Licensing Big Data

eDesiderata Forum 2016
Forum Agenda

Session I: Business & Financial Data  12:00-1:00 p.m. CT
Session II: Public Opinion & Population Data  1:00-2:00 p.m. CT
Session III: Geospatial Data  2:00-3:00 p.m. CT
Wrap up & outlook  3:00-3:15 p.m. CT
Session III: Geospatial Data Agenda

Overview of landscape Bernard Reilly, Center for Research Libraries

Presentation John Faundeen, United States Geological Survey

Presentation Amber Leahey, Ontario Council of University Libraries

Presentation Julie Sweetkind-Singer, Stanford University

Conversation
  • Bernard Reilly (moderator)
  • John Faundeen
  • Amber Leahey
  • Julie Sweetkind-Singer
Session III Geospatial Data: An Overview

Bernard Reilly
President, Center for Research Libraries
What is geospatial data?

- Geospatial data has explicit geographic positioning that ties it to a place or region on the earth.
  - Latitude/longitude, boundaries, address, zip code
How is geospatial data collected?

- Active data collection through direct use of a GPS device, observation, satellite imagery, or direct data collection such as through decennial census.
- Passive data collection through sensors such as stream flow sensors, wildlife tagging, cell phones, and weather stations.
- Created and distributed by individuals, research groups, all levels of government, NGOs, private companies, etc.

Monterey Bay Bathymetry, MBARI, 1997
Geospatial information and decision-making

- Emergency preparedness: What houses lie within a 100-year flood plane?
- Climate change impacts: How will climate change impact the location of rice fields in China?
- Health care: What is the predicted rate and spatial spread of the Zika virus over the next year?
- Travel: How long will it take me to get from my house to the airport at 6pm on a Friday?

Predicted Zika distribution, eLife, NPR Health
Geospatial Data

John Faundeen, Archivist, U.S. Geological Survey (USGS), Earth Resources Observation and Science (EROS) Center
Outline

- Context
- Three Case Studies
- Summary
Context

- **Department of the Interior**
  - U.S. Geological Survey
    - Earth Resources Observation and Science Center
- **Established 1972**
  - Landsat Satellite Program
  - Aerial Photography Missions
- **Archives and Science Co-located**
Context

• **Satellite Observations**
  • Film: > 4,000,000 frames
    • 1960 – 1992
  • Digital: > 9,400,000 images
    • 1960 – Present

• **Aerial Observations**
  • Film: > 6,600,000 frames
    • > 50,000 Rolls
    • 1937 – Present
  • Digital: 1937 – 2012
Case Study 1: SPOT

- SPOT (Satellite Pour l’Observation de la Terre)
  - French
- North American Distributer SICORP
- SPOT Satellites 1-3
  - 1986-1998
  - 1,000,000 images
- Transfer to USGS late 1990s
  - $600 per image ($300 royalty & $300 cost recovery)
    - Unsuccessful
    - Evolved to free and open distribution
Case Study 2: GeoEye OrbView-3

- **American Hi-Resolution Satellite Company**
  - OrbView-3 Platform
    - 2003 – 2007
    - 179,000 Image strips

- **Department of Commerce**
  - U.S. Government first rights of refusal
  - National Satellite Land Remote Sensing Data Archive

- **Free and Open Distribution**
Case Study 3: Planet Labs

- American Hi-Resolution Satellite Company
- 2015 Purge Notification to DoC
- Over 100 Satellites in Orbit
  - 2011 – 2015 in offer
  - 2,000,000 images
- USGS EROS Appraisal Process
  - Archivist, Day-to-Day Manager, Scientist, Senior Management
  - Offer not accepted
    - Would have resulted in free, open access
Summary

• U.S. Government
  • Interaction with private sector

• Majority of holdings free and open/public domain

• U.S.-based Satellite Companies
  • Department of Commerce License
  • U.S. Government first right of refusal

• Offers viewed as eventual free and open distribution (May involve Sunset Dates)
Geospatial Data

Licensing geospatial data in Ontario Libraries

Amber Leahey, Data and Geospatial Librarian
Scholars Portal, Ontario Council of University Libraries
Acknowledgements

- Marcel Fortin, Head, Map and Data Library, University of Toronto Libraries
- Carol Stephenson, OCUL Licensing Officer
- OCUL Geo Community
Consortia approaches to licensing

- Coordination between collections librarians and subject librarians

- Different kinds of license models:
  - Individual user license
  - Institutional license
  - Consortium license

- Canadian examples:
  - Ontario Geospatial Data Exchange (OGDE), Land Information Ontario (LIO) [Consortium, Institutional license, SP/OCUL distribution]
  - DMTI Inc. [National consortium; University of Toronto led, SP/OCUL distribution]
  - Data Liberation Initiative (DLI), Statistics Canada [National consortium; institutional license, SP/OCUL distribution]
  - Municipal data - City of Mississauga [OCUL license, will be SP/OCUL distributed]
Benefits to consortia purchasing

- Moves away from the ad hoc approach
- Reduce producer’s distribution burden
- Leverage negotiation from the position of multiple institutions buy-in
  - Cost sharing and/or reduced price
  - Terms of use
  - Redistribution rights
Negotiating academic use

Data vendors: academics aren’t their key audience!

