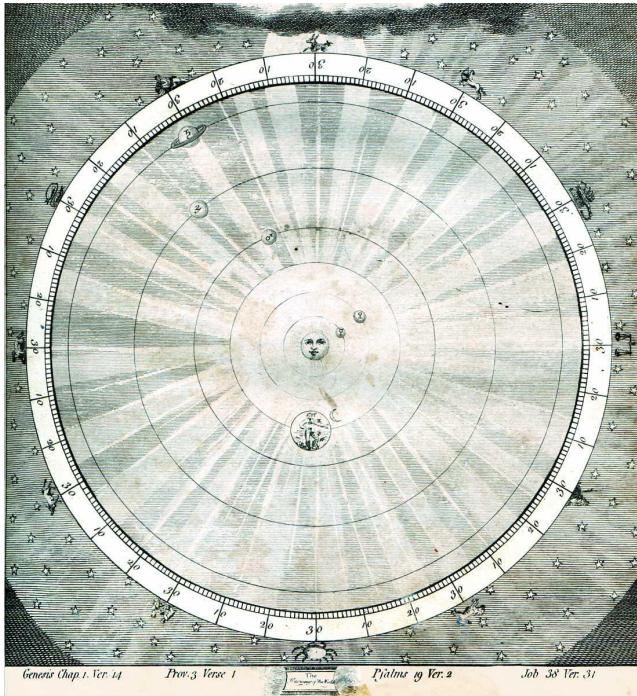


Challenges in Harmonizing Geospatial Metadata



Karen Majewicz
University of Minnesota
United States

Kimberly Durante
Stanford University
United States

Andrew Battista
New York University
United States

Taylor Hixson
New York University, Abu Dhabi
United Arab Emirates

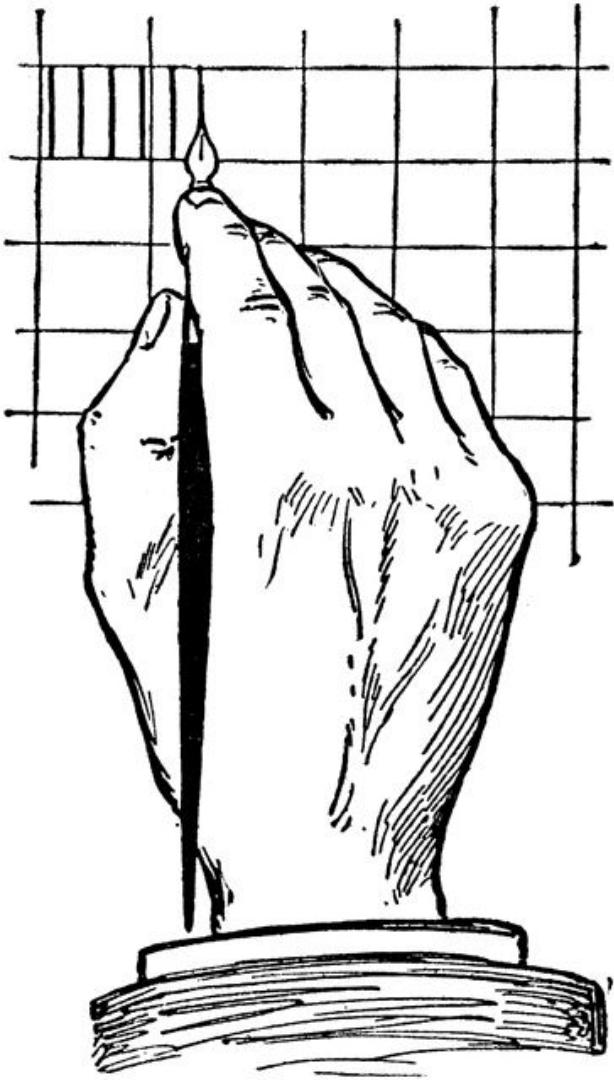
Melinda Kernik - moderator
University of Minnesota
United States

IASSIST & Carto 2018
Friday, June 1, 2018
9:00-10:15



Outline of presentations

1. Overview of Geospatial Metadata Challenges
2. Development of a Geospatial Discovery System
3. Harmonizing Geospatial Metadata: A Use Case
4. Digital Colonialism?



Overview of Geospatial Metadata Challenges

Karen Majewicz
majew030@umn.edu



What are some important elements of geospatial metadata?

DESCRIPTIVE & ADMINISTRATIVE

- Titles, descriptions, subjects, dates
- Usage constraints
- Authors, publishers

CONTENT

- Database field types
- Field definitions

SPATIAL

- Spatial reference system
- Bounding box extents
- Geometry type

DATA QUALITY

- Accuracy
- Omissions
- Lineage and processing history



How is geospatial metadata created?

Sample Workflow for Data Curator

1

Gather information
that can't be
determined, especially
about data quality,
attribute table
definitions, from data
creator

2

Query the dataset to
extract spatial
information

3

Batch insert templated
information

4

Manually write the
remaining descriptive
information



Workflow: What are some of the barriers?

- Many geospatial resources do not have any metadata
- Not seen as important throughout much of the professional GIS community. Labor outweighs benefits.
- Lack of overlapping skill-sets: little training in metadata for GIS specialists



Workflow: What are possible improvements?

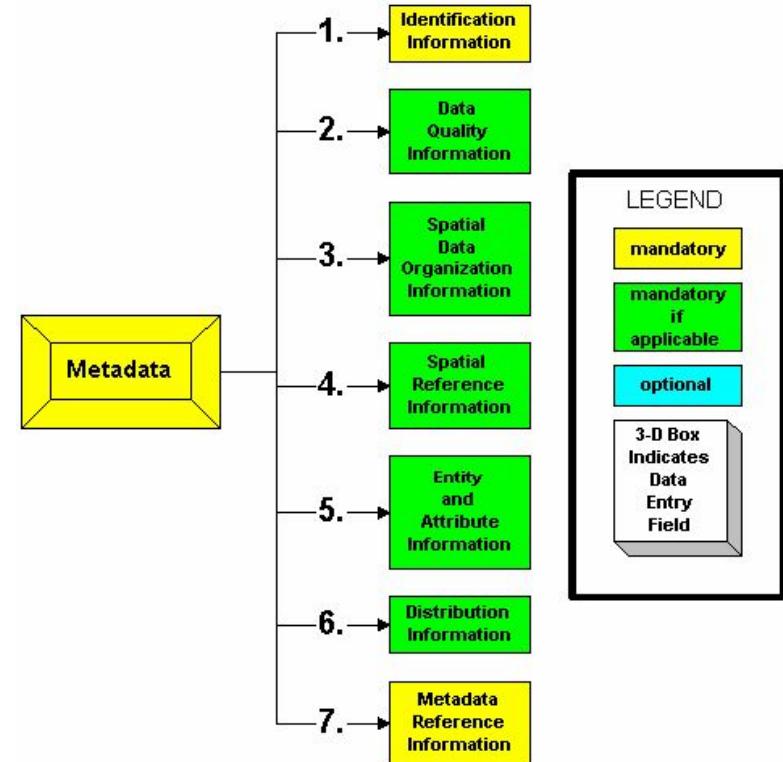
- Encourage data producers to practice good data management and document key information as it is gathered and created
- Perform outreach to demonstrate the value of metadata for organizations in terms of long term investments that will save labor and costs in the future



What are the primary geospatial metadata standards?

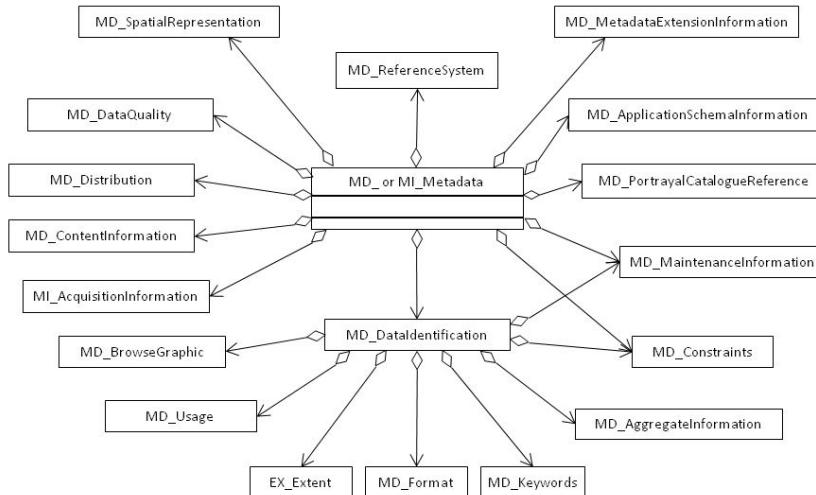
FGDC CSDGM

- clear, easy to understand structure
- developed in the 1990s, before web services and geoportal infrastructures
- still widely used in the USA, especially by state and local governments





What are the primary geospatial metadata standards?



ISO Metadata Objects

ISO 191xx series

- flexible structure
- splits repeatable metadata containers, such as coordinate system or attribute table definitions into separate files
- mandatory for federal data in the USA, commonly used internationally



Standards: What are some of the challenges?

- The standards are described by many metadata authors as complex and arduous. Some agencies abandon them for simpler standards.
- Both standards are expressed in XML format, which can be easily corrupted by missing tags or nesting errors
- Difficulty of ISO/FGDC to easily transform to indexed discovery metadata



Flexibility of ISO Standard

DISTRIBUTION INFO

→ Distribution

→ **Transfer Options**

DISTRIBUTION INFO

→ Distribution

→ **Distributor**

→ Distributor Transfer Options

Digital Transfer Options
Online Resource
Linkage

Digital Transfer Options
Online Resource
Linkage



What are some tools for creating geospatial metadata?

Dedicated, out-of-the-box applications:

- Proprietary: ArcCatalog
- Open Source: GeoNetwork, CatMDEdit
- New & proposed developments: mdEditor.org, plugins for QGIS



Tools: What features are lacking?

- Automatic extraction of spatial metadata
- Easy batch editing (find and replace, CSV updates)
- Customized crosswalking between formats and standards



Tools: What are alternatives?

- Hire a developer to create customized editors that interact with preservation and discovery systems (ex. GéolIndex+)
- Rely on scripting (ex. Python, Ruby) for batch transformations and editing



The role of Open Data Portals for Public Data

The screenshot shows the Government of Canada Open Government Portal. It features a top navigation bar with links for Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, and More services. A search bar is present, along with a "Français" link. Below the header, there's a breadcrumb trail: Home > Open Government > Search Open Government. The main content area displays a search bar and a results summary: "3,253 records found Order by: Relevance".

CKAN

The CKAN screenshot shows a search results page for "Cadastral Information". It lists three datasets: "Cadastral Information for UTM Zone 7 Northwest Territories", "Cadastral Information for Taku Site", and "Cadastral Information for Village Island 1". Each dataset has a brief description, resource formats (e.g., KML, Shapefile), and a "View Details" button.

Socrata

The Socrata screenshot shows a search results page for "Street Sweeping". It lists three datasets: "Street Sweeping - 2017 - Map", "Street Sweeping - 2018 - Map", and "Street Sweeping - 2013 - Map". Each dataset includes a map preview, a brief description, update information (e.g., "Updated April 13, 2017"), and a "View Details" button. A sidebar on the left lists categories like Administration & Finance, Buildings, Community & Economic Development, Education, Environment & Sustainable Development, Sanitation, and View Types.

The ArcGIS Hub screenshot shows a map titled "Health India" focusing on South Asia. The map highlights India with a blue overlay representing "Recycling Service Areas". The ArcGIS interface includes a search bar, user navigation links (ArcGIS Hub, Community, Initiatives, Open Data), and a legend. Below the map, there are tabs for Overview, Data, and API Explorer.

About

Open Data
Shared By: Vaibhav.Singh
Data Source: esrinid-
coe.nitt-tech.in

Custom License 3/29/2018 Spatial Dataset 693 Rows
This map provides details of District-Wise Health Care Infrastructure for India 2015 and 2016, Percentage of women received Janani Suraksha Yojana (JSY) benefits, Personal habit of Age 15 and above, Quality of Family Planning services. Reported prevalence of Morbidity and Unmet need for
More ▾



Portals: What are some of the problems?

- Data portals don't always feature the option to include standards metadata.
- Ephemeral Data
 - Priority given to most up to date version of data
 - Why devote time to documenting data that will soon be purged?



Portals: What are better practices?

- Choose geoportal applications that incorporate external metadata and codebooks
- Archive public data into stable spatial repositories



Discovery Metadata

- Data portals use discovery metadata that is indexed for searching
- Discovery metadata can be harvested by other portals

```
// http://maps-semcog.opendata.arcgis.com/data.json
{
  "@context": "https://project-open-data.cio.gov/v1.1/schema/catalog.jsonld",
  "@type": "dcat:Catalog",
  "conformsTo": "https://project-open-data.cio.gov/v1.1/schema",
  "describedBy": "https://project-open-data.cio.gov/v1.1/schema/catalog.json",
  "dataset": [
    {
      "@type": "dcat:Dataset",
      "identifier": "http://maps-semcog.opendata.arcgis.com/datasets/ff498b1c85bd4b49aa730eb7a8c8c683_0",
      "title": "ParkData",
      "description": "Tabular data containing park amenities.",
      "keyword": [
        "land"
      ],
      "issued": "2018-05-11T12:01:05.000Z",
      "modified": "2018-05-11T12:06:07.000Z",
      "publisher": {
        "name": "Southeast Michigan Council of Governments"
      },
      "contactPoint": {
        "@type": "vcard>Contact",
        "fn": "SEMCOG Staff",
        "hasEmail": "mailto:infocenter@semcog.org"
      },
      "accesesLevel": "public",
      "distribution": [
        {
          "@type": "dcat:Distribution",
          "title": "ArcGIS Open Dataset",
          "format": "Web page",
          "mediaType": "text/html",
          "accessURL": "http://maps-semcog.opendata.arcgis.com/datasets/ff498b1c85bd4b49aa730eb7a8c8c683_0"
        },
        {
          "@type": "dcat:Distribution",
          "title": "Esri Rest API",
          "format": "Esri REST",
          "mediaType": "application/json",
          "accessURL": "https://services1.arcgis.com/xUx8EjNc6egUPYWh/arcgis/rest/services/ParkData/FeatureServer/0"
        },
        {
          "@type": "dcat:Distribution",
          "title": "GeoJSON",
          "format": "GeoJSON",
          "mediaType": "application/vnd.geo+json",
          "downloadURL": "http://maps-semcog.opendata.arcgis.com/datasets/ff498b1c85bd4b49aa730eb7a8c8c683_0.geojson"
        }
      ]
    }
  ]
}
```



Discovery Metadata: What are some of the impediments?

- Consistency problems in spellings and terms
- Metadata not written for external users (no place name given)
- Variances in local practices: reliance upon acronyms, custom schemas



Discovery Metadata: Possible Solutions

- Develop robust metadata application profiles with entry guidelines
- Implement validation scripting into the workflow to verify consistency
- Use controlled vocabularies
- Encourage sharing discovery metadata between institutions for collaborative remediation

Development of a Geospatial Data Discovery System

Kim Durante
Stanford University Libraries
IASSIST & Carto 2018

Issues - Data Discovery and Access

- Data are hard to find
- Links are ephemeral
- Rights are indeterminable
- Metadata often missing or incomplete

SearchWorks catalog

All fields

books & media



Library services ▾

Advanced search Course reserves Selections (0) ▾

Back to results 1 of 974 Next →

Cite Send to Select

ESRI data & maps [electronic resource]

RESPONSIBILITY

Environmental Systems Research Institute, Inc.

IMPRINT

[Redlands, CA] : The Institute, 1996.

PHYSICAL DESCRIPTION

1 CD-ROM ; 4 3/4 in.



Earth Sciences Library (Branner)

Today's hours: 4p - 9p

Reference

G70.212 .E77 1996

In-library use

More options

Find it at other libraries via WorldCat

Creators/Contributors

CONTRIBUTOR

Environmental Systems Research Institute (Redlands, Calif.)

Contents/Summary

SUMMARY

Contains map data and many scales of geography for the world, Mexico, Canada, as well as general and detailed data for the United States.

Subjects

SUBJECT

ArcView.

Maps > Databases.

- areacode.shp
- blkgrp.shp
- blockpop.shp
- cbsa.shp
- cd113.shp
- cities.shp
- cities_dtl.shp
- counties.shp
- dtl_cnty.shp
- dtl_cnty_ln.shp
- dtl_st.shp
- dtl_st_ln.shp
- placeply.shp
- places.shp
- senate.shp
- states.shp
- tracts.shp
- urban.shp
- urban_dtl.shp

areacode
shapefile

Thumbnail
Not
Available

Tags

There are no tags for this item.

Summary

There is no summary for this item.

Description

There is no description for this item.

Credits

There are no credits for this item.

Use limitations

There are no access and use limitations for this item.

Extent

West -178.227822 **East** -65.244128
North 71.390482 **South** 17.926706

Scale Range

There is no scale range for this item.

ArcGIS Metadata ►

Topics and Keywords ►

* CONTENT TYPE Downloadable Data

Hide Topics and Keywords ▲

Citation ►

* TITLE areacode

PRESENTATION FORMATS *digital map

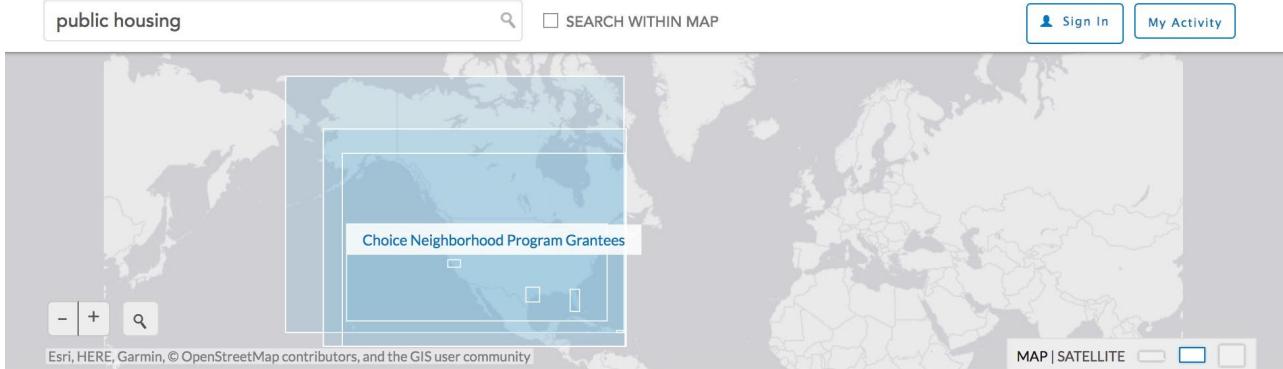
Hide Citation ▲

Resource Details ►

DATASET LANGUAGES * English (UNITED STATES)

**HUD.GOV**U.S. Department of Housing and Urban Development
Secretary Ben Carson

Search HUDUser

[eGIS Home](#)[eGIS Applications ▾](#)[eGIS Data ▾](#)[eGIS API and Data Documentation](#)

1–10 of 18 results

Relevance ▾

[Share](#)

104 attributes | 3527 locations |

**Public Housing Authorities** (from Open Data – Rental Assistance Programs)Shared by [HUD.Official.Content](#)

Public Housing was established to provide decent and safe rental housing for eligible low-income families, the elderly, and persons with disabilities. Public housing comes in all sizes and types, from scattered single family houses to high-rise apartments for elderly families. There are approximately 1.2 million households living in public housing units, managed by over 3,300 housing agencies (...)

My Favorites

click on any dataset to add to your favorites

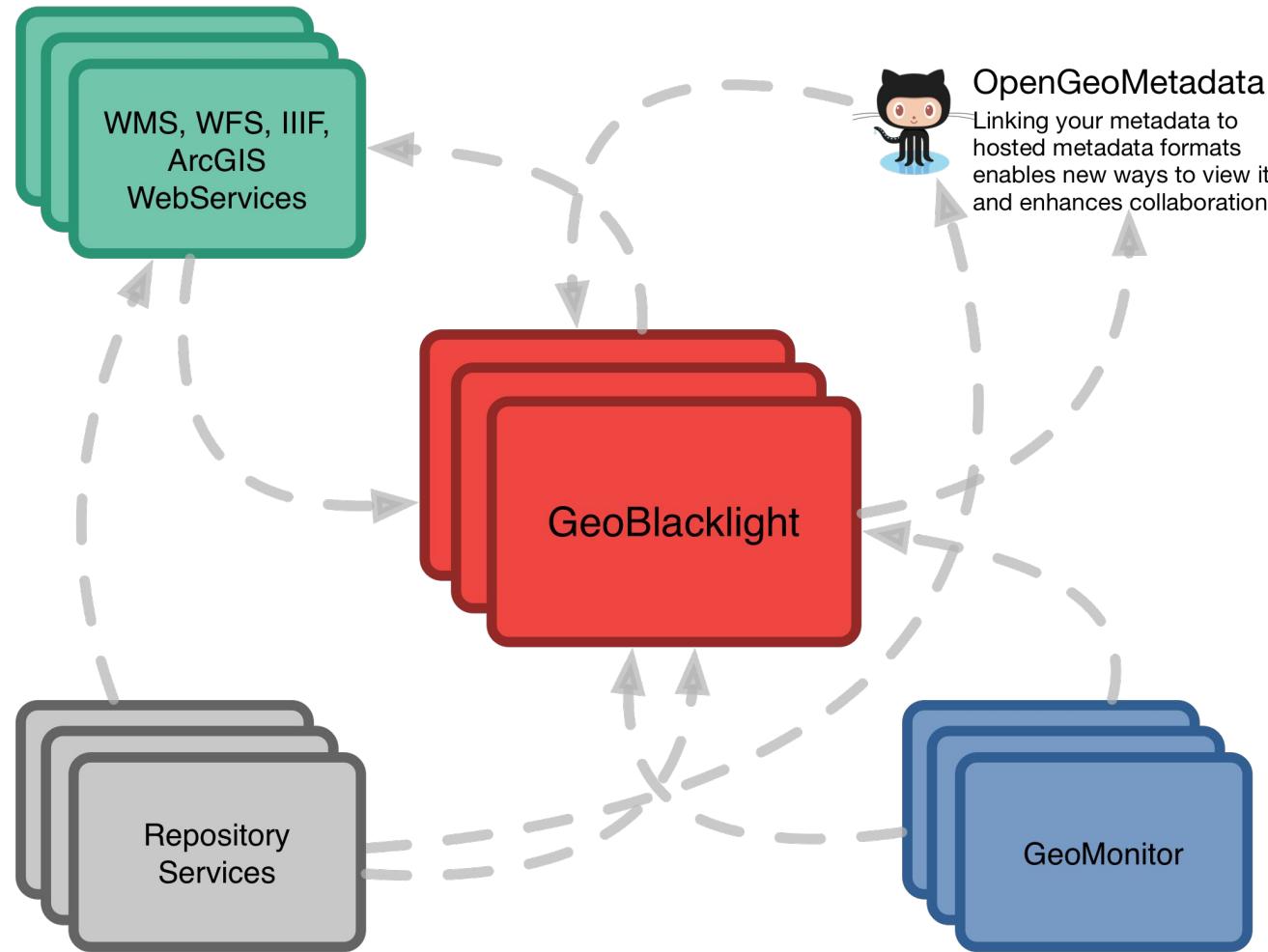
146 attributes | 201921 locations |

**Public Housing Buildings** (from Open Data – Rental Assistance Programs)Shared by [HUD.Official.Content](#)

HUD administers Federal aid to local Housing Agencies (HAs) that manage housing for low-income residents at rents they can afford. Likewise, HUD furnishes technical and professional assistance in planning, developing, and managing the buildings that comprise low-income housing developments. This dataset provides the location, and resident characteristics of public housing d...

138 attributes | 6618 locations |

**Public Housing Developments** (from Open Data – Rental Assistance Programs)Shared by [HUD.Official.Content](#)



public housing

Search 

Limit your search

Institution Author Publisher 

United States. Department of
Housing and Urban Development  54

Subject Place 

United States  54

Year Access Data type 

Polygon  54

You searched for: public housing  Publisher > United States. Department of Housing and Urban Development [Start Over](#)Place > United States Data type > Polygon Bounding Box > -117.949219 15.45368 -80.771484 49.95122 [« Previous](#) | **1 - 10 of 54** | [Next »](#)[Sort by relevance](#) 10 per page [Save this search](#)

- ▶ 1. [Housing Choice Vouchers by Tract, 2016](#)  
- ▶ 2. [ACS 5 Year Housing Data by State, 2008-2012](#)  
- ▶ 3. [ACS 5 Year Housing Data by Place, 2008-2012](#)  
- ▶ 4. [ACS 5 Year Housing Data by County, 2008-2012](#)  
- ▶ 5. [ACS 5 Year Housing Data by Tract, 2008-2012](#)  
- ▶ 6. [Choice Grantees, 2016](#)  
- ▶ 7. [Fair Market Rents, 2016](#)  
- ▶ 8. [CDBG Activity by Tract, 1996-2016](#)  
- ▶ 9. [HOME Activity by Tract, 2016](#)  



[« Previous](#) | **26 of 66** | [Next »](#)[Back to Search](#) [Start Over](#)

Fair Market Rents, 2016

Author(s) United States. Department of Housing and Urban Development

Description Fair Market Rents (FMRs) are primarily used to determine payment standard amounts for the Housing Choice Voucher program. FMRs are also used to determine the initial renewal rents for some expiring project-based Section 8 contracts, initial rents for housing assistance payment (HAP) contracts in the Moderate Rehabilitation Single Room Occupancy program (Mod Rehab), and to serve as a rent ceiling in the HOME Investment Partnership Program (HOME) for rental assistance. HUD annually estimates FMRs for 530 metropolitan areas and 2,045 nonmetropolitan county FMR areas. By law the final FMRs for use in any Fiscal Year must be published and available for use at the start of that Fiscal Year.

[Read More](#)

Publisher United States. Department of Housing and Urban Development

Place(s) [United States](#)

Subject(s) [Housing, Housing Choice Voucher Program \(U.S.\)](#), [Boundaries](#), and [Society](#)

Year 2016

Held by [Stanford](#)

More details at <http://purl.stanford.edu/jx876vw5031>

Use and reproduce... This item is in the public domain. There are no restrictions on use.

Copyright This work is in the Public Domain, meaning that it is not subject to copyright.



Attribute	Value
fid	jx876vw5031.2110
objectid	512
fmr_0bdr	1412
fmr_1bdr	1814
fmr_2bdr	2289
fmr_3bdr	2987
fmr_4bdr	3556
fmr_areana	San Francisco, CA HUD Metro FMR Area
fmr_code	METRO41860MM73
fmr_4plusb	3556

Tools

[Bookmark](#)

 [Email](#)

[Web services](#)

 [Metadata](#)

[Download Shapefile](#) ▾

 [Open in Carto](#)

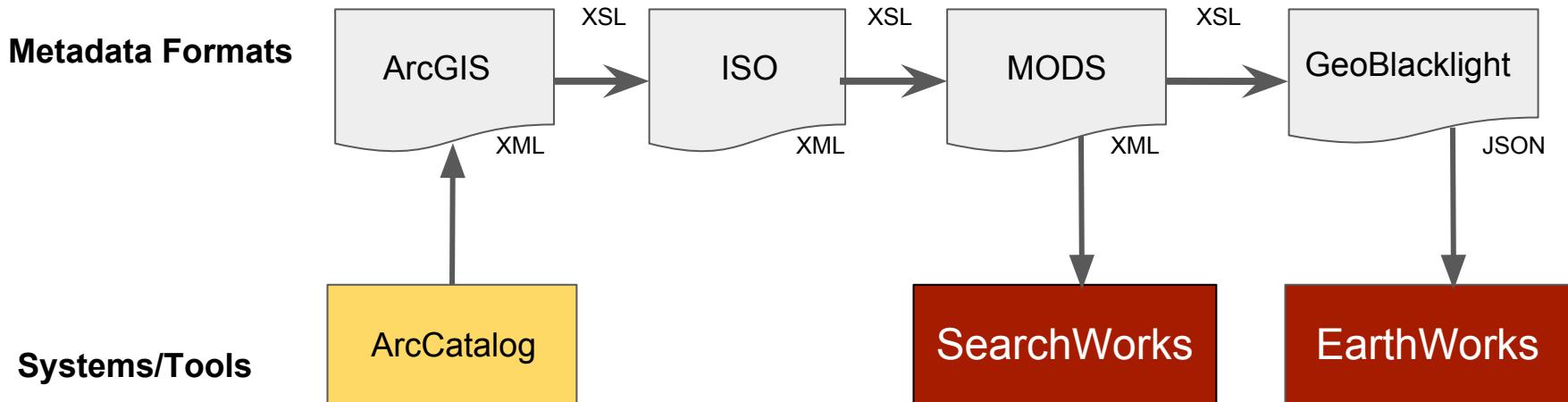
GeoBlacklight Metadata Schema

Title	Fair Market Rents, 2016
Description	Fair Market Rents (FMRs) are primarily used to determine payment standard amounts for the Housing Choice Voucher program ...
Publisher	United States. Department of Housing and Urban Development
Date issued	2016
Spatial	United States
Subject	Housing
Temporal	2016
Geometry type	Polygon
Geometry	ENVELOPE (-178.227822, 146.154418, 71.390482, -14.601813)
Rights	Public

GeoBlacklight Metadata Schema

Identifier	http://purl.stanford.edu/jx876vw5031
Layer ID	druid:jx876vw5031
Layer Slug	stanford-jx876vw5031
Download URL	http://stacks.stanford.edu/file/druid:jx876vw5031/data.zip
MODS Metadata	http://purl.stanford.edu/jx876vw5031.mods
ISO Metadata	https://raw.githubusercontent.com/OpenGeoMetadata/edu.stanford.purl/master/jx/876/vw/5031/iso19139.xml
Web Mapping Service	https://geowebervices.stanford.edu/geoserver/wms

Metadata Workflow



Stanford Digital Repository

Fair Market Rents, 2016

PREFERRED CITATION

United States. Department of Housing and Urban Development. (2016). Fair Market Rents, 2016.
United States. Department of Housing and Urban Development. Available at
<http://purl.stanford.edu/jx876vw5031>.

COLLECTION

U.S. Department of Housing and Urban Development Maps and GIS Data



Access conditions

USE AND REPRODUCTION

This item is in the public domain. There are no restrictions on use.

COPYRIGHT

This work is in the Public Domain, meaning that it is not subject to copyright.

Description

[Back to results](#) [Send to](#) [Select](#)

U.S. Department of Housing and Urban Development Maps and GIS Data [electronic resource]

IMPRINT Washington, D.C.: United States. Department of Housing and Urban Development,

[View all 66 items](#)

- [ACS 5 Year CHAS Data by County, 2008-2011 \[2016\]](#)
United States. Department of Housing and Urban Development
- [ACS 5 Year CHAS Data by Place, 2008-2012 \[2016\]](#)
United States. Department of Housing and Urban Development
- [ACS 5 Year CHAS Data by State, 2008-2012 \[2016\]](#)
United States. Department of Housing and Urban Development

[View all 66 items >](#)

Collection details

DIGITAL CONTENT
[66 items](#)

COLLECTION PURL
<https://purl.stanford.edu/wc590wy7435>

Creators/Contributors

CORPORATE AUTHOR

[United States. Department of Housing and Urban Development.](#)

Subjects

SUBJECT

[Housing > United States > Maps](#)
[Grants-in-aid > United States > Maps](#)
[Community development > United States > Maps](#)

GENRE

[Geospatial data](#)
[Geographic information systems data](#)

Encoding Best Practices

- Titles and Descriptions - Brief/full summarizations of the data
- Spatial Keywords - Library of Congress Authorities, GeoNames
- Thematic Keywords - LC Authorities, ISO Topic Categories
- Personal and Organizational Names - LC Authorities
- Rights - Creative Commons, Open Data Commons Open Database License
- URLs and References - Stable, persistent, relevant

Automating Workflows

- Metadata transformation (XSLT)
- XML metadata creation and enhancement (Python)
- Validation (Python)
- Data management (Python)

Open GeoMetadata

Columbia University

Cornell University

Harvard University

Massachusetts Institute of Technology

New York University

Princeton University

Tufts University

University of California at Berkeley

University of Pennsylvania

University of Virginia

Virginia Tech

Big Ten Academic Alliance



<https://github.com/OpenGeoMetadata>

Aggregating Metadata

FGDC

ISO 19139

MODS

OpenGeoportal Schema

GeoBlacklight Metadata Schema

Institution		
B	Baruch CUNY	126
UCB	Berkeley	351
C	Columbia	3,265
H	Harvard	10,448
I	Illinois	129
Psi	Indiana	291
Globe	Iowa	343
MIT	MIT	7,166
Maryland	Maryland	32
MassGIS	MassGIS	599
M	Michigan	802
Michigan State	Michigan State	43
M	Minnesota	624
NYU	NYU	3,502
Princeton	Princeton	28,872
P	Purdue	235
S	Stanford	10,606
T	Tufts	3,092
UVa	UVa	1,256
W	Wisconsin	750
« Previous Next »		
A-Z Sort		Numerical Sort

 Search or jump to... / Pull requests Issues Marketplace Explore

OpenGeoMetadata / edu.nyu

Code Issues 1 Pull requests 2 Projects 0 Wiki Insights

Branch: master edu.nyu / handle / 2451 / 3 / 38 / 76 / Create

sgbalogh adds dspace internal ids

..

geoblacklight.json adds dspace internal ids

 Search or jump to... / Pull requests Issues Marketplace

OpenGeoMetadata / edu.stanford.purl

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Branch: master edu.stanford.purl / bh / 042 / xf / 3440 /

drh-stanford upgrade to GeoBlacklight Schema 1.0

..

geoblacklight.json upgrade to GeoBlacklight Schema 1.0

iso19110.xml latest export

iso19139.html regenerated HTML rendering of ISO 19139 metadata

iso19139.xml latest export

mods.xml latest export

preview.jpg latest export

 Search or jump to... / Pull requests Issues Marketplace

OpenGeoMetadata / big-ten

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Branch: master big-ten / datasets / 02a-01 / 03 / 80 / 56 / 28d8dc4ff5978bef555ce4477c /

KarenMajewicz Added Illinois Geospatial Data Clearinghouse (illinois.edu)

..

geoblacklight.json Added Illinois Geospatial Data Clearinghouse (illinois.edu)

iso19139.xml Added Illinois Geospatial Data Clearinghouse (illinois.edu)

 Search or jump to... / Pull requests Issues Marketplace Explore

OpenGeoMetadata / edu.cornell

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Watch 11 ★

Branch: master edu.cornell / 00 / 16 / 26 / Create new file Upload file

kgjenkins snapshot 2017-12-15 Latest commit 7c

..

fgdc.xml snapshot 2017-12-15

geoblacklight.json snapshot 2017-12-15



Search or jump to...

Pull requests Issues Marketplace Explore



OpenGeoMetadata / GeoCombine

[Unwatch](#) 14[Star](#) 15[Fork](#) 10[Code](#)[Issues 9](#)[Pull requests 3](#)[Projects 0](#)[Wiki](#)[Insights](#)[Settings](#)

A Ruby toolkit for managing geospatial metadata [https://github.com/OpenGeoMetadata/Geo...](https://github.com/OpenGeoMetadata/GeoCombine)

[Edit](#)[Add topics](#)[153 commits](#)[7 branches](#)[9 releases](#)[7 contributors](#)

Branch: master ▾

[New pull request](#)[Create new file](#)[Upload files](#)[Find file](#)[Clone or download](#) ▾

mejackreed Merge pull request #83 from OpenGeoMetadata/faster-file-search ...

Latest commit 934e86f on Jan 10

bin	Add command line interface using thor	3 years ago
lib	Merge pull request #83 from OpenGeoMetadata/faster-file-search	5 months ago
spec	add upgrade_to_v1 logic to convert pre_v1 to v1 schema	6 months ago
.coveralls.yml	add coveralls.yml	3 years ago
.gitignore	initialize gem	3 years ago
.rspec	adds in metadata transformations	3 years ago
.travis.yml	Update rubies to current supported versions	10 months ago
Gemfile	improve slugification logic	6 months ago
LICENSE.txt	initialize gem	3 years ago
README.md	add repo as argument for geocombine:clone/pull and use clone_url rath...	5 months ago
Rakefile	adds in metadata transformations	3 years ago
geo_combine.gemspec	use net-http-persistent for geocombine:index task	5 months ago

Harmonizing Geospatial Metadata: A Use Case

IASSIST & Carto 2018 - Montreal

Slides: <http://tiny.cc/iassist2018>

Andrew Battista
Librarian for Geospatial Information Systems
New York University - ab6137@nyu.edu





GeoBlacklight Metadata



OpenGeoMetadata

Simplified metadata schema to facilitate data discovery

JSON records that incorporate Dublin Core elements, plus some elements for spatial component

Easier to author from scratch; hard to transform from existing ISO 19131 or FGDC standