2018 International Workshop: A Portal for Japanese Old Maps
Ritsumeikan University
Kyoto

WorldMap
HARVARD UNIVERSITY

Infrastructure to Support Collaboration

Ben Lewis
Harvard Center for Geographic Analysis

Center for Geographic Analysis
Harvard University

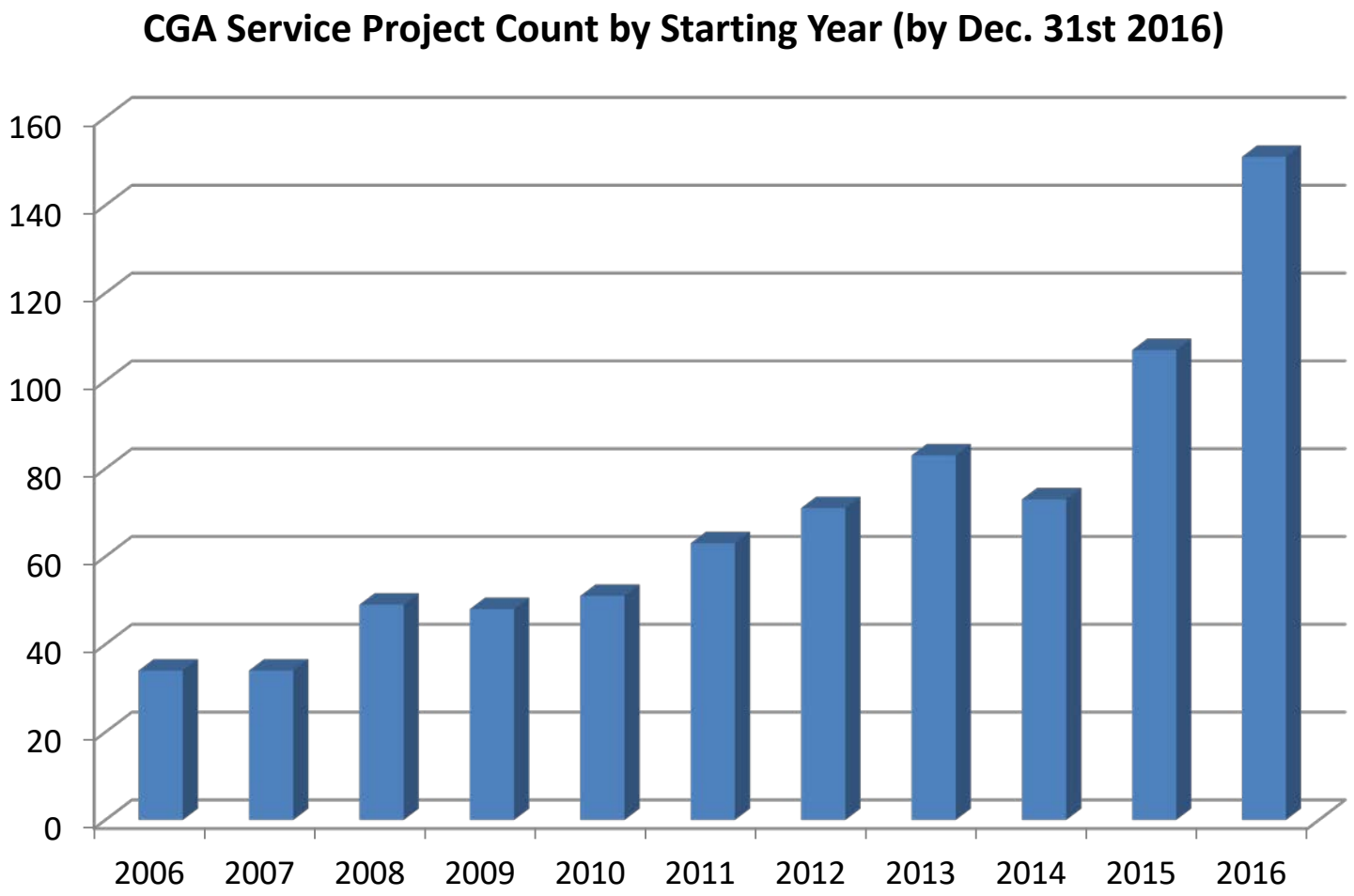
VE RI TAS
HARVARD

The Institute for Quantitative Social Science
at Harvard University

Center for Geographic Analysis (CGA)

- Founded in 2006 by Professor Peter Bol, a historian of medieval China
- Center created to support research in all things Geospatial, and available to scholars across the university, all schools and departments
- Based in the Institute for Quantitative Social Science (IQSS)

Small Team Supports Many Projects



We Develop Platforms to Scale Ourselves

- WorldMap (customized geonode)
- HHypermap (wrote ourselves)
- Warp.worldmap (mapwarper instance)
- Dataverse / WorldMap integration
- ArcGIS Server (installed instance)
- Omeka Neatline (installed instance)
- Billion Object Platform (wrote ourselves)
- MapD (project started at CGA)

WorldMap In a Nutshell

- Designed to lower barriers for researchers who wish to use geospatial technology
- Web-based, cloud hosted
- Made available to the world to use
- Service oriented architecture
- Open source software with a global developer community.

Allows researchers to…

- **Organize:** their own (large) mapping datasets and share them
- **Visualize:** maps with data-driven symbology
- **Publish:** data to the world or to just a few collaborators
- **Discover:** maps by searching across systems
- **Mashup / Combine:** one's own data with data provided by others
- **Collaborate:** by letting several people edit the same map

Open standards-based input

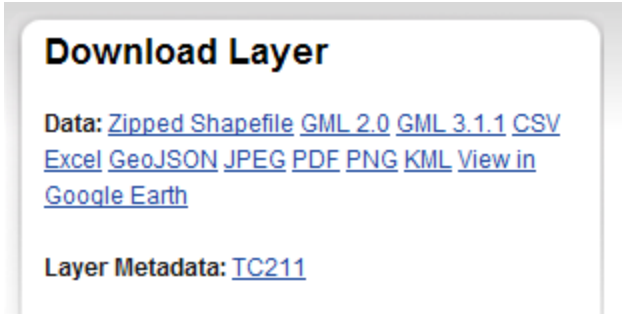
- ESRI REST Image services
- OGC WMS map services
- OGC Catalog services
- ESRI Shapefile, zip compressed or not
- GeoTIFF
- Styled Layer Descriptors (SLD)

Open standards-based output

Service Types:
- WMS
- WFS

Download Formats:
- Zipped Shapefile
- KML
- GML
- GeoJSON
- CSV
- Excel
- JPEG
- PDF
- PNG

Metadata Format:
- ISO TC211



Interoperability with ArcGIS

Database

- Postgres/PostGIS

Web Services

- Esri Open REST
- WMS
- WFS

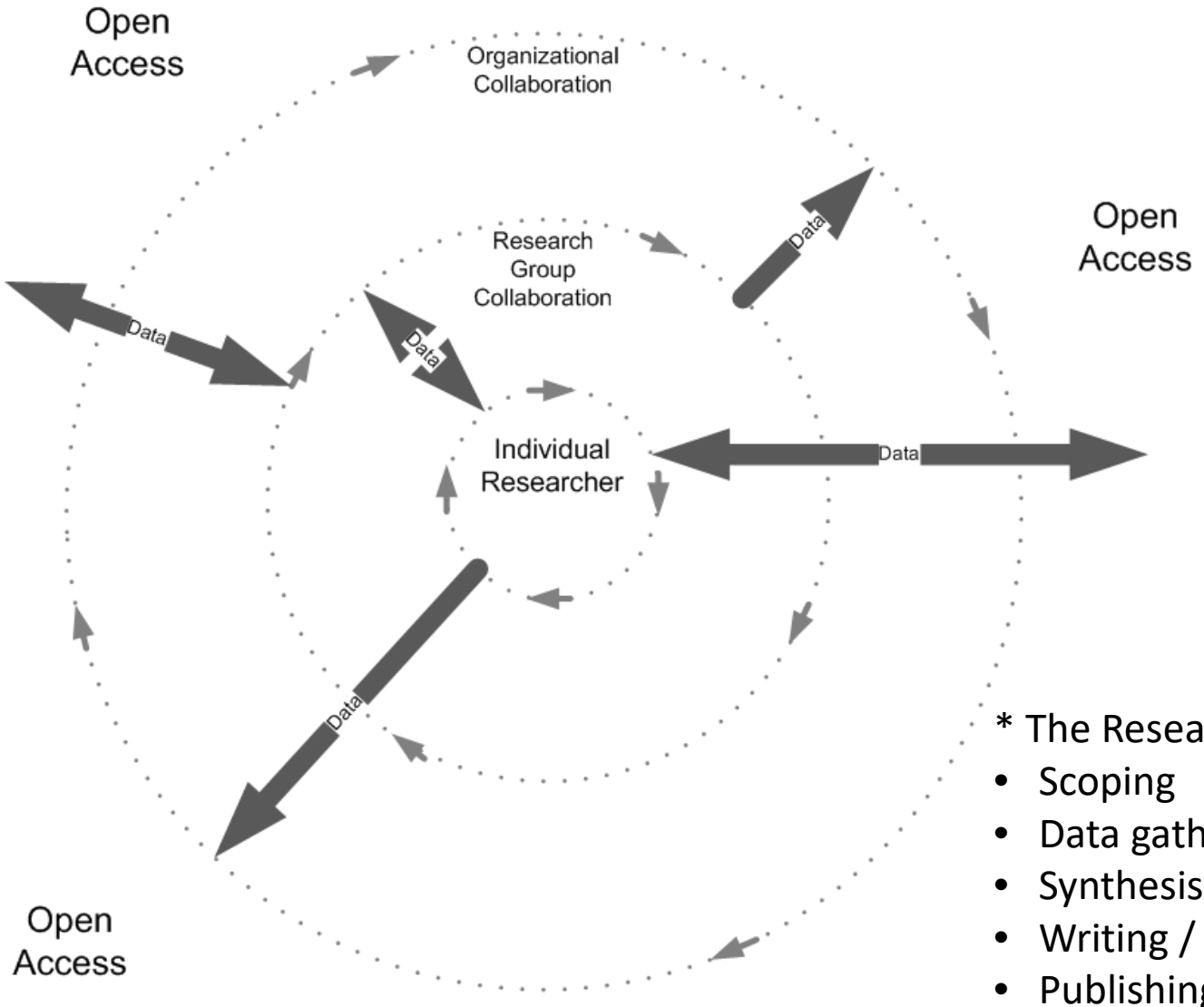
Data Input

- Esri Shapefile
- GeoTIFF

Data Output

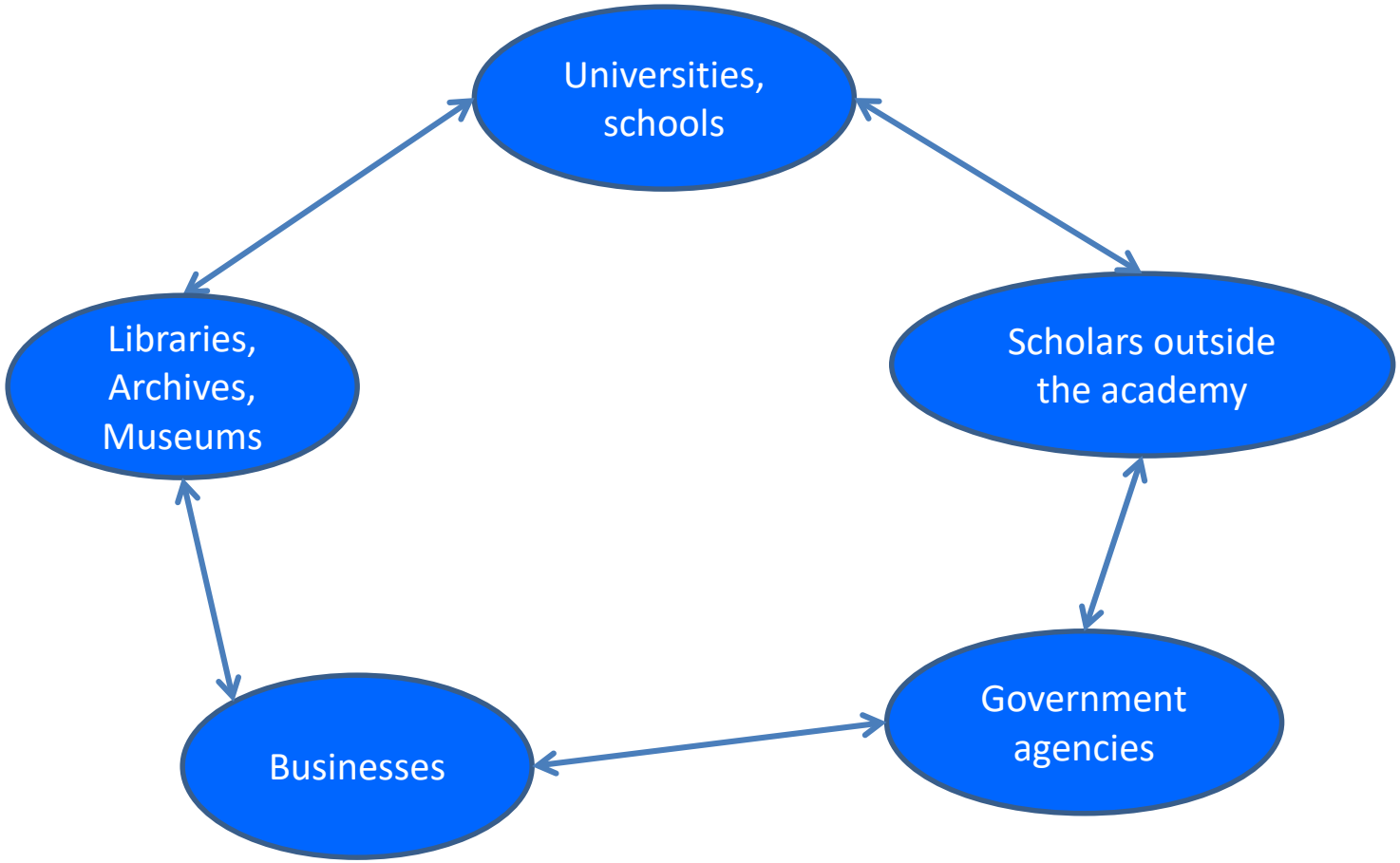
- Esri Shapefile
- JSON
- GeoTIFF
- KML, GML

Open but with Access Controls



* The Research Lifecycle:
- Scoping
- Data gathering / exploration
- Synthesis / analysis
- Writing / communication
- Publishing / delivery

The best data is outside any single organization

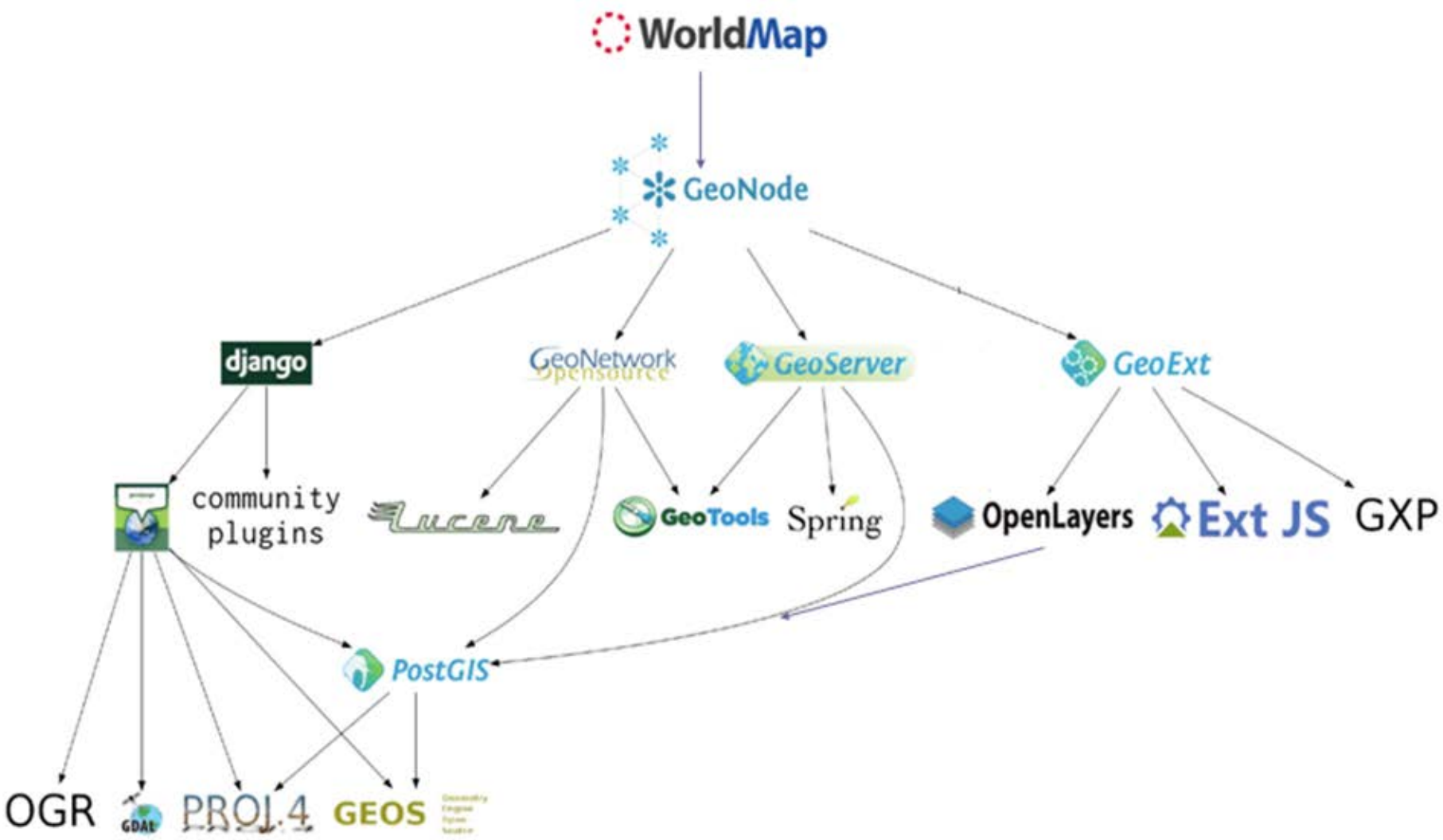


Built on GeoNode

http://geonode.org/

- WorldMap is part of the GeoNode development community which includes:
  - World Bank
  - United Nations World Food Program
  - European Commission
  - U.S. State Department
  - and many other organizations
- WorldMap forked from GeoNode in 2012 and after much work are now reuniting with the main project, having contributed many new features to GeoNode.

WorldMap software stands on many strong shoulders



Contributing Organizations (WorldMap and GeoNode)

WorldMap

- **Ritsumeikan University** – Japanese Old Maps Online
- **Zhejiang University –** Chinese Academic Mapping Platform
- **Boston Area Research Initiative** - BostonMap
- **Cornell University** – Global Health Map
- **UN University** – Wildlife Enforcement Monitoring
- **Virtue Foundation** – Women in the World
- **Amazon** - Hardware
- Others…

GeoNode

- **World Bank** – GFDRR, Dominode, Risiko
- **U.S. State Dept.** – ROGUE, HIU, Syria Damage Assessment
- **NOAA** - GeoCloud
- **UN World Food Program** – WFP Geonode
- **Australian Govt.** - AIFDR, TsuDAT
- **MapStory Foundation** – MapStory App
- Others…

Hosted on a New Public Cloud

- Massachusetts Open Cloud (MOC)
- Led by Boston University, the MOC is a collaborative effort among
  - Boston University
  - Harvard
  - MIT
  - University of Massachusettes
  - Northeastern University
  - Massachusetts Green High-Performance Computing Center (MGHPCC)
  - Oak Ridge National Laboratory (ORNL).

openstack.

WorldMap Stats

- Number of registered users: 23,000
- Number of maps: 5,400
- Number of local layers: 26,000
- Number of remote layers: 90,000
- Number of visitors: 1.5 million

The Problem of Search

- Millions of geo-services exist a click away, but no good way to discover them.
- It’s still like the early days of the web before search engines.

Are Maps a Web or a Haystack?

- Tim Berners Lee chose term “World Wide Web” because hypertext (HTML) can link any page to any other page and create an information web.
- This enabled Google to create a self-curating registry using PageRank (which it built an empire around).
- Web pages point to each other but maps and geospatial datasets don’t.
- So “geospatial haystack” is a better term.
- This is one reason we don’t yet have good search engines for geospatial data.

HHypermap: Toward a solution

- Platform to create, maintain, and deploy a **large registries** of web map services
- Support **search by time and space**
- **Scalable to millions or billions** of objects
- **Open API** to enable any mapping system to find data and use it
- Funded by U.S. National Endowment for the Humanities

Why Focus on Map Service Discovery Rather Than Data/Metadata Discovery?

- Some advantages:
  - Everything in system has a live interactive map behind it
  - All layers can be instantly viewed in context and used at least for display, overlaid, etc.
  - A vast set of layers created by thousands of organizations, containing millions of layers becomes discoverable
  - As layers are discovered and used, and saved in map views, crowd curation can begin, a sort of PageRank for maps.
- Some disadvantages such as:
  - Metadata may be minimal (we are not a library initiative)
  - The data behind the service may and may not be available
  - Maps which are not yet georeferenced are not included

Ways to populate a registry

- Crawl web to find them – returns many endpoints but quality varies greatly
- User submissions – fewer but endpoints tend to be more reliable

Crawling the web

- OGC Services
  - Look for "?request getcapabilities" and not “test” in the href URL
- ESRI Rest Services
  - Look for “/arcgis/rest/services” in the target-DOMAIN-URI of the WARC Response Header text
- KML or KMZ files
  - Look for an href URL ending in .kml or .kmz files
- Compressed shapefiles
  - Look for “shape” or “shp” and string ending with “.zip” in the href URL
- Tile Servers
  - Look for “tile” or “tiles” and string ending with “.png” in the href URL

Some of the larger known catalogs

- David Rumsey Collection
- Various mapwarpers
- New York Public Library Collection
- Geoblacklight - Platform developed by Stanford and other universities to provide fast search access to geospatial library holdings.
- OpenGeoPortal – Platform developed by Tufts and other universities to provide fast search access to geospatial library holdings.
- ArcGIS Open Data – Esri collection of 44,000 open datasets and growing.
- Geodata.gov, Geoplatform.gov – The U.S. Federal government has built a data sharing platform for U.S. data using CKAN software and ArcGIS Online.
- INSPIRE Geoportal – Spatial data portal of the European Commission.
- GEOSS registry – Group on Earth Observations registry of 850 map service collections.
- Geopole.org – CSW catalogue service providing access to 400,000 layers.
- Geonetwork – Geospatial catalogue maintained by the Food and Agriculture Organization of the United Nations.
- Spatineo.com – Commercial service which is currently monitoring 40,000 web services containing 899,000 layers.
- Many CKAN portals
- Many Thredds servers

Improving search experience through visualization of distributions

- Temporal dimension offers huge opportunity to improve search
- Opportunity to provide visibility into the spatial distributions of collections

Taking advantage of the time dimension

- Most spatial data describes events in time though often not explicitly.
- When data does have a time component it is often not easily accessed.
- An opportunity exists to improve search by:
  1. Making latent temporal information explicit using enrichment techniques
  2. Implementing UI/backend enhancements on existing systems

TimeMiner – to enrich metadata

- Temporal metadata for geospatial datasets is often weak.
- In a crowd-sourced data repository, data creators and contributors often do not create detailed metadata.
- Many data sets have temporal properties, but time is often ambiguously defined, mentioned as unstructured text in the title, abstract, and elsewhere.
- Time is often not referred to using a standard date/time format such as ISO 8601, but as descriptive text.

Initial simple TimeMiner Logic

1. Look for date in the date range section of the metadata and choose the earlier date. (Date: from Metadata)
2. If there is no #1 above, look for 4 digit numbers in title first, then abstract, which are less than or equal to 2018 (present year) (Date: Detected)
3. If there IS a date in #2 above, check to see whether there is a CE or AD or BCE or BC after it and apply math accordingly (Date: Detected)
4. If there IS NO #2 above, look for 1, 2, or 3 digit numbers with associated CE, AD, BCE, BC, and apply math accordingly (Date: Detected)

Another simple technique: historic periods Example: Chinese dynasty names

- Xia, Hsia — ca. 2100-1600 BCE
- Shang — ca. 1600-1050 BCE
- Zhou, Chou — ca. 1046-256 BCE
- Qin, Ch'in — 221-206 BCE
- Han — 206 BCE-220 CE
- Sui — 581-618 CE
- Tang, T'ang — 618-906
- Song, Sung — 960-1279
- Yuan — 1279-1368
- Ming — 1368-1644
- Qing, Ch'ing — 1644-1912

Source: http://afe.easia.columbia.edu/timelines/china_timeline.htm

Perhaps a similar approach could be used for Japanese Periods

538–710	Classical Japan	Asuka		Imperial government
710–794		Nara		
794–1185		Heian		
1185–1333	Medieval Japan	Kamakura		Kamakura shogunate
1333–1336		Kenmu Restoration		Imperial government
1336–1392		Muromachi	Nanboku-chō period	Ashikaga shogunate
1392–1467				
1467–1573			Sengoku period	Ashikaga shogunate and *sengoku daimyōs*
1573–1603		Azuchi–Momoyama		Oda Nobunaga, Toyotomi Hideyoshi and Tokugawa Ieyasu
1603–1868	Early Modern Japan	Edo	Tokugawa period	Tokugawa shogunate
1868–1912	Modern Japan	Meiji	Pre-war	Imperial government
1912–1926		Taishō		
1926–1945		Shōwa (Prewar)		
1945–1952	Contemporary Japan	Shōwa (Occupied Post-war)	Post-war	GHQ/SCAP
1952–1989		Shōwa (Post-occupation)		Parliamentary democracy
1989–present		Heisei		

Source: https://en.wikipedia.org/wiki/History_of_Japan

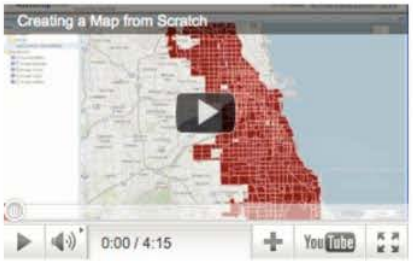
WorldMap

blewis | Log out | Create Map | View Map | Help | Admin | GeoServer

Create a Map | View a Map | About

ChinaMap

Build your own mapping portal and publish it to the world or to just a few collaborators. WorldMap is open source software.



Watch the WorldMap Quick Start video

Create a Map | View a Map | About

ChinaMap

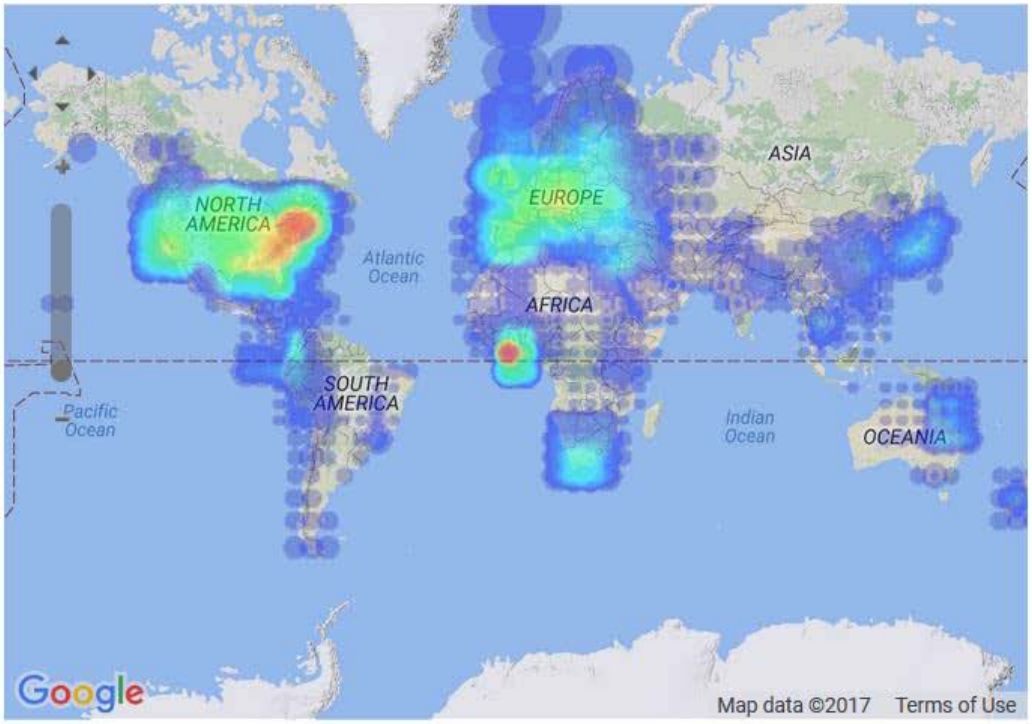
Build your own mapping portal and publish it to the world or to just a few collaborators. WorldMap is open source software.

Search BETA

SEARCH | Upload Layer | Create Layer | Rectify Layer | Submit a Map Service

Keyword | Source | All Layers | Search | Reset

from year 5000M BCE to year Future



Title	Source	Date
IndexLetIdent	gis.icao.int	2016
OSM: 4000 un-tagged probable…	worldmap.ha…	None
asianame	gis.icao.int	2016
Major World Watersheds	water.discom…	2016
Major World Watersheds	water.discom…	2016
Major World Watersheds	water.discom…	2007
WRI Major Watersheds of the …	worldmap.ha…	2016
{ERS_port_nb}	maratlas.dis…	2016
00	maratlas.dis…	2016
{ERS_port_nb}	maratlas.dis…	2016
657 Crude Oil Refineries - Retr…	worldmap.ha…	2006
Oil Refineries from IndustryAbout	worldmap.ha…	2016
AGEAR	gis.icao.int	2016
AGEAR	gis.icao.int	2016

Prev | Next | Showing 1-200 of 112557

No Layers Selected

Clear Selected | Add To Map

Enter search... | Search | Rese

Google

2000 km
1000 mi

1 : 139770641

Center for Geographic Analysis

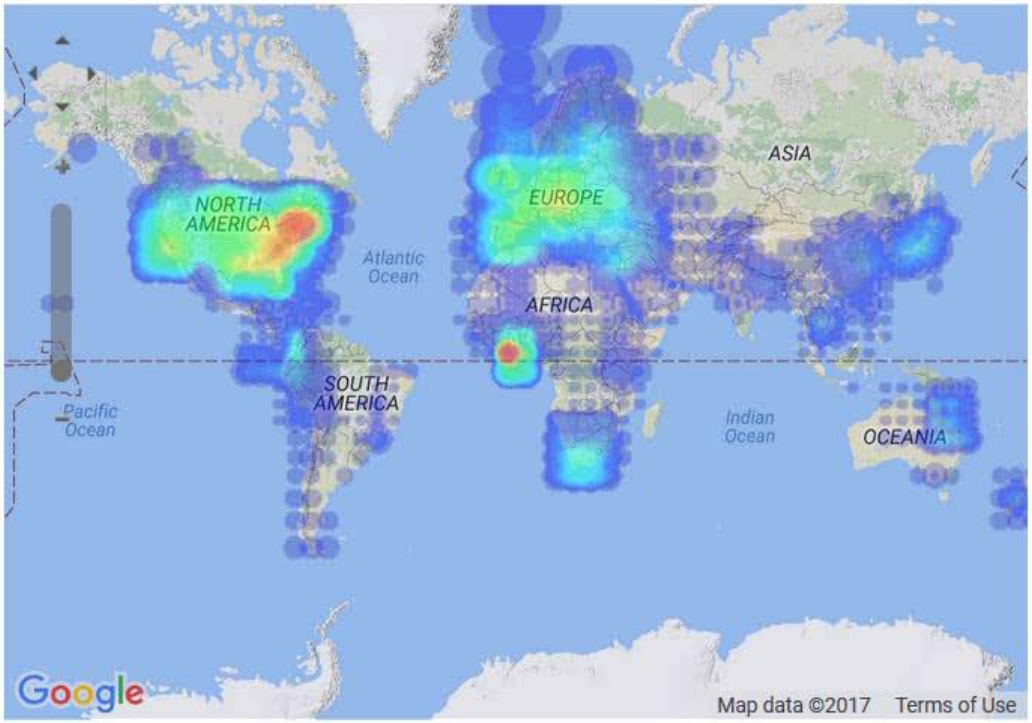
Map data ©2017 Terms of Use

Search BETA

SEARCH | Upload Layer | Create Layer | Rectify Layer | Submit a Map Service

Keyword | Source | All Layers | Search | Reset

from year 5000M BCE to year Future



Title	Source	Date
IndexLetIdent	gis.icao.int	2016
OSM: 4000 un-tagged probable...	worldmap.ha...	None
asianame	gis.icao.int	2016
Major World Watersheds	water.discom...	2016
Major World Watersheds	water.discom...	2016
Major World Watersheds	water.discom...	2007
WRI Major Watersheds of the ...	worldmap.ha...	2016
{ERS_port_nb}	maratlas.dis...	2016
00	maratlas.dis...	2016
{ERS_port_nb}	maratlas.dis...	2016
657 Crude Oil Refineries - Retr...	worldmap.ha...	2006
Oil Refineries from IndustryAbout	worldmap.ha...	2016
AGEAR	gis.icao.int	2016
AGEAR	gis.icao.int	2016

Prev | Next | Showing 1-200 of 112557

No Layers Selected

Clear Selected | Add To Map

Enter search... | Search | Rese

2000 km
1000 mi
1 : 139770641

Center for Geographic Analysis

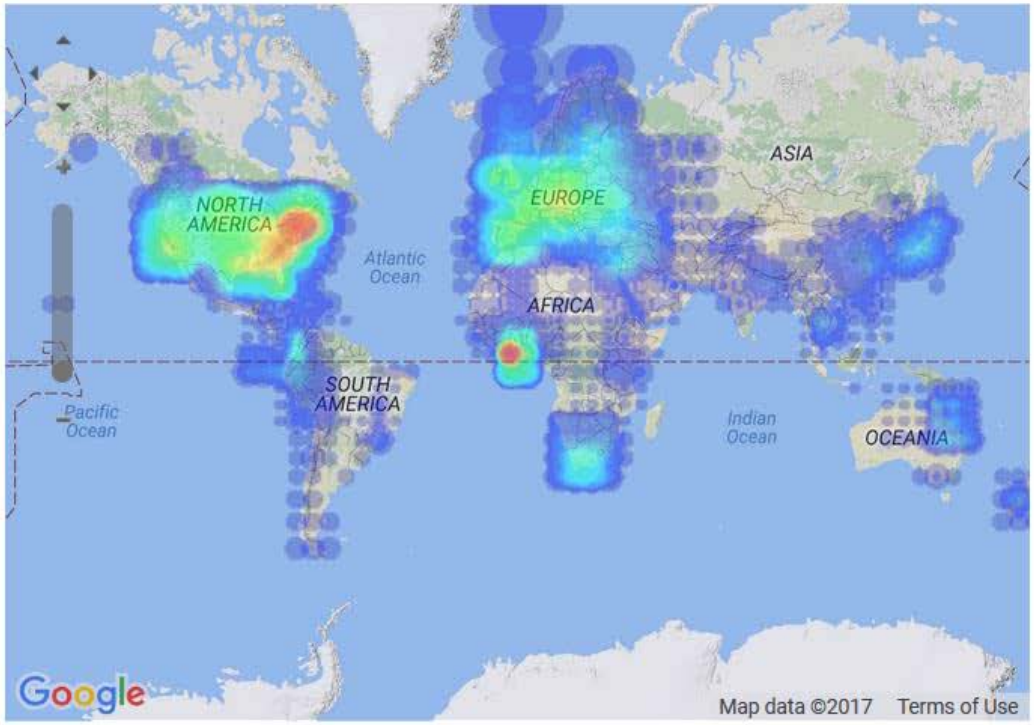
Map data ©2017 Terms of Use

Search BETA

SEARCH | Upload Layer | Create Layer | Rectify Layer | Submit a Map Service

Keyword | Source | All Layers | Search | Reset

from year 5000M BCE to year Future



Title	Source	Dat...
Champlain's map of New Fran...	maps.nypl.org	1632
Russia Isaac Massa 1632	mapwarper.net	1632
Poland Lithuania 1635	warp.worldm...	1635
America / Jodocus Hondius ex...	maps.nypl.org	1635
Terra Firma et Novum Regnum...	mapwarper.net	1635
Bern, Switzerland 1638 Merian	warp.worldm...	1638
America Septentrionalis.	maps.nypl.org	1639
Vingboons map of Manhattan, ...	maps.nypl.org	1639
Nova Anglia, Novum Belgium, e...	maps.nypl.org	1639
Heda1640s	worldmap.ha...	1640
Kalf 1640s	worldmap.ha...	1640
WillemHeda1640	worldmap.ha...	1640
Claesz1640s	worldmap.ha...	1640
DeHeem1640s	worldmap.ha...	1640

Prev | Next | Showing 1001-1200 of 112557

No Layers Selected

Clear Selected | Add To Map

Enter search... | Search | Rese

Google

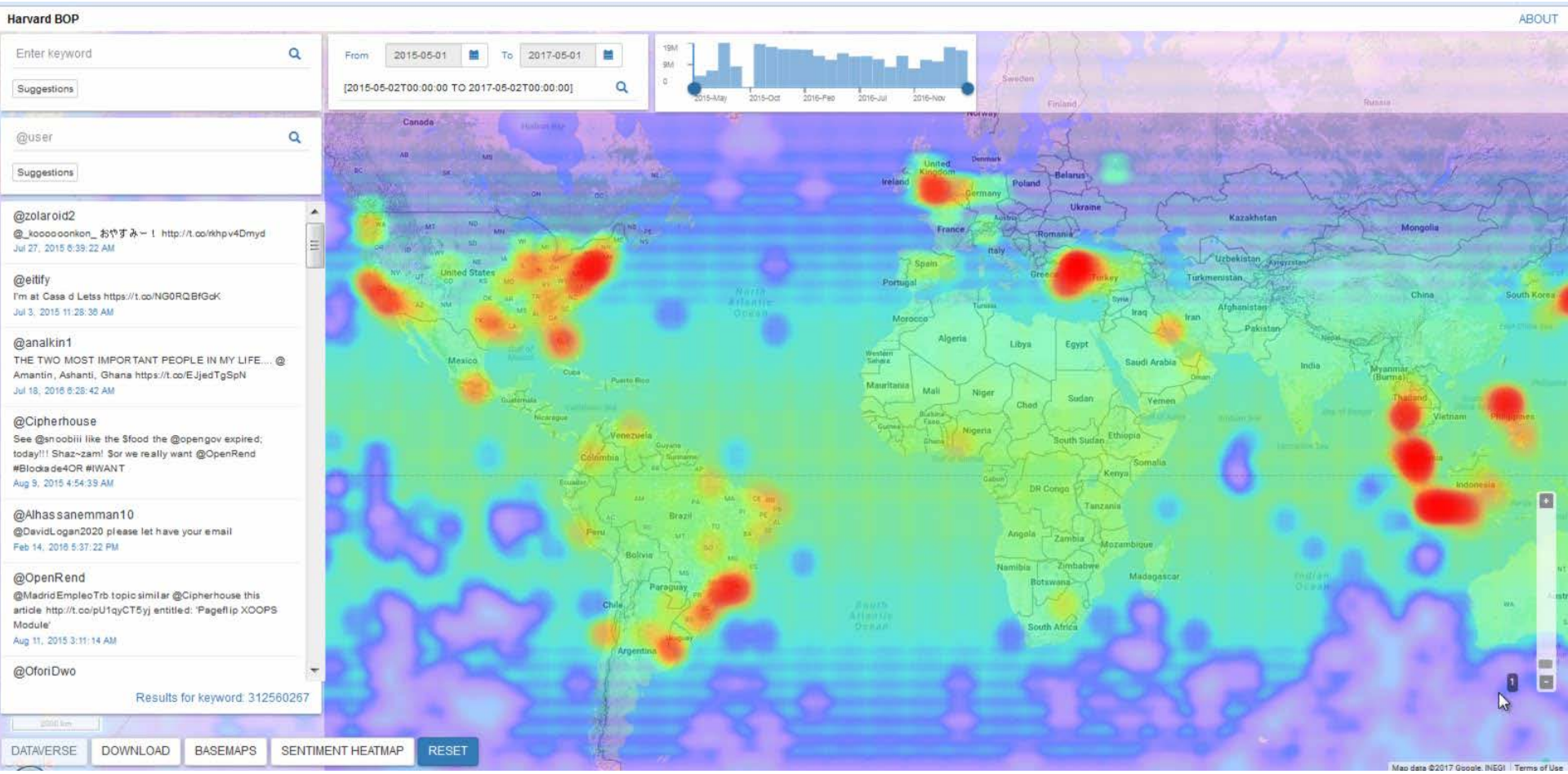
2000 km
1000 mi

1 : 139770641

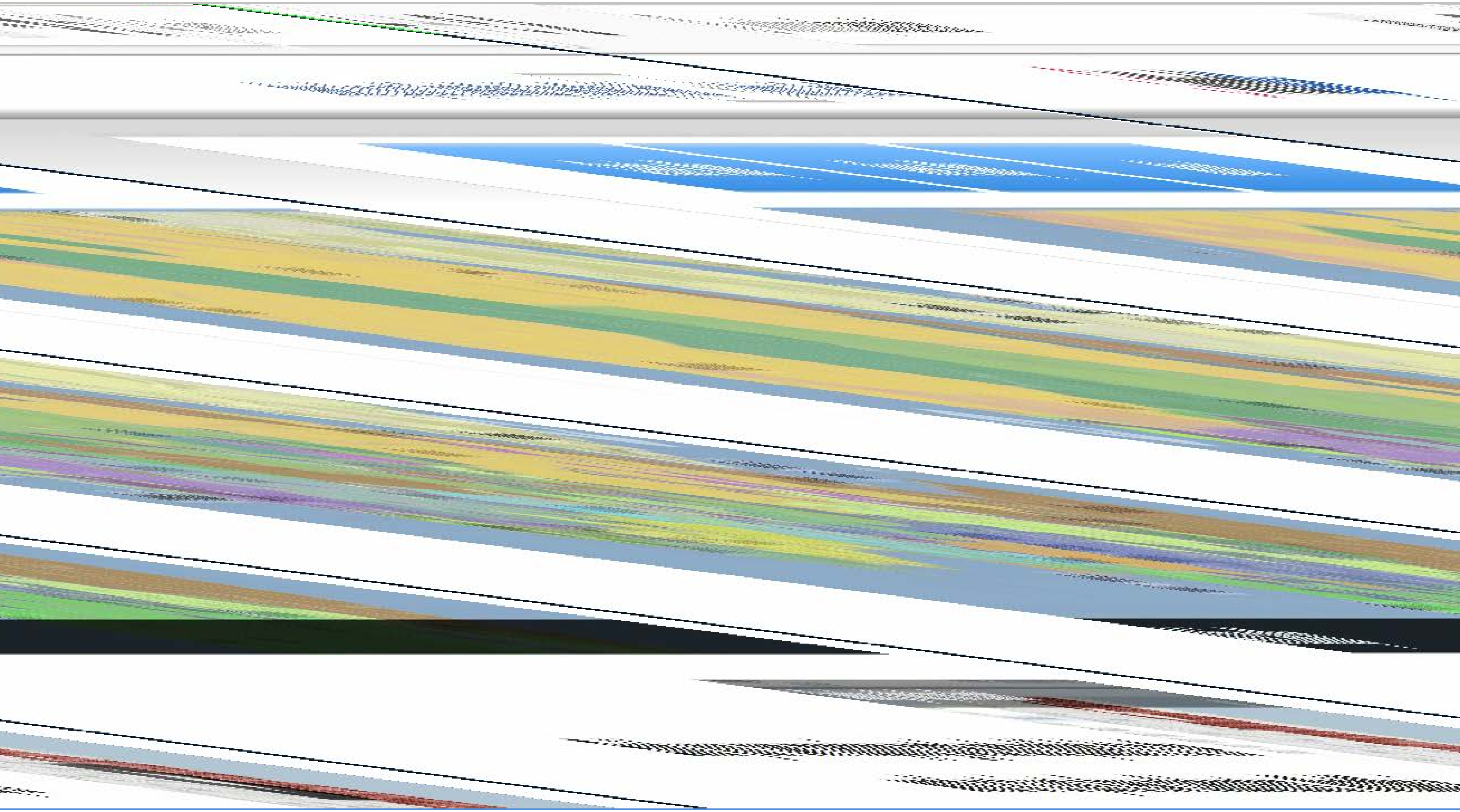
Center for Geographic Analysis

Map data ©2017 Terms of Use

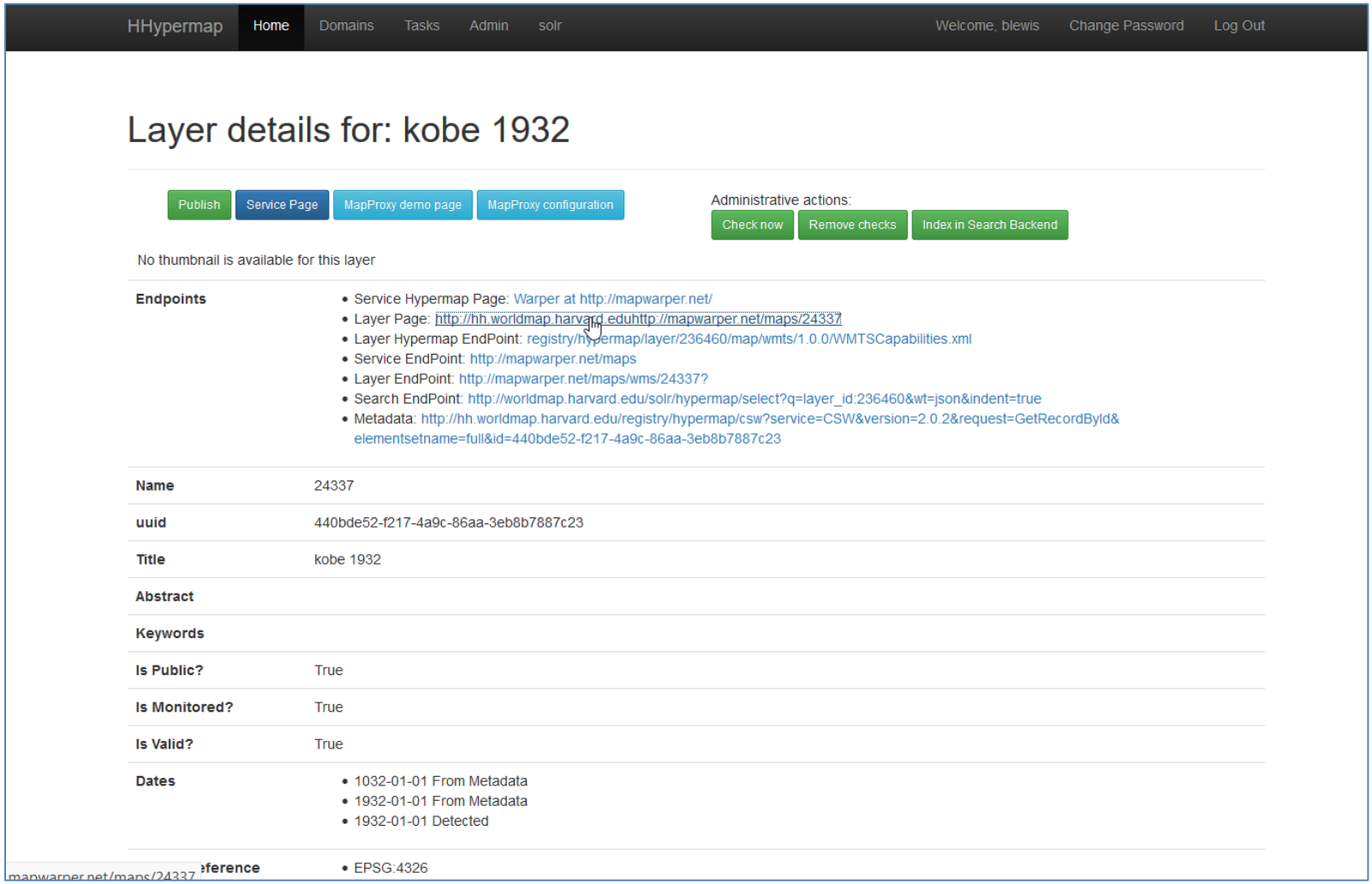
BOP Brexit


Harvard BOP
ABOUT
Enter keyword
Suggestions
From 2015-05-01 To 2017-05-01
[2015-05-02T00:00:00 TO 2017-05-02T00:00:00]
@user
Suggestions
@zolaroid2
@_kooooonkon_ おやすみー！ http://t.co/rkhpv4Dmyd
Jul 27, 2015 6:39:22 AM
@eitify
I'm at Casa d Letss https://t.co/NG0RQBfGcK
Jul 3, 2015 11:28:38 AM
@analkin1
THE TWO MOST IMPORTANT PEOPLE IN MY LIFE.... @ Amantin, Ashanti, Ghana https://t.co/EJjedTgSpN
Jul 18, 2015 6:28:42 AM
@Cipherhouse
See @snoobiii like the $food the @opengov expired; today!!! Shaz~zam! $or we really want @OpenRend #Blockade4OR #IWANT
Aug 9, 2015 4:54:39 AM
@Alhassanemman10
@DavidLogan2020 please let have your email
Feb 14, 2016 5:37:22 PM
@OpenRend
@MadridEmpleoTrb topic similar @Cipherhouse this article http://t.co/pU1qyCT5yj entitled: 'Pageflip XOOPS Module'
Aug 11, 2015 3:11:14 AM
@OforiDwo
Results for keyword: 312560267
DATAVERSE
DOWNLOAD
BASEMAPS
SENTIMENT HEATMAP
RESET

HHypermap for Exploring Specific Collections



Layer Page for Kobe 1932

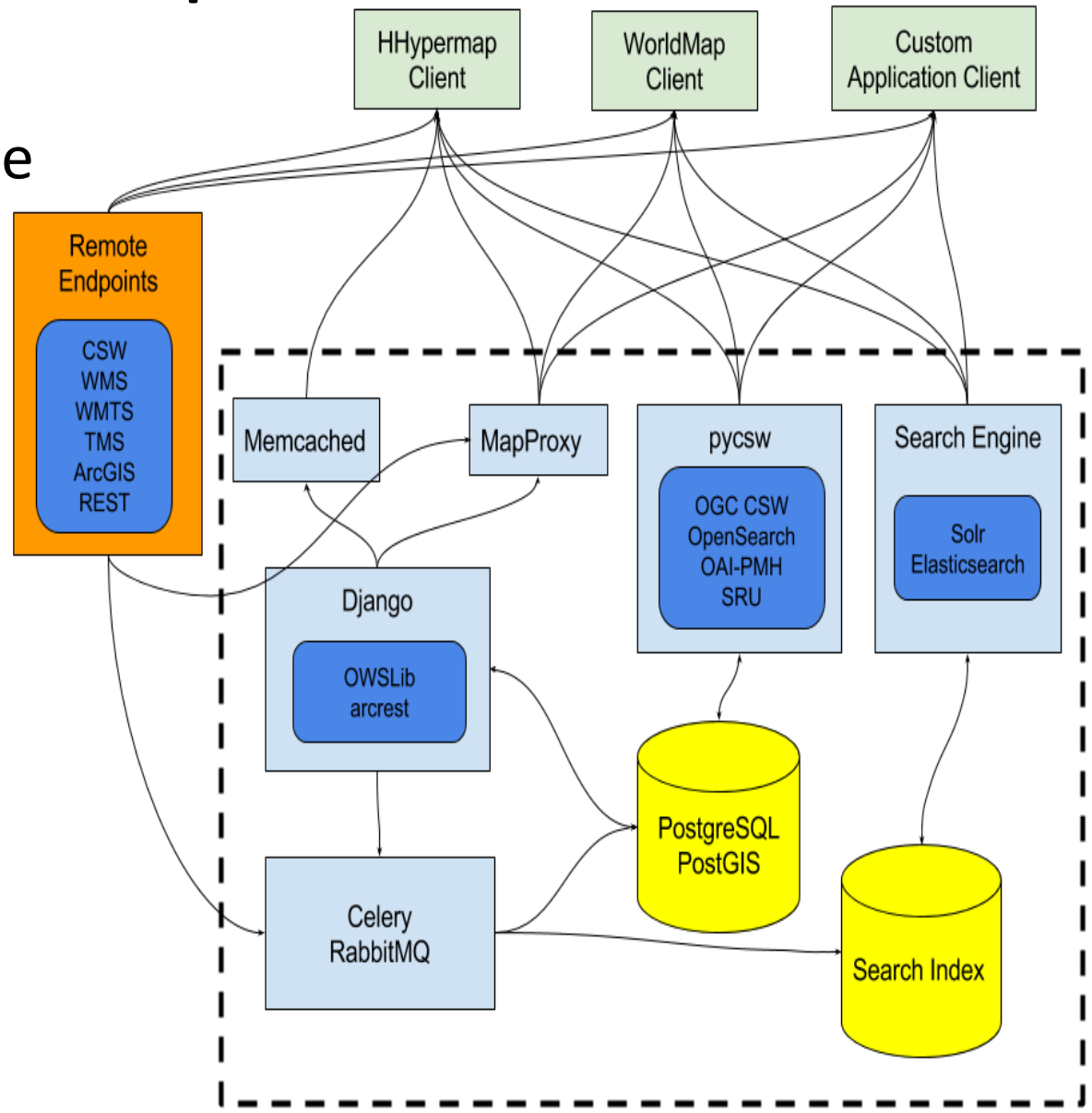


From Layer page one can get to source page on remote server mapwarper.net

HHypermap Architecture

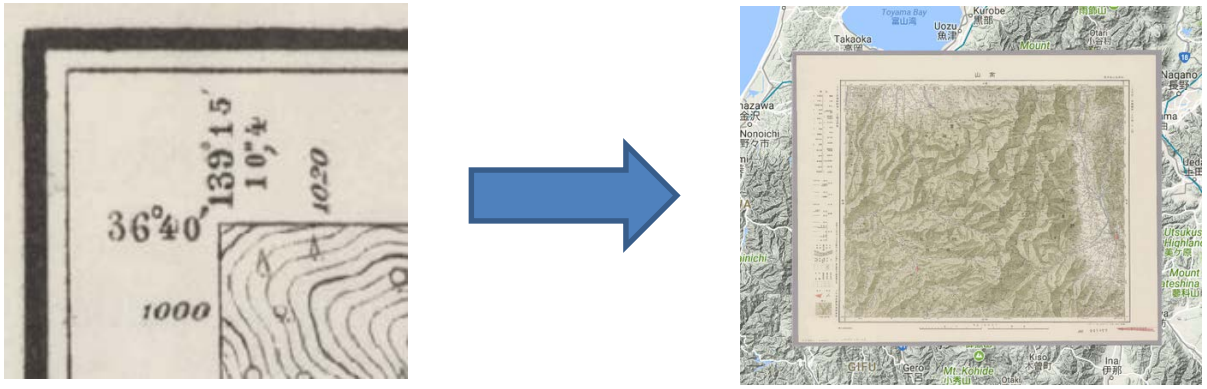
Built on open source software:

- Celery
- RabbitMQ
- Django
- Lucene
  - Solr
- MapProxy
- Memcached
- OWSLib
- PostgreSQL
- PostGIS
- pycsw



Automatic Georeferencing

- These 1300 high quality layers from Stanford were georeferenced automatically using software written by Ryo Kamata at Ritsumeikan U. on the mapwarper platform.



- This is a great example of how ungeoreferenced maps can flow through an online enrichment pipeline. One can imagine other forms such as place name extraction.
- This approach has application to the vast collections of large scale paper map series held by map libraries around the world.

A couple other projects

Dataverse <–> WorldMap Integration

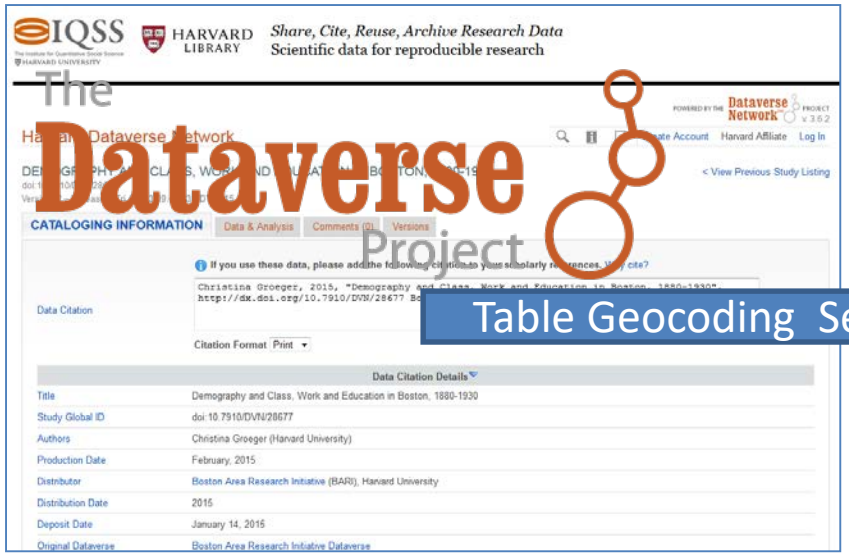
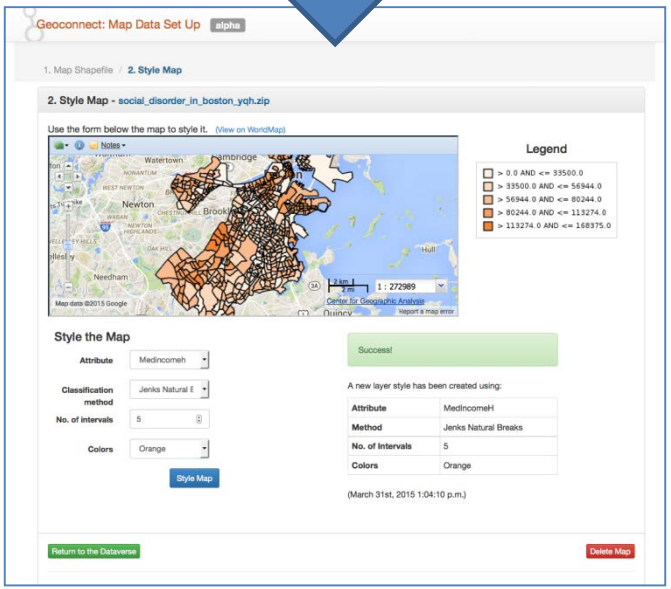


Table Geocoding Service

WorldMap



- Dataverse is industrial-strength archive for social science data.
- WorldMap is scalable geovisualization platform.
- Connect them to let researchers visualize their data spatially without leaving the social science platform.

Spatial GPU Database Project

- MapD is a database 1000 times faster than a standard RDBMS.
- MapD with origins at CGA in 2012, is now a successful startup.
- MapD has open sourced the GPU database and is working with CGA to extend it.
- The project is happening within the Spatio-temporal Innovation Center composed of Harvard, George Mason, and U.C. Santa Barbara.
- We will focus initially on testing hydrologic models which have very large outputs.

MAPD

In Conclusion

- The amount of historical spatial information on the web is large and continues to grow
- But it is not well handled by traditional search engines
- It is not possible to perform a quick search and find a particular dataset (or determine one does not exist)
- This lack of discoverability remains a major challenge for our community but is one that is solvable if we continue to work together on it.

Some Links…

- Center for Geographic Analysis
  - http://gis.harvard.edu
- WorldMap
  - http://worldmap.Harvard.edu
  - https://github.com/cga-harvard/worldmap
- HHypermap
  - http://hhypermap.worldmap.harvard.edu
  - https://github.com/cga-harvard/HHypermap
- Dataverse
  - https://dataverse.harvard.edu/
  - https://github.com/IQSS/dataverse
- Open Geoportal http://opengeoportal.org
- Geoblacklight http://geoblacklight.org/

Thank you

Benjamin Lewis
blewis@cga.harvard.edu

Center for
Geographic Analysis

Harvard University

http://cga.harvard.edu
http://worldmap.harvard.edu