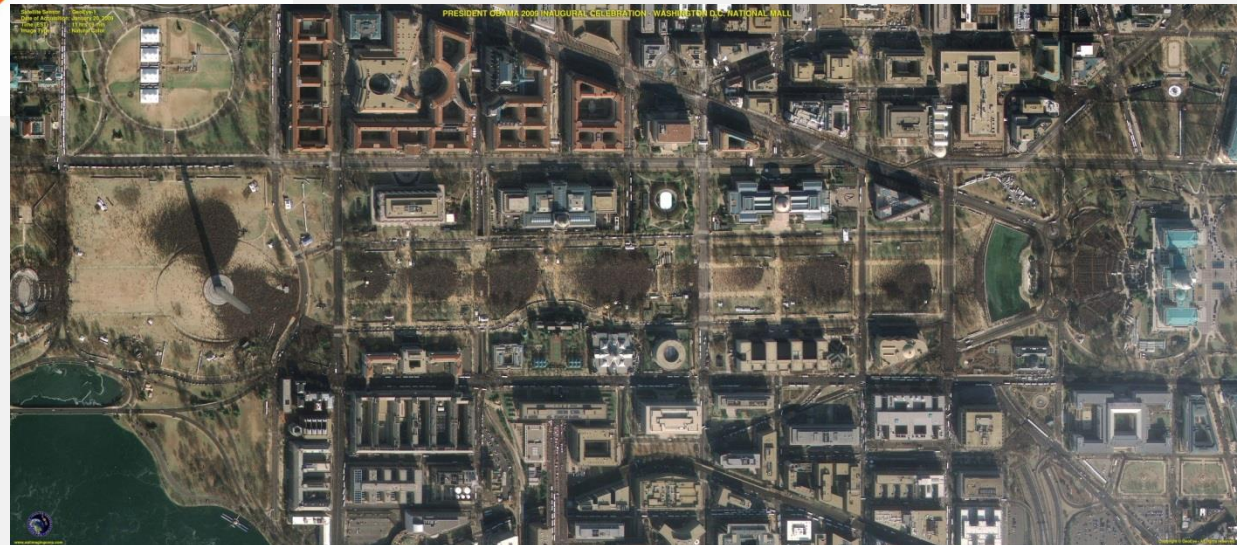
Challenges in Geospatial Data Preservation



WTC – LIDAR Three Dimensional Model, September 15, 2001. New York State, Office for Technology (c2001) and EarthData International. Geography and Map Division, Library of Congress.

Preservation for **spatial analysis** as well as **historical record**

President Obama Inauguration - GeoEye-1 Satellite Image. Geoeye-1, Jan. 20, 2009, Natural Color. Satellite Imaging Corporation.



Challenges in Geospatial Data Preservation



October 29, 2012

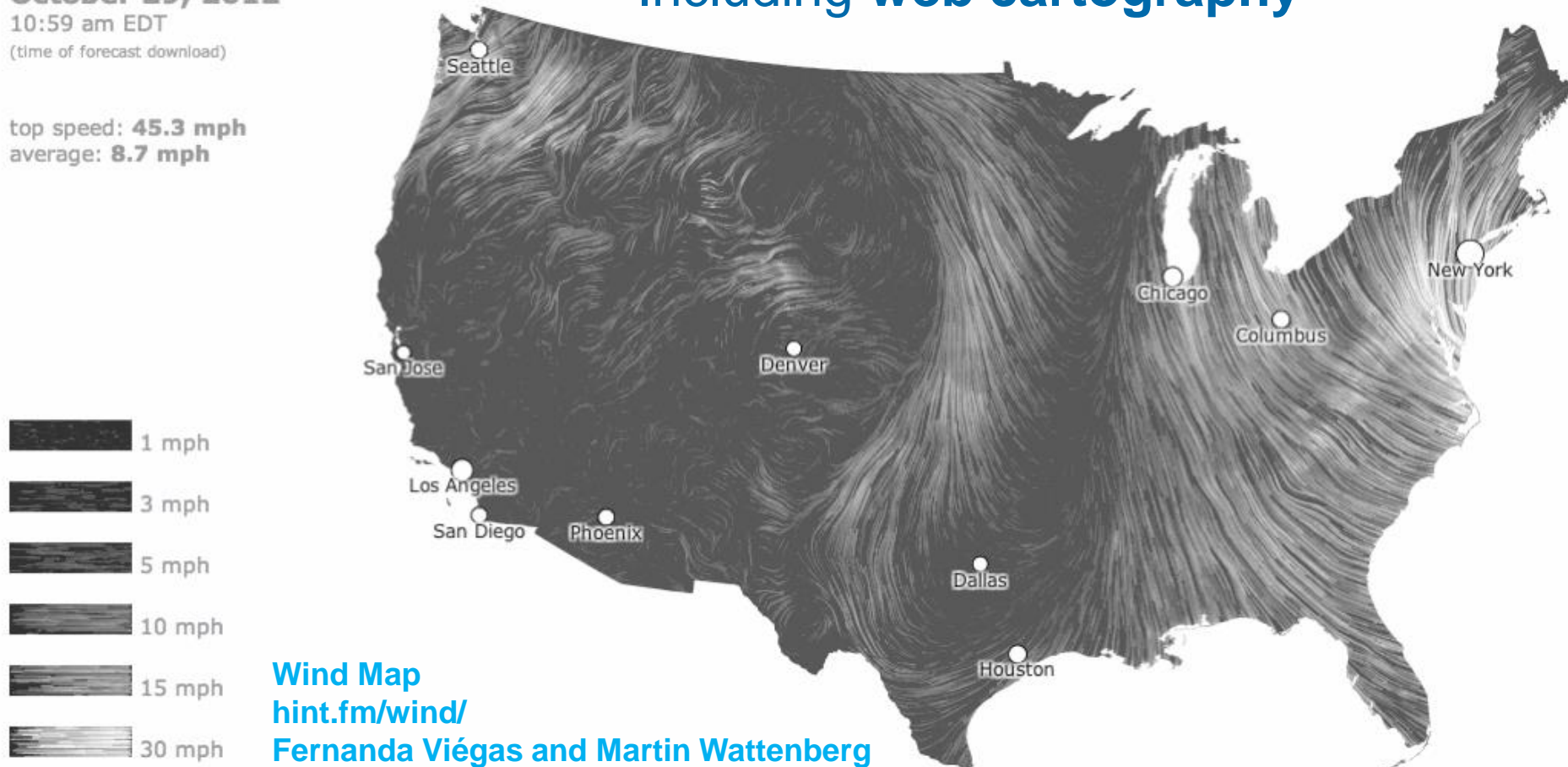
10:59 am EDT

(time of forecast download)

top speed: **45.3 mph**

average: **8.7 mph**

Including web cartography

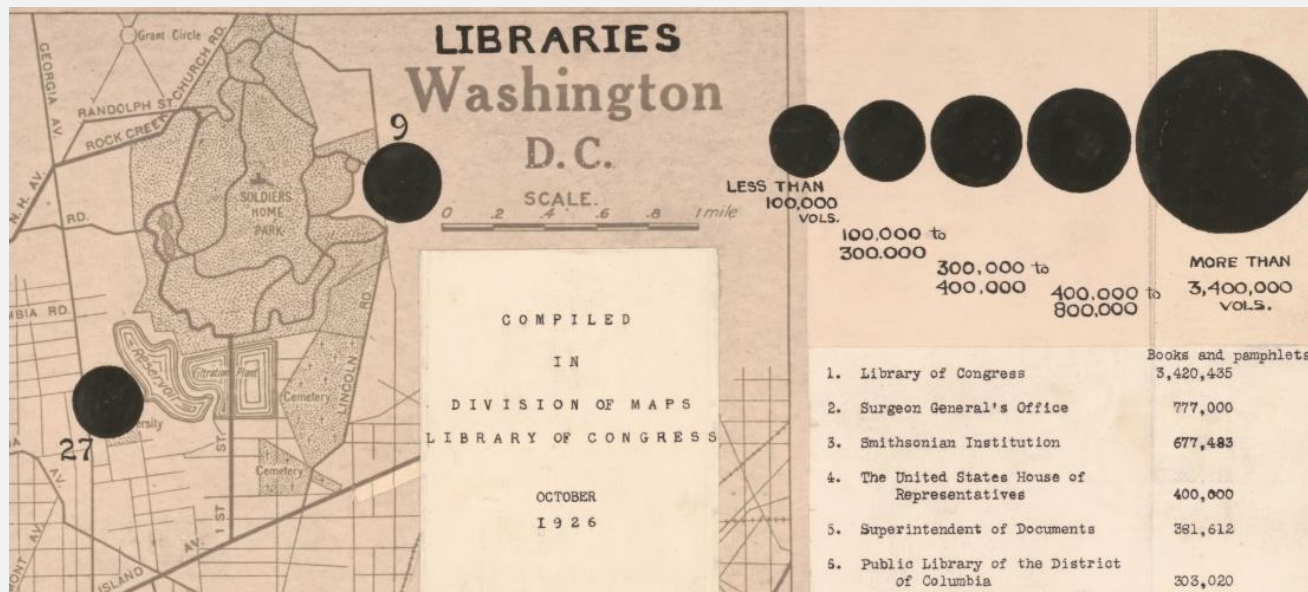


Wind Map
hint.fm/wind/
Fernanda Viégas and Martin Wattenberg



Challenges in Geospatial Data Preservation

- Challenges
 - Variety of data formats
 - Proprietary data structures
 - Metadata and data context
 - Web services and real-time data
 - Ensuring access to data



Detail of "Libraries, Washington D.C." by Division of Maps, Library of Congress, 1926. Geography and Map Division, Library of Congress. [LCCN: 87694100](#)

Geospatial Hosting Environment



- Digital environment for geospatial data discovery, sharing, analysis, and web mapping across the Capitol Hill Campus and beyond
- Based on:
 - **ArcGIS Online and Portal for ArcGIS:** geospatial analysis, web map creation and dissemination
 - **Voyager Search:** geospatial dataset indexing and discovery software
 - **FME:** data transformation and management

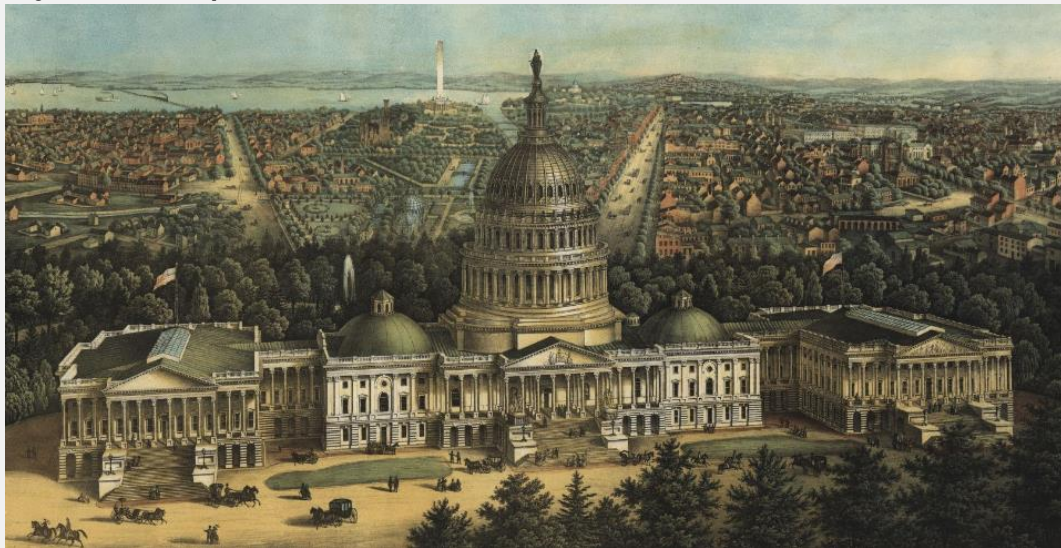


Geospatial Hosting Environment



■ USES

- Producing digital data and polished web map applications for Congress
- Creating web maps for a public audience that highlight collections of Geography and Map Division and other parts of LC
- Improving organization and stewardship of geospatial data holdings in Geography and Map Division



Detail of "View of Washington City" by E. Sachse & Co., 1871. Geography and Map Division, Library of Congress. [LCCN: 90683459](#)



Jacqueline Nolan, jnol@loc.gov

Colleen Cahill, cstu@loc.gov

Tim St. Onge, tong@loc.gov

Questions?

THANK YOU!