• Terms of use
• Access for who? 1-2 researchers, departments, whole university, public…
• Delivery and access mechanism (user name and password, local ftp, web portal…)
• Metadata (repositories, catalogue, etc.)
• Redistribution rights for loading in web portals
• Perpetual rights, historical copies
• Preservation
Access and dissemination

Ontario’s Scholars GeoPortal (http://geo.scholarsportal.info)
Challenges providing access

• Large datasets (imagery in the TBs)
• Annual editions
• Data streams?
• Derivative products (Stereo imagery, DEMs, etc.)
• Multiple access conditions and restrictions
• Metadata

Storage, hardware, resources, expertise…
Long-term preservation

- In 2013, Scholars Portal became a Trusted Digital Repository for OCUL’s journal article collections
- Plans for extending this to data are underway
- Variety of geospatial collections including scanned maps, vector data, non-spatial geographic data, raster imagery, etc.
- Metadata in ISO 19115 standard
- Thinking about open, stable, formats for geospatial data
  - TIFF, JPEG 2000
  - SHP
  - KML
  - XML
  - ....
Geospatial data

Julie Sweetkind-Singer
Assistant Director of Geospatial and Cartographic Services
Head of Branner Earth Science Map Library & Map Collections
Stanford University
Swimming in a sea of data

- Data from a wide variety of producers – finding it may be a lot of work.
- Data may be of unknown provenance, of unknown quality and with scant metadata.
- Access may be difficult due to the size of the data, especially with satellite imagery.
- Data producers may not be used to working with universities.
Know what you’re buying

- Request samples of the data
- Request samples of the metadata
- Ask for names of others who have purchased the data and talk to them
- Ask for a copy of the license right away
Vet the license carefully

- Perpetual access
- Ability to share within your university
- Ability to incorporate into a spatial data infrastructure
- Ability to perpetually archive
- Ability to publish and how
- Clear wording on necessary attribution
- Clear wording on updates and data changes
- Think ahead! Allow for flexibility later
Traffic Bottlenecks, California, 2011

Author(s): California Department of Transportation

Description: This line shapefile represents traffic bottlenecks, or heavy congestion during peak a.m. and p.m. travel periods, on California’s State Highway system during 2011. These data were derived from the California Department of Transportation (Caltrans) HQ Traffic Operations Asset Management Branch and prepared by Caltrans as part of the Mobility Performance Report (MPR). The MPR is prepared by Caltrans and district staff providing detailed data about highway system performance related to congestion and mobility. Caltrans collects vehicle counts and calculates speeds at all hours of the days and the week of major metropolitan areas throughout California via the Caltrans Performance Measurement System. This information helps identify congestion bottlenecks and results in more cost-effective investments to improve performance of the State Highway System. This layer is part of a collection of GIS data created by the California Department of Transportation (Caltrans). This dataset is intended for researchers, students, and policy makers for reference and mapping purposes, and may be used for basic research.

Publisher: California Department of Transportation

Place(s): California

Subject(s): Traffic density, Traffic flow, Roads, and Transportation

Year: 2011

Held by: Stanford

More details at: [http://purl.stanford.edu/sz080jh0301](http://purl.stanford.edu/sz080jh0301)

Use and reproduction: This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. [http://www.gnu.org/licenses/](http://www.gnu.org/licenses/)

Copyright: Copyright resides with the originator.
Conversation: Geospatial Data

Bernard Reilly, Center for Research Libraries (Moderator)
John Faundeen, United States Geological Survey
Amber Leahey, Ontario Council of University Libraries
Julie Sweetkind-Singer, Stanford University

#BigDataForum_CRL
Contact Information: Geospatial Data

**Bernard Reilly**  
President, Center for Research Libraries  
Reilly@crl.edu

**John Faundeen**  
Archivist, U.S. Geological Survey (USGS), Earth Resources Observation and Science (EROS) Center  
faундеен@usgs.gov

**Amber Leahey**  
Data and Geospatial Librarian, Scholars Portal, Ontario Council of University Libraries  
amber.leahey@utoronto.ca

**Julie Sweetkind-Singer**  
Assistant Director of Geospatial and Cartographic Services, Head of Branner Earth Science Map Library & Map Collections, Stanford University  
juliets@stanford.edu
Licensing Big Data: Where do we go from here?
Upcoming CRL events

**Webinar: CLOCKSS, LOCKSS and the "Long Tail," Preserving at-risk Publishers**
Thursday, December 1
2:00 to 3:00 p.m. Central Time

**CRL Annual Council of Voting Members and Collections Forum**
Friday, April 21, 2017
(to be broadcast as virtual event)

Visit [www.crl.edu/events](http://www.crl.edu/events) to register for the webinar
For More Information

- Fill out our follow-up survey at [http://www.surveymonkey.com/r/CRLWebinarFollowup](http://www.surveymonkey.com/r/CRLWebinarFollowup)

- Find descriptions and analysis of individual Big Data resources in [www.edesiderata.crl.edu](http://www.edesiderata.crl.edu)

- Sign up for *CRL Connect*: [www.crl.edu/connect](http://www.crl.edu/connect)

- Find CRL on Facebook and Twitter