Center *for* Research Libraries
GLOBAL RESOURCES NETWORK

May 18, 2018

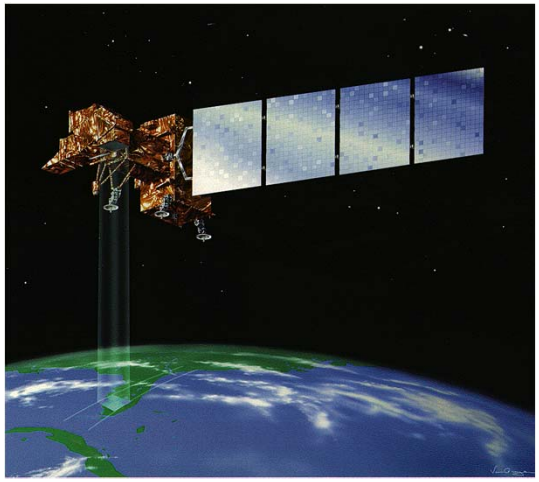
Social Science Research and the Challenges of Big Geospatial Data

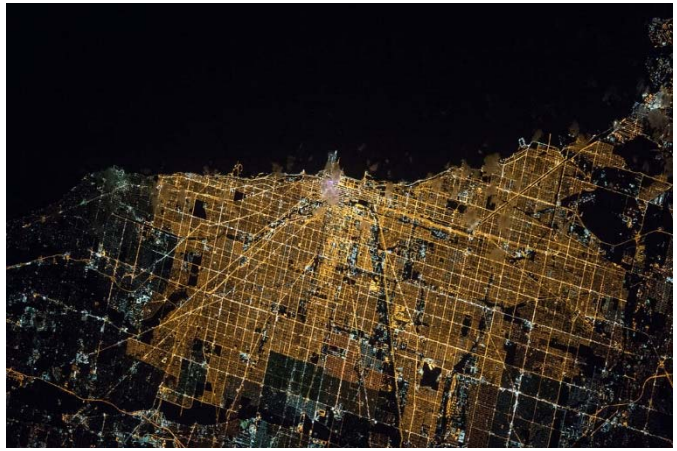
Julie Sweetkind-Singer
Head, Branner Earth Sciences Library and Map Collections
Stanford University Libraries

THE *NEW*
GLOBAL
RESOURCES
SUPPLY
CHAIN

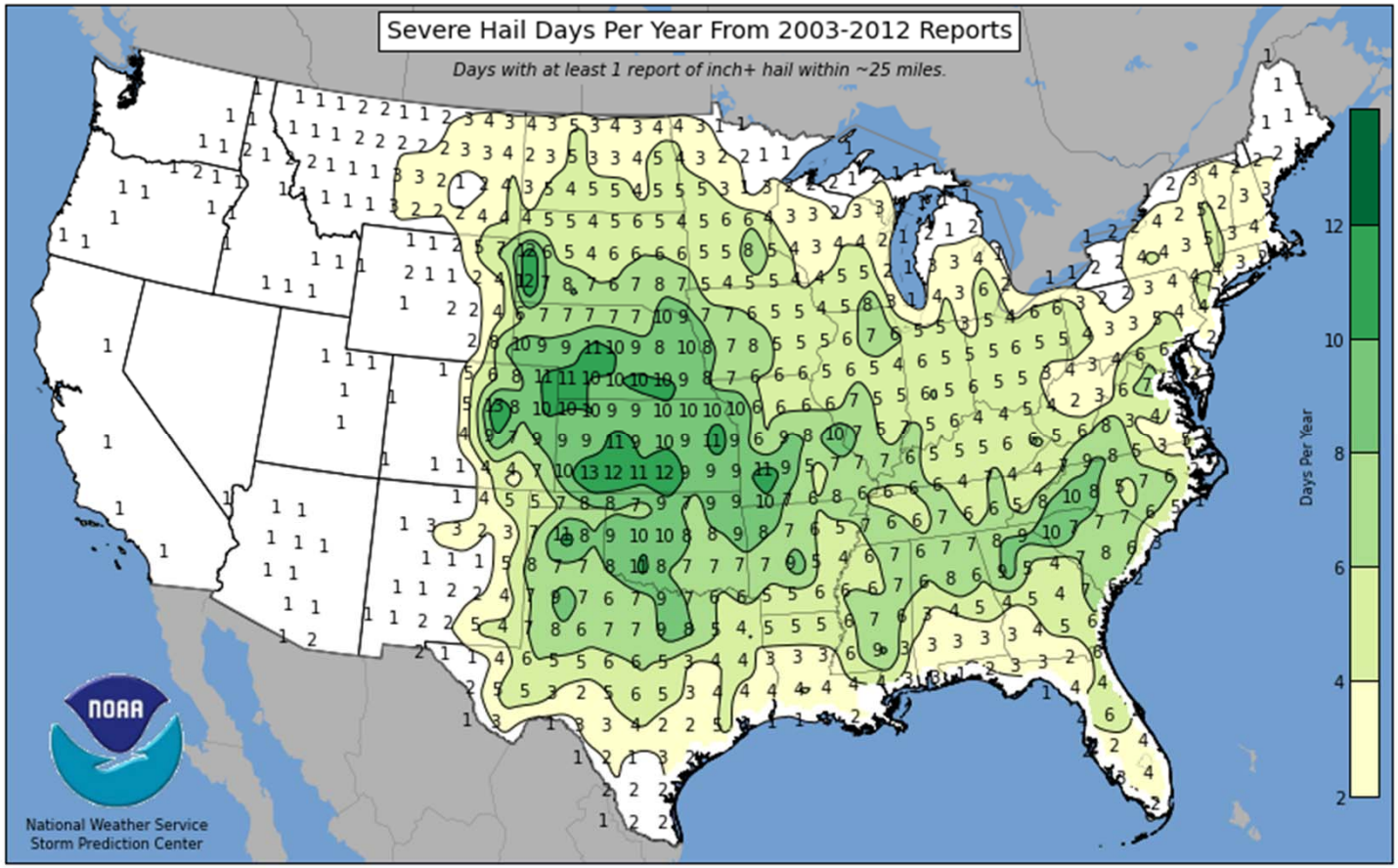
Overview

- Introduction to geospatial data
- Collection issues
- Rise of external data sources
- Management of data in-house





Left: Landsat 7 satellite. **Right:** Chicago at night. Images: NASA



GEOSPATIAL DATA

NOAA, 2012.

What is Geospatial Data?

- Geospatial data has explicit geographic positioning that ties it to a place or region on the earth. It includes latitude and longitude.

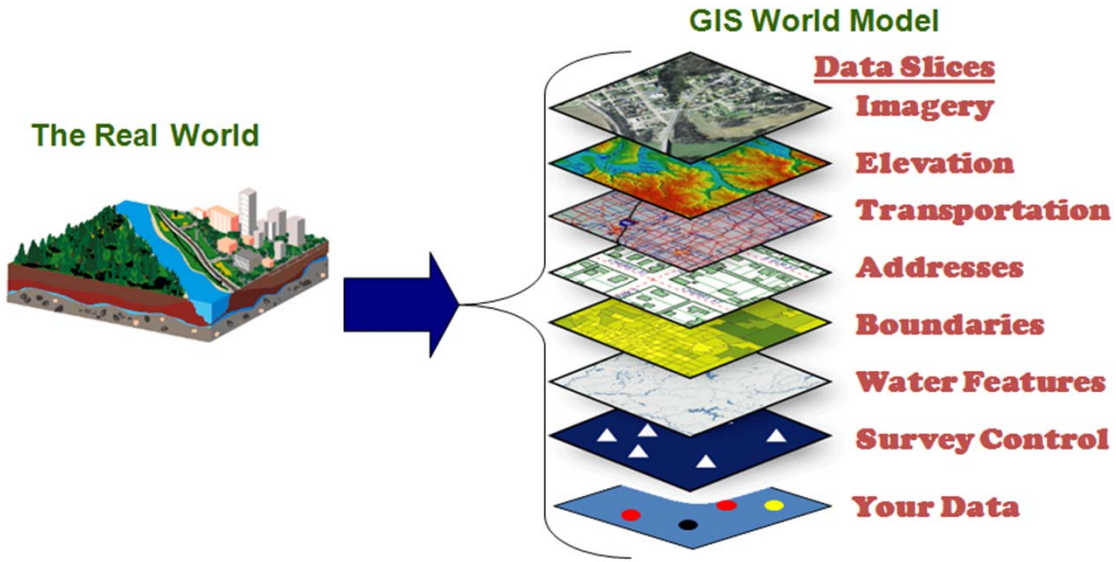
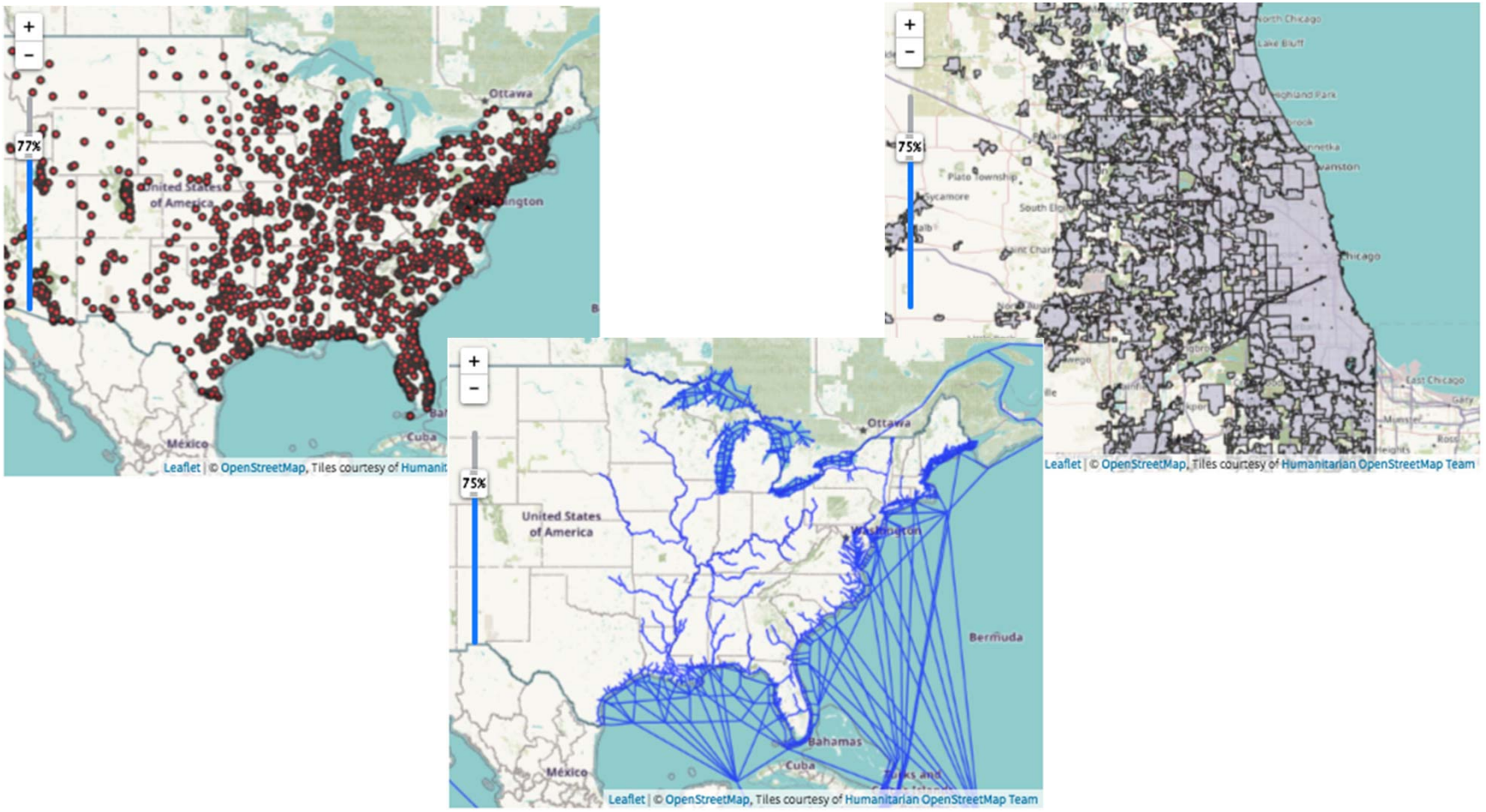


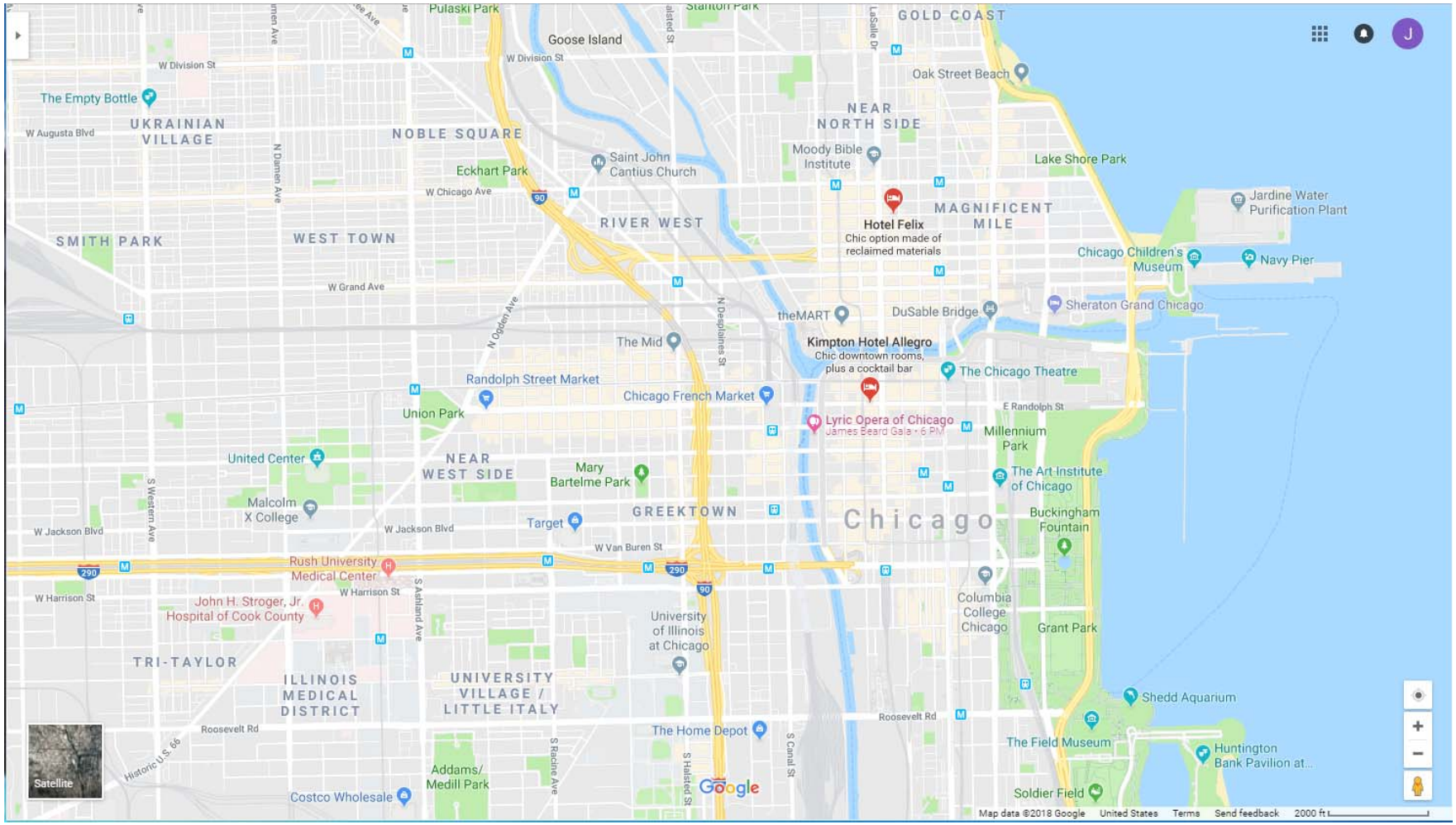
Image from the Indiana Geographic Information Office.

Vector Data

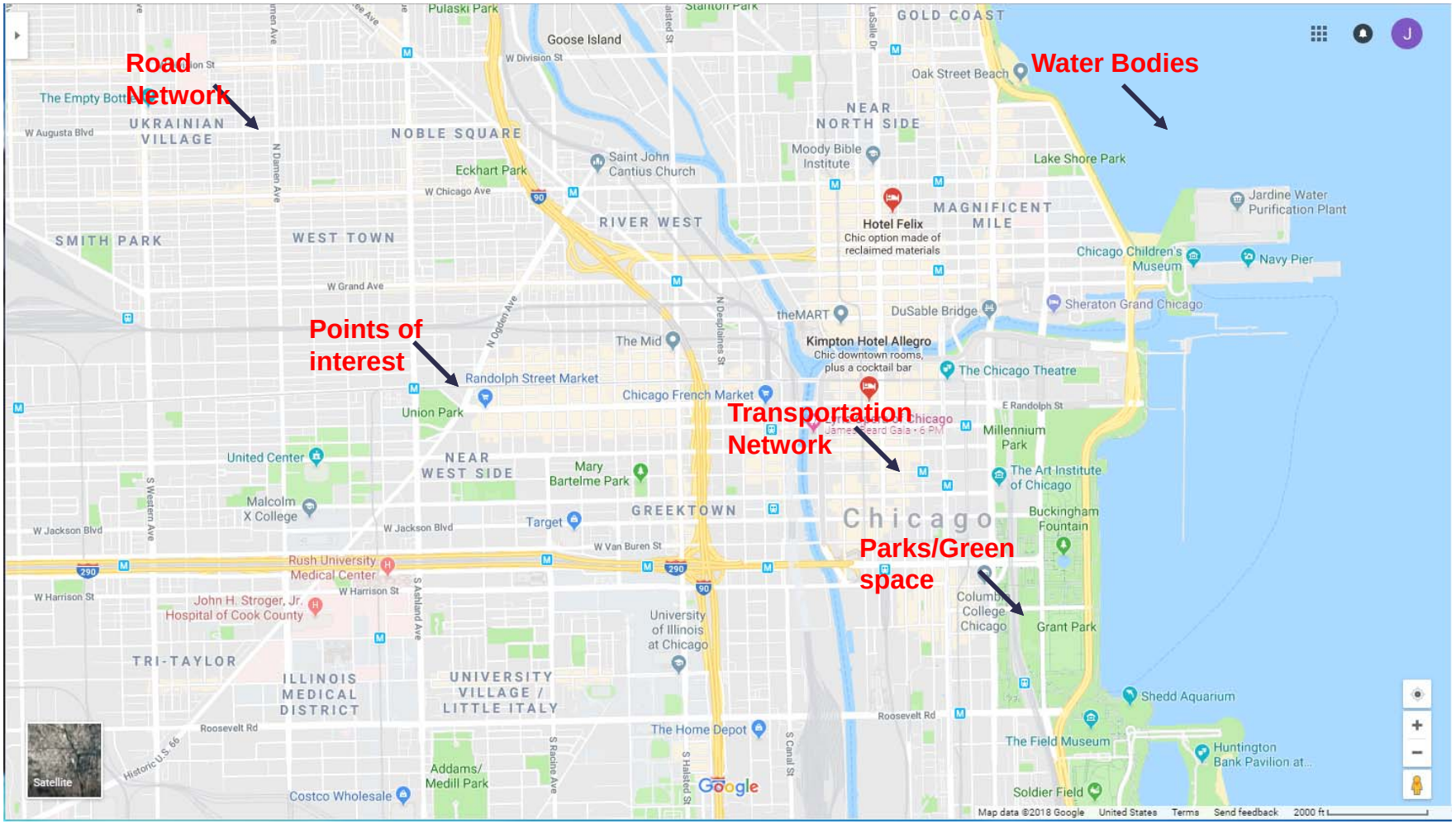
- Vector data are points, lines, and polygons.



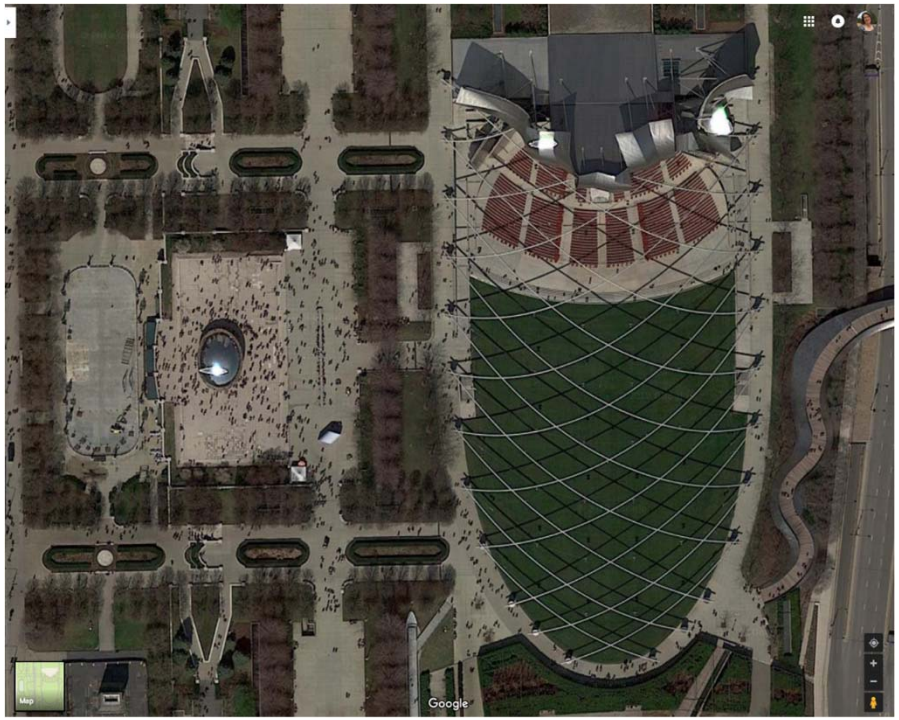
Left: United States cities 2010, ESRI. **Center:** Navigable Waterways, 2011, U.S Army Engineer Institute for Water Resources. **Right:** Census Places, 1990 – Illinois, US Census Bureau.



Google maps: Chicago

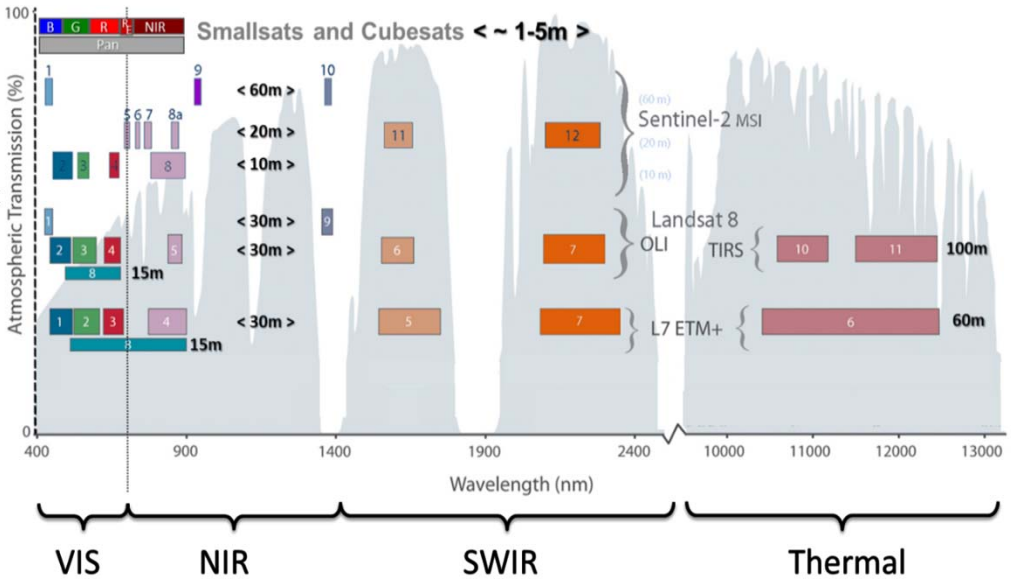


Google maps: Chicago

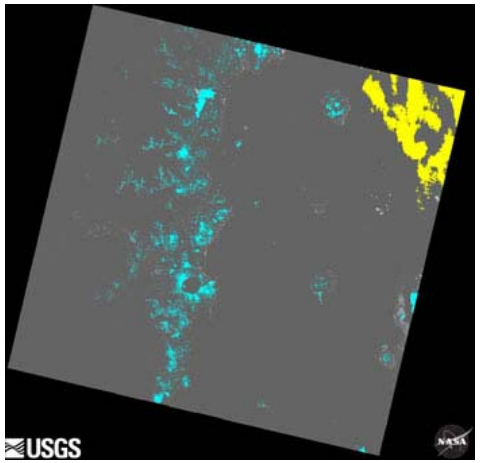


Top: Chicago, 2000, NASA

Right: Chicago, 2018, Google Earth



Red = Forest
Gray = Clear cutting



Blue = Snow/Ice
Yellow = Clouds

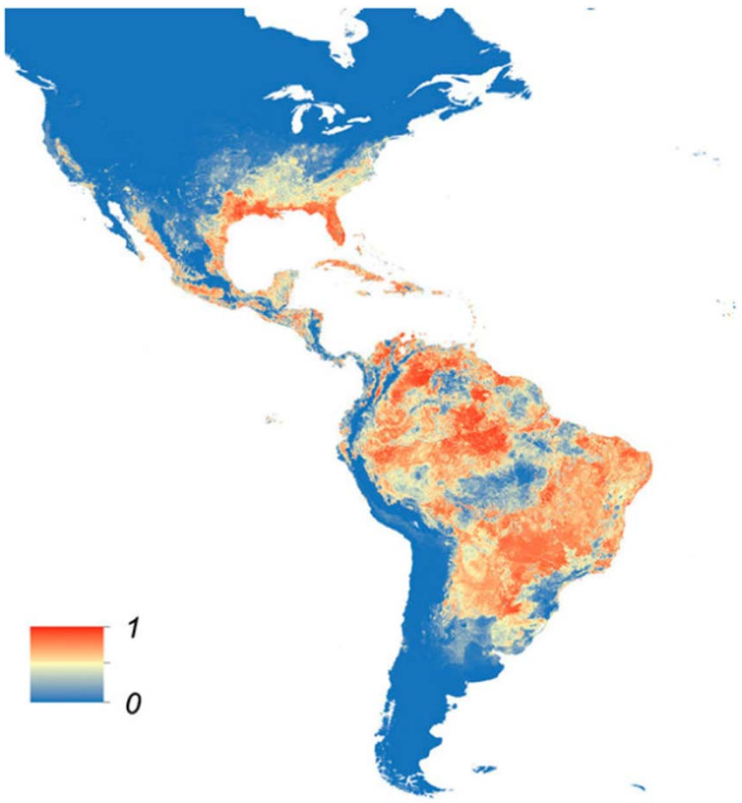
How is Geospatial Data Collected?

- 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.

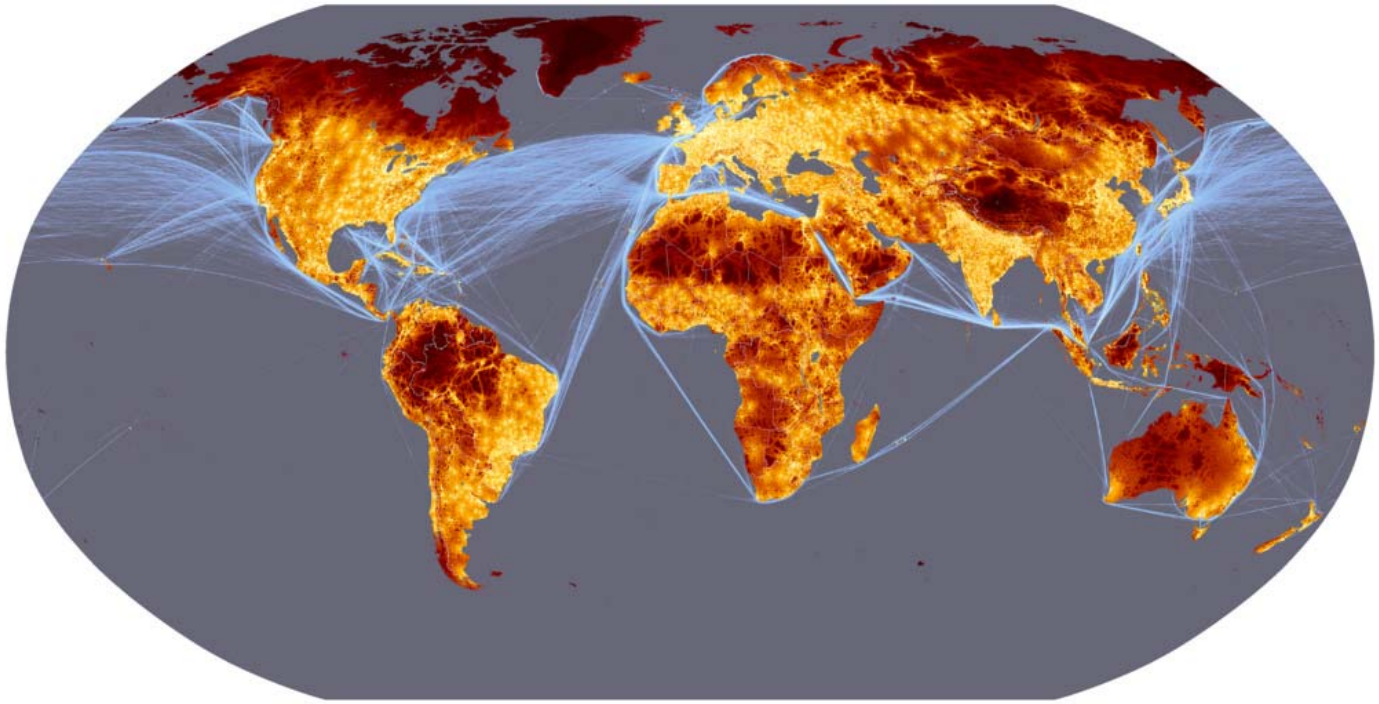
Monterey Bay Bathymetry, MBARI, 1997

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?

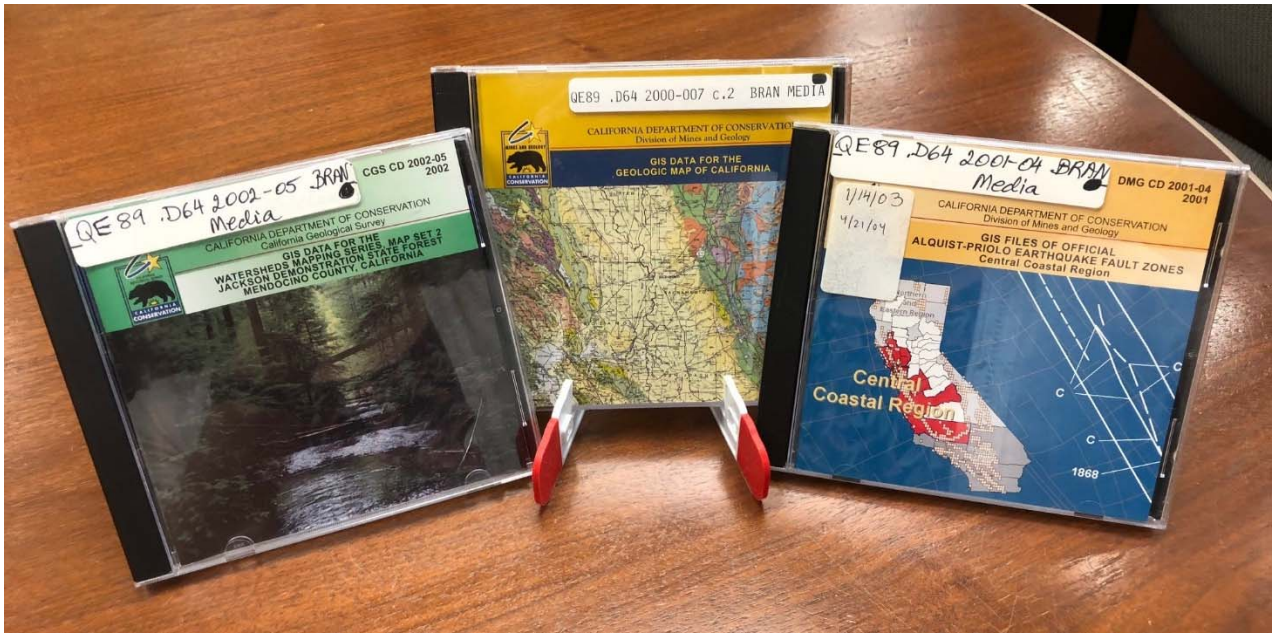


Predicted Zika distribution, eLife, NPR Health



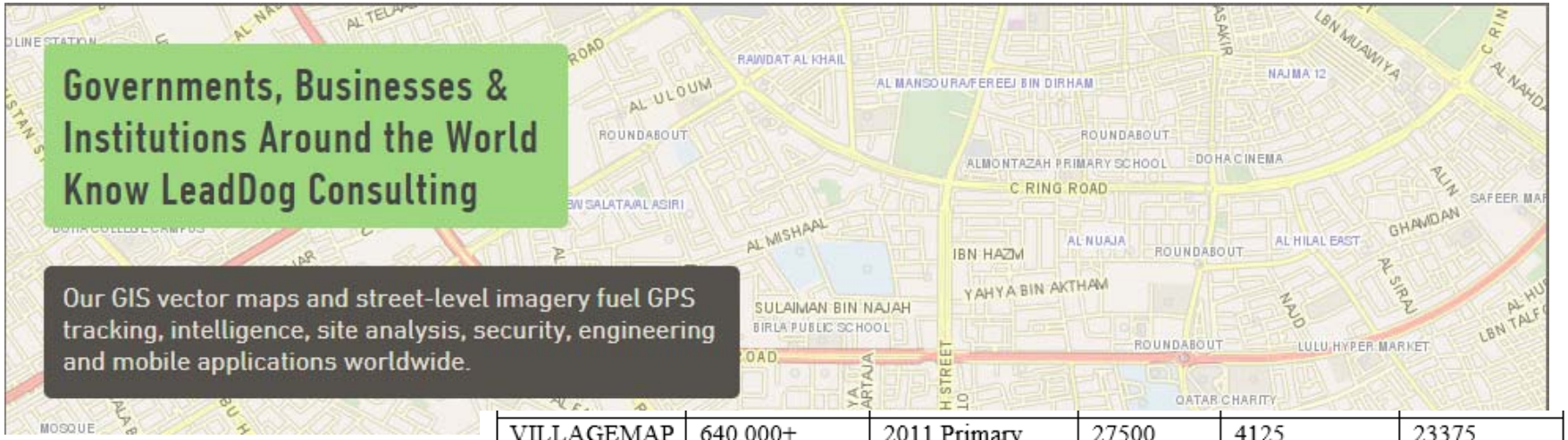
COLLECTION ISSUES

Travel time to major cities, World Bank, 2009. Flickr Commons.



California Geological Survey

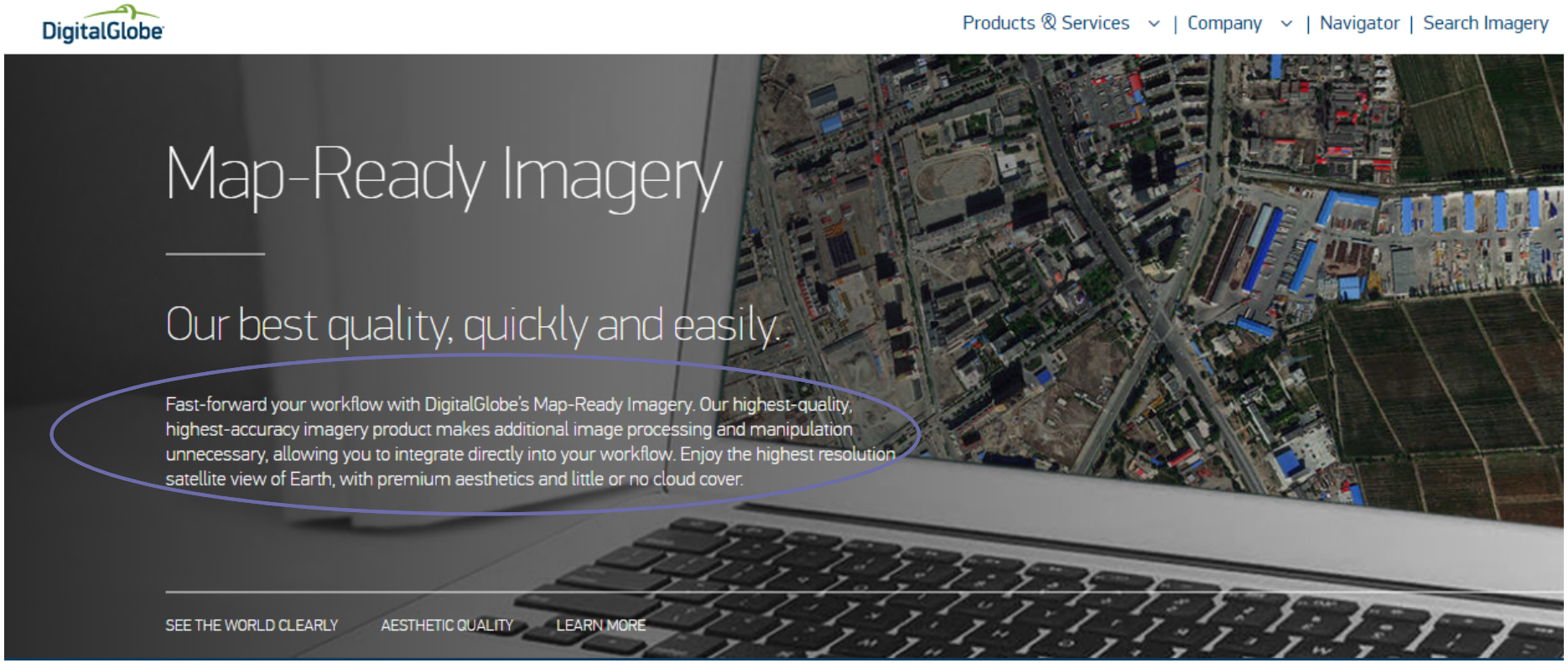
Geologic Map of California	CGS	This edition of the Geologic Map of California was prepared in recognition of the California Geological Survey's 150th Anniversary. It is an all-digital product built on the original compilation of C.W. Jennings published in 1977, with some additions and modifications. The Geologic Map of California presents an overview of the geology and structure of the state. It represents the geologic features that one would find on a visit to any locality in the state. The restraints of scale limit the detail that can be shown, but the most important geologic features are portrayed. The distribution of the major rock types and the major structural elements are shown with sufficient detail to be useful for many purposes.	Metadata	Download	2010-12-31
Geologic Map of the Lake Tahoe Basin, California and Nevada	CGS	To provide basic geologic information to those concerned with land use planning and decision-making in this part of California.	Metadata	Download	2005-01-01
Geologic Map of the Monterey 30'x60' Quadrangle and Adjacent Areas, California	CGS	To provide basic geologic information to those concerned with land use planning and decision-making in this part of California.	Metadata	Download	2002-01-01



VILLAGEMAP	640,000+ villages boundaries	2011 Primary Census Abstract	27500	4125	23375
LANDUSE /LANDCOVER	Physical boundary units and points	Level 1 classification in 45 types of landuse/landcover	7000	1050	5950
CLIMATE	District	Rainfall (mm) Mean Monthly for 4 years i.e. 2012, 2013, 2014, 2015	3500	500	3000
	District	Temperature (Degree Celsius) Minimum & Maximum Mean Monthly for 2014	3500	500	3000

Vendor-driven data sources:

- Vector
- Raster
- Streaming



- Pay to access model -> stop paying, no access
- Access to certain classes of people
- Reluctance to provide all of the data -> use their platforms
- Onerous terms -> Reaxsys

United Nations Committee of Experts on
Global Geospatial Information Management

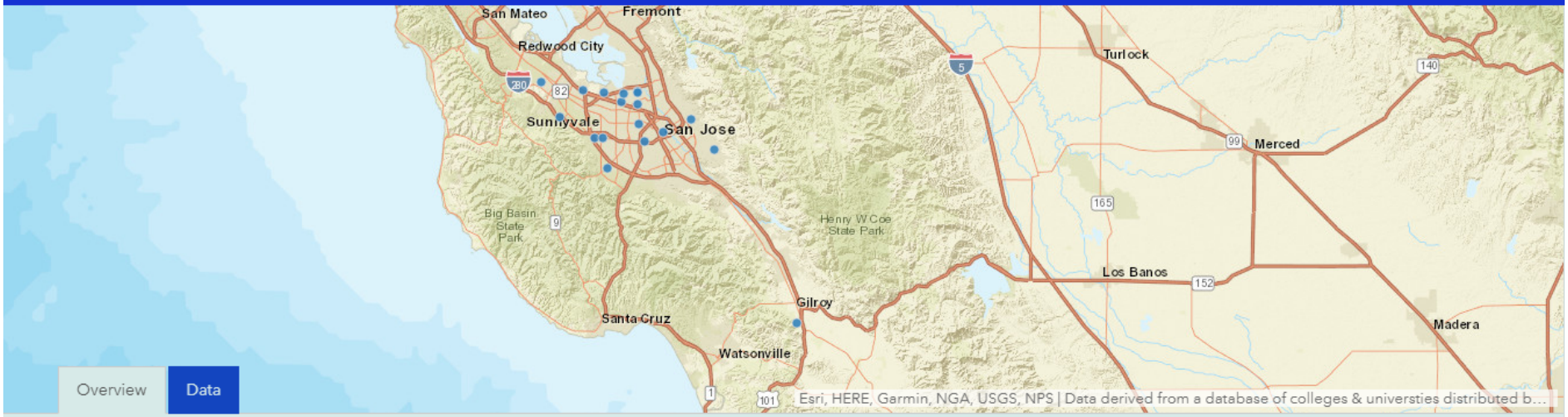
Compendium on Licensing of Geospatial Information

(final draft as at June 2017)

Available at: https://bit.ly/2I4UpmO

Santa Clara County Information Services Department GIS Data Portal

Find | My Data | Sign In



Overview | Data

SCC Schools CollegeUniversity

☆ Favorite | Download | APIs

Custom License | 6/15/2016 | Spatial Dataset | 18 Rows

This service compilation is made from a variety of data sources, most notably, County of Santa Clara tax rate area for district boundaries, CA Dept of Education for school point data information, County of Santa Clara for school land area information (Year 2015).

Attributes

Chart • Map Visualization

ADDRESS	Category	CITY	PLACENAME	PLACETYPE	SOURCE	UNIQUEID	ZIP	SHOW MORE
Text	Text	Text	Text	Text	Text	Text	Text	3 Attributes

About

Shared By:
Sampa.Patra@isd.sccgov.org_sccgov
Data Source: services.arcgis.com

View Metadata
Create Webmap
Create a Story Map

Camp Lemonnier (top) and a suspected CIA base
(bottom left) in Djibouti. Strava heatmap

Volume and Complexity

Satellite imagery

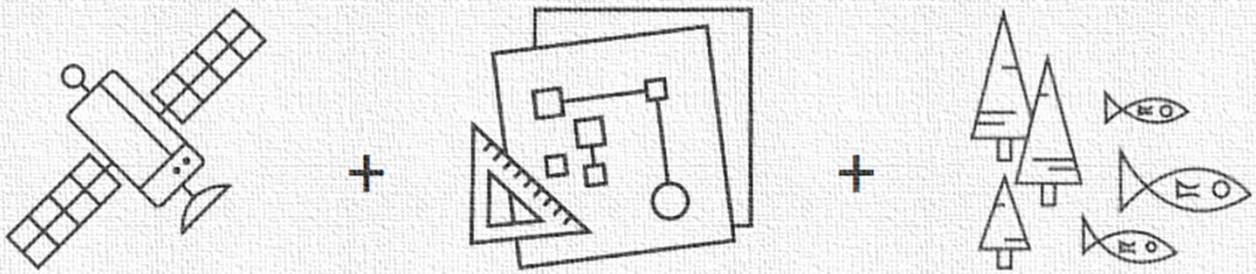
- Landsat 7 and Landsat 8 Data Acquisition Status – 1.3 petabytes of data
  - The Landsat 7 and Landsat 8 satellites collect about 1,200 new scenes per day. In addition, the Landsat Global Archive Consolidation effort continues to add numerous international scenes to the archive. The entire USGS Landsat archive now holds over 6.8 million scenes!
- Planet – 7+ Petabytes with 7+ Terabytes added daily
  - 88 new satellites recently launched
- DigitalGlobe and CNES/Astrium – probably bigger than Planet's holdings

Non-satellite imagery

- Population
- Health
- Housing
- Weather
- Elevation
- Land cover
- Climate
- Cropland

Meet Earth Engine

Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities and makes it available for scientists, researchers, and developers to detect changes, map trends, and quantify differences on the Earth's surface.



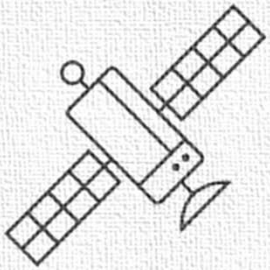
SATELLITE IMAGERY

YOUR ALGORITHMS

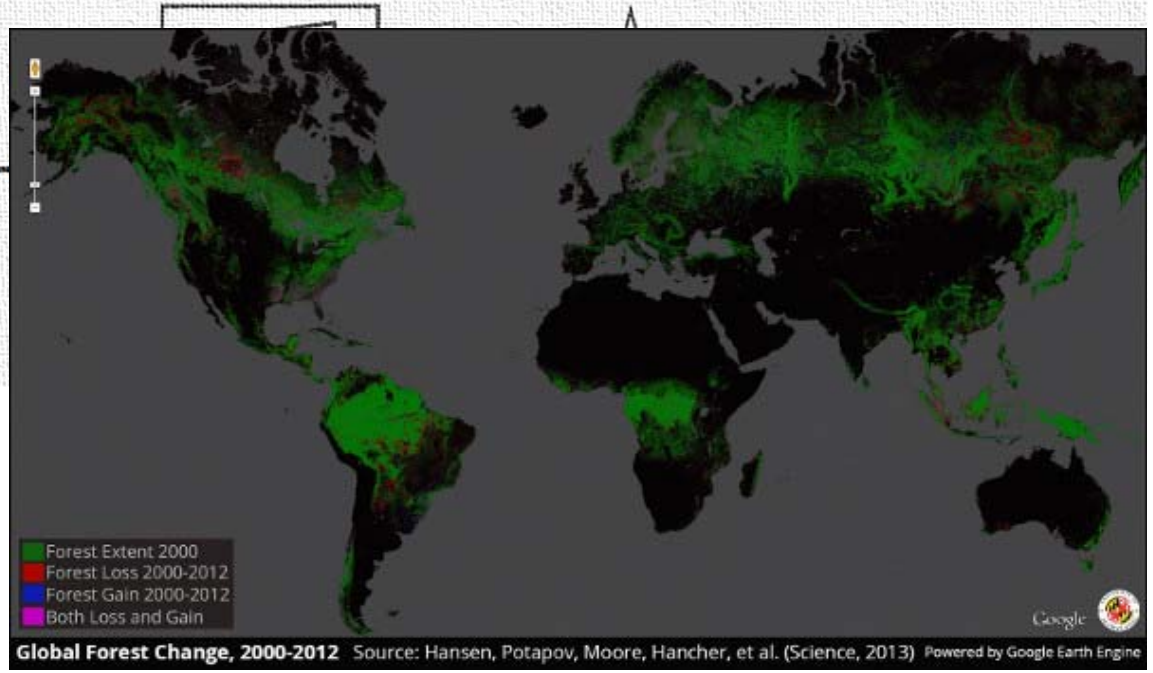
REAL WORLD APPLICATIONS

Meet Earth Engine

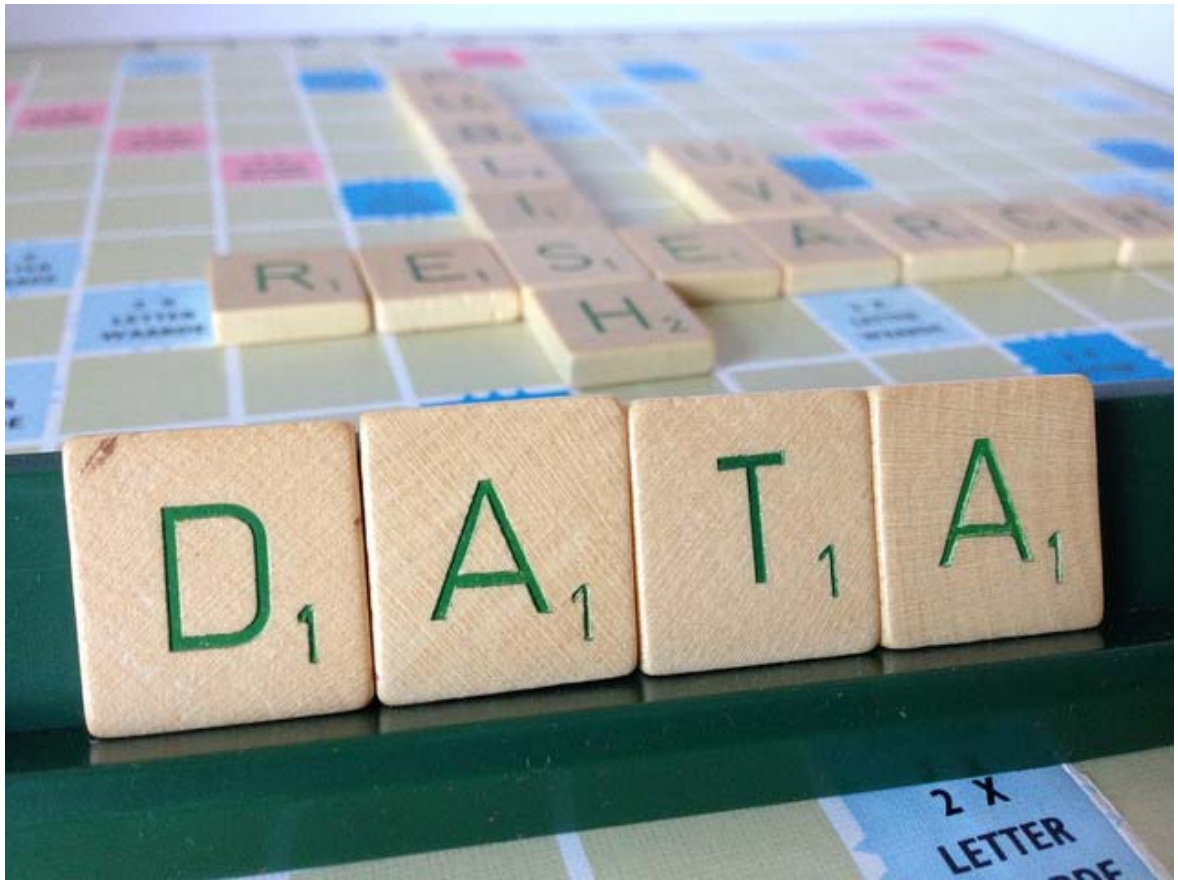
Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities and makes it available for scientists, researchers, and developers to detect changes, map trends, and quantify differences on the Earth's surface.



SATELLITE IMAGERY

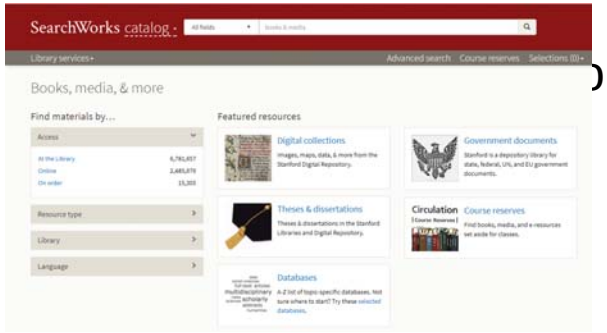


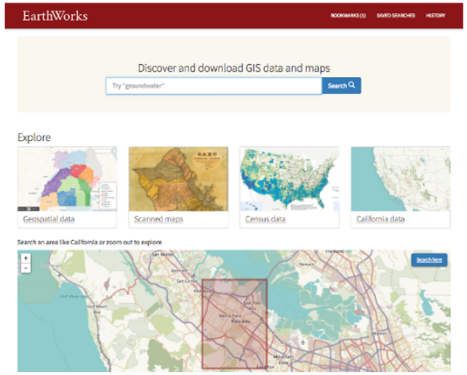
"This is the first map of forest change that is globally consistent and locally relevant. What would have taken a single computer 15 years to perform was completed in a matter of days using Google Earth Engine computing." -
- Professor Matt Hansen, University of Maryland

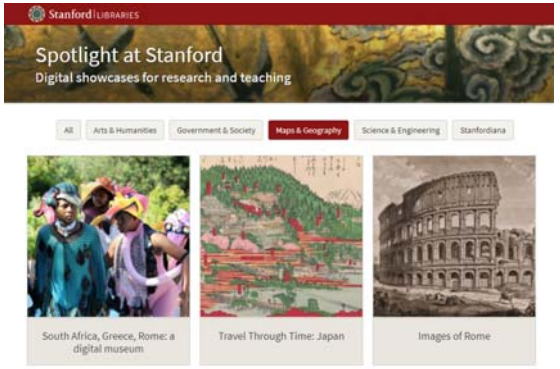


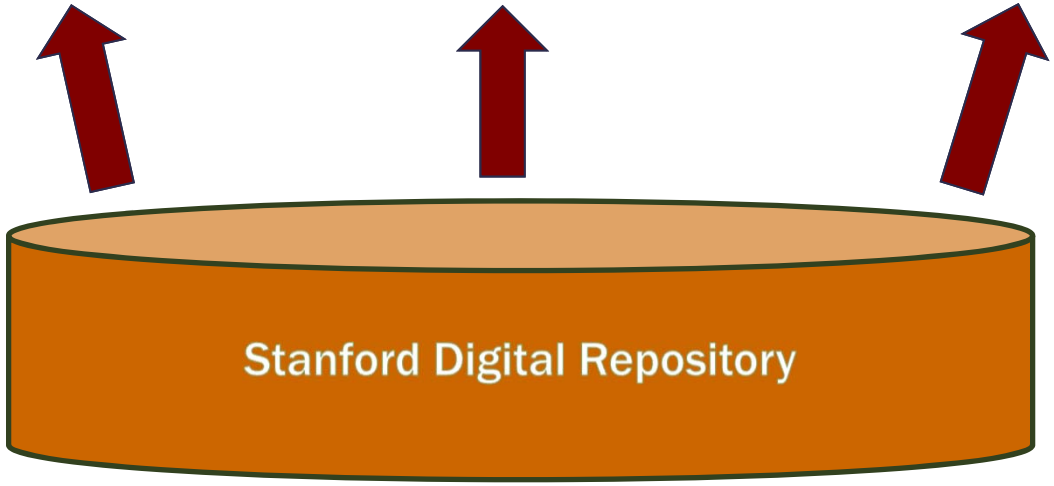
DATA MANAGEMENT FOR LIBRARIES

Janneke Staaks, Research Data Management. Flickr Commons.



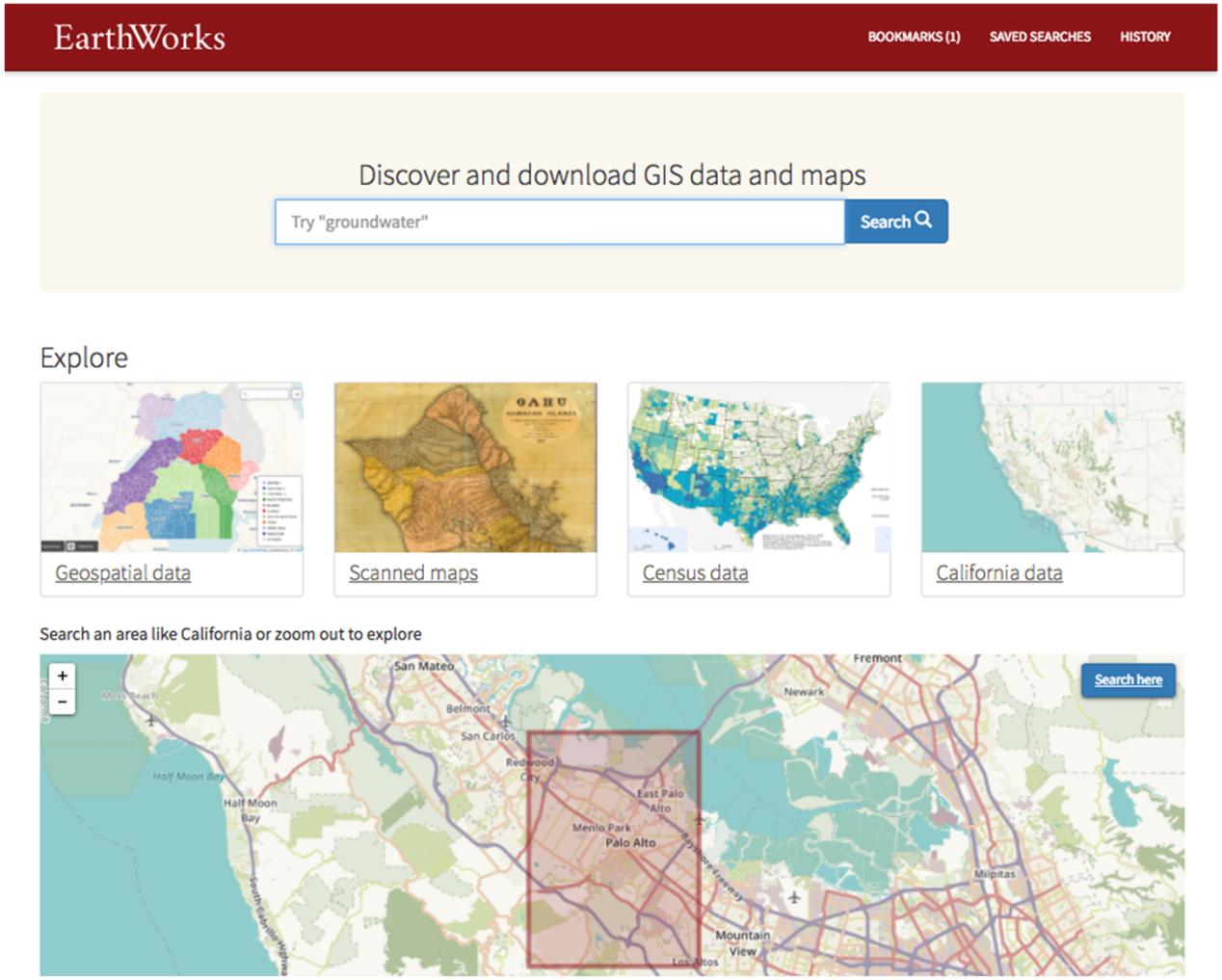






https://library.stanford.edu/research/stanford-digital-repository

Scalable, Robust, Shared Management and Preservation Services



https://earthworks.stanford.edu

chicago Search

Limit your search

Institution

- Michigan 24
- Harvard 15
- Stanford 11
- MIT 8
- Indiana 5
- Illinois 2
- Princeton 2
- Minnesota 1

Author

Publisher

Subject

Place

Year

Access

Data type

Format

You searched for: chicago ✖ Start Over

« Previous | **1 - 10** of **68** | Next »

Sort by relevance | 10 per page | Save this search

1. [Chicago road map].
2. Chicago, Illinois, 1853 (Raster Image)
3. Chicago, Illinois, Great Fire, 1871 (Raster Image)
4. Administrative Boundaries, Chicago, United States, ...
5. Central Business District, Chicago, United States, 1990
6. Urban Development, Chicago, United States, 1990-...
7. Urban Footprint, Chicago, United States, 1990
8. Urban Footprint, Chicago, United States, 2000
9. Urban Land Cover, Chicago, United States, 1990
10. Urban Land Cover, Chicago, United States, 2000

« Previous | Next » | 1 | 2 | 3 | 4 | 5 | 6 | 7

Chicago, Illinois, Great Fire, 1871 (Raster Image)

Description This layer is a georeferenced raster image of the historic paper map entitled: Map of Chicago showing the burnt district, presented by Freeman Burr. It was published by Freeman Burr in 1871. Scale [ca. 1:29,300]. The image inside the map neatline is georeferenced to the surface of the earth and fit to the Illinois East State Plane Coordinate System NAD83 (in Feet) (Fipszone 1201). All map collar and inset information is also available as part of the raster image, including any inset maps, profiles, statistical tables, directories, text, illustrations, index maps, legends, or other information associated with the principal map. This map shows features such as the area burned by the Great Chicago Fire of 1871, roads, railroads, railroad stations, drainage, city wards, and more. This layer is part of a selection of digitally scanned and georeferenced historic maps from The Harvard Map Collection as part of the Imaging the Urban Environment project. Maps selected for this project represent major urban areas and cities of the world, at various time periods. These maps typically portray both natural and manmade features at a large scale. The selection represents a range of regions, originators, ground condition dates, scales, and purposes.

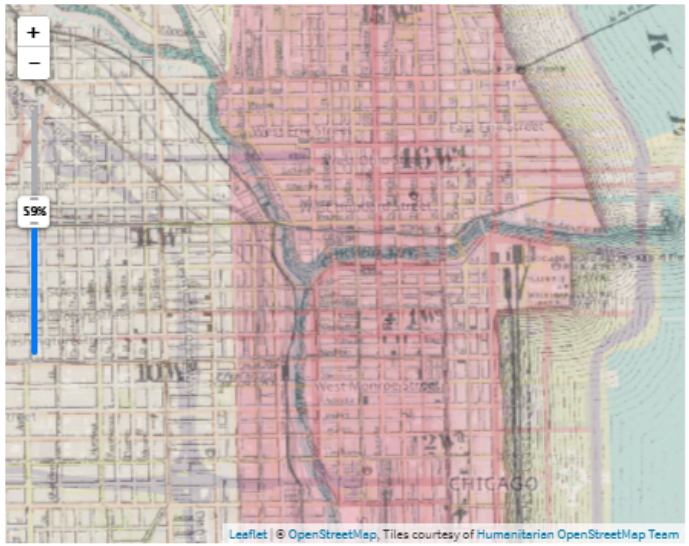
Publisher Harvard Map Collection, Harvard College Library

Place(s) Illinois and Chicago

Subject(s) Maps, Fires, Great Fire, Chicago, Ill., 1871, Human settlements, Cities and towns, Land use, Landforms, Infrastructure (Economics), Transportation, Bodies of water, and imageryBaseMapsEarthCover

Year 1871-01-01T01:01:01Z

Held by Harvard



Attribute	Value
Click on map to inspect values	

Tools

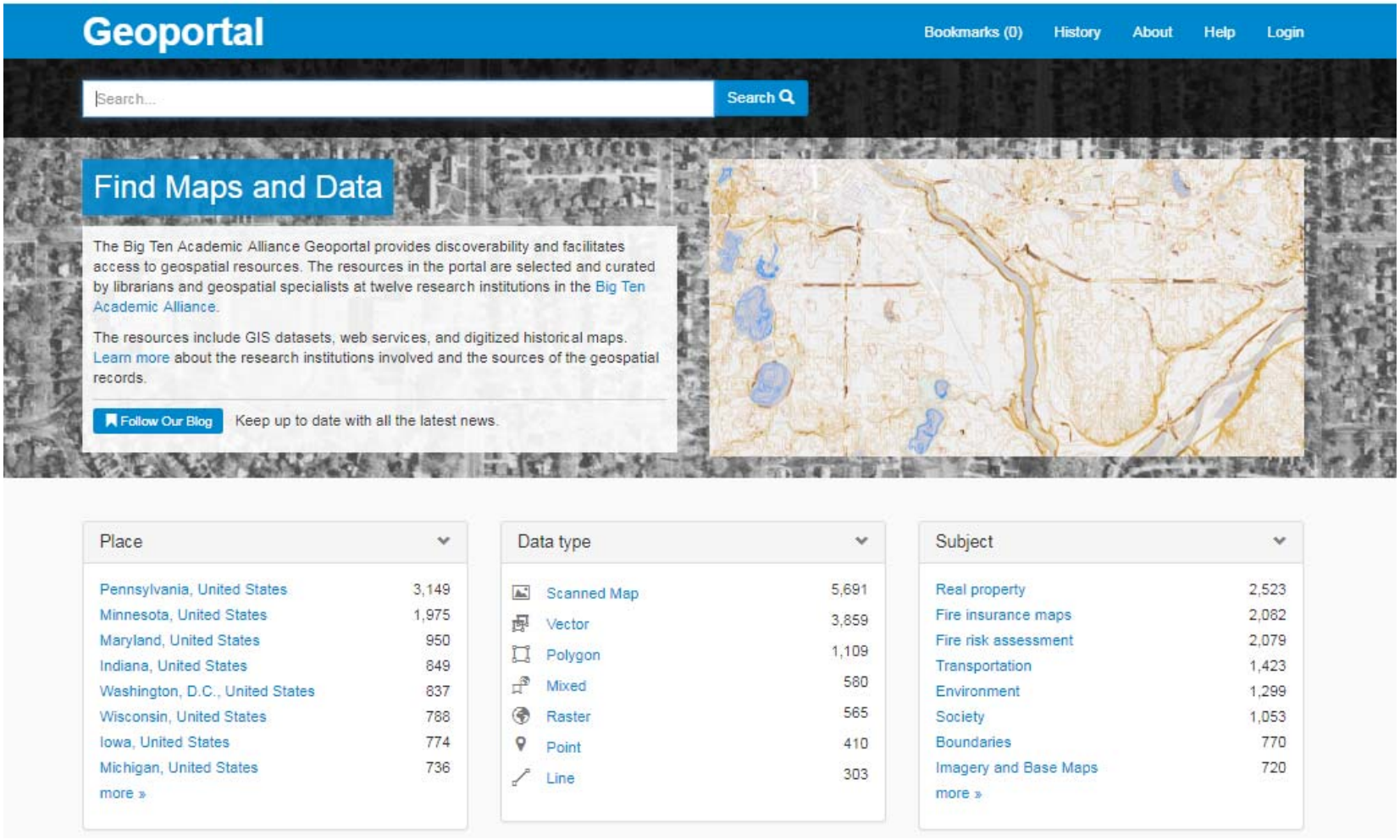
Bookmark

Email

Web services

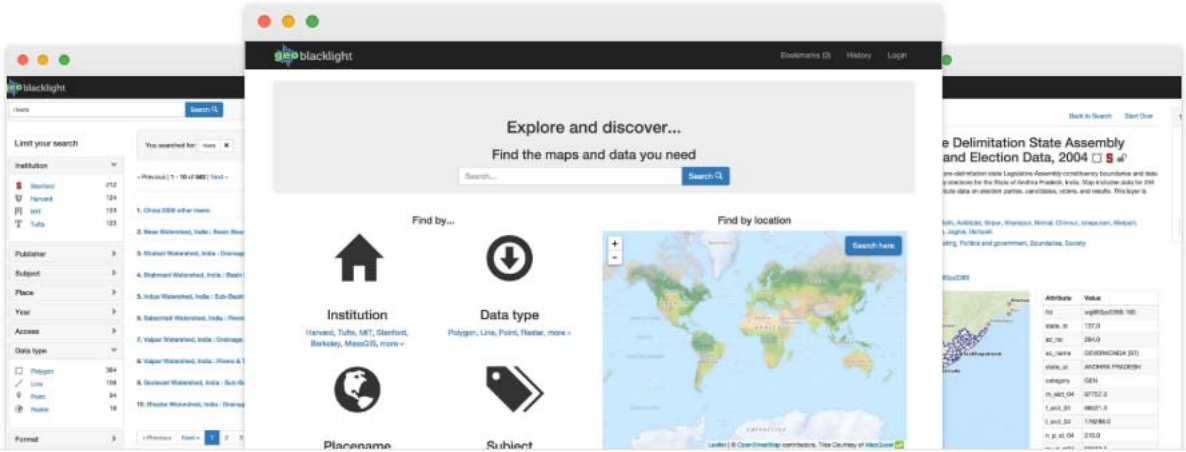
Download GeoTIFF

https://geo.btaa.org



GeoBlacklight

A multi-institutional open-source collaboration building
a better way to find and share geospatial data.



Easy to install

Quick and easy to get going with GeoBlacklight. Follow the simple installation instructions or checkout one of the tutorials.

Easy to customize

GeoBlacklight uses SASS and Twitter Bootstrap to enable adopters to easily customize their installation.

Built for longevity

Modularized software projects that enhance each other. Test coverage and documentation is a priority.

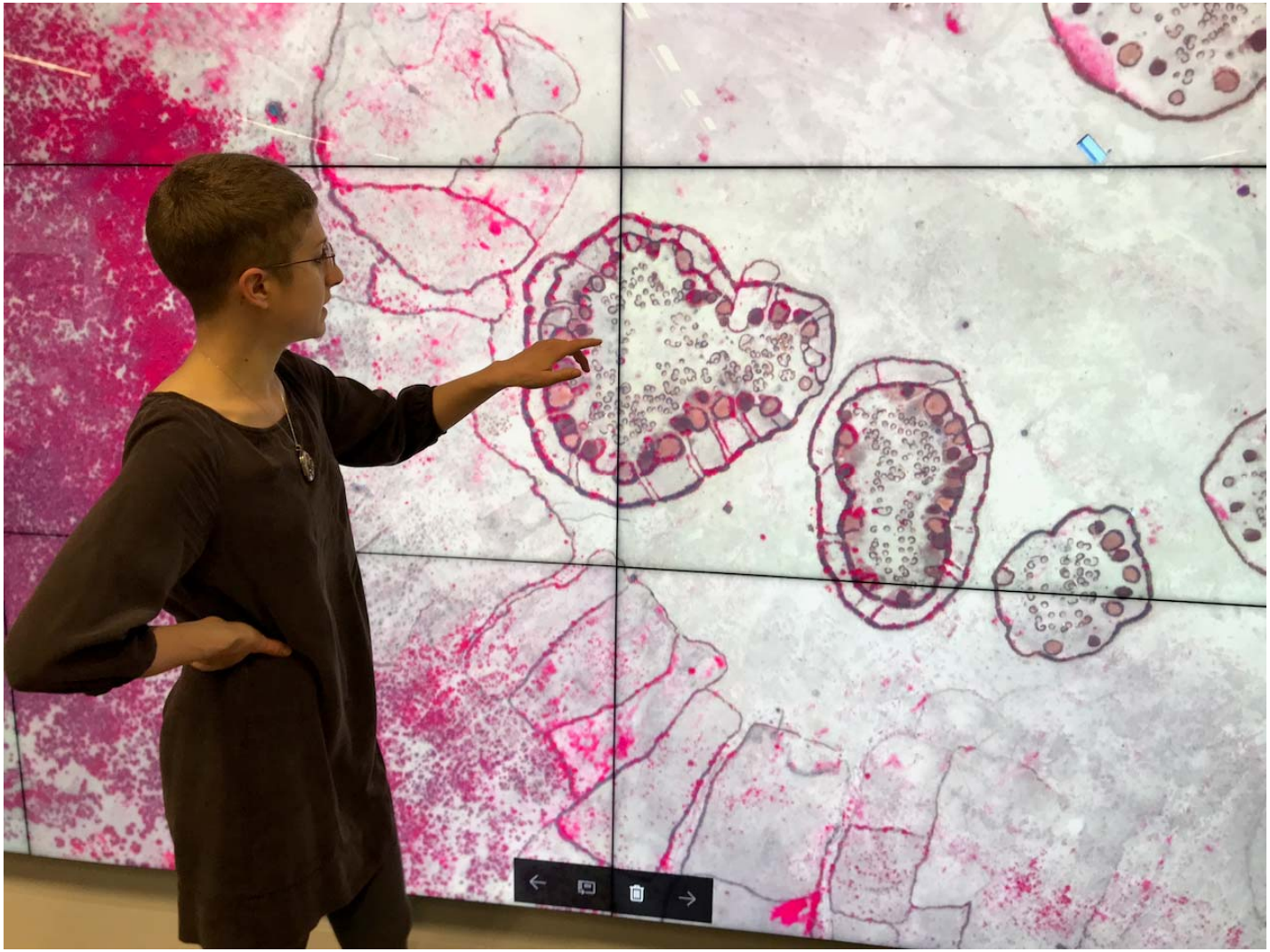
https://geoblacklight.org

David Medeiros and Stace Maples.
Stanford Geospatial Center staff
Credit: Danielle T. Tucker

- Data acquisition
- Data management
- Software management
- Training classes
  - Software
  - Data collection
- Management of services
  - Geocoding servers
  - Secure computers
- Programming/computer skills
  - Python
  - Java Scripting
  - R
  - Github

Credit: Stace Maples

Credit: Stace Maples



Credit: Stace Maples

Conclusion

- Increasingly research questions across a huge array of disciplines have a spatial component.
- Complex data are increasing in type, volume, and rapidity of output.
- Few repositories are capable of collecting and managing the content.
- Increasing need in libraries to provide high level staff expertise for data collection, licensing negotiations, creating access portals, and providing end-user support.
- Pushing questions to the cloud create amazing opportunities and yet troubling issues around ownership and reproducibility of research.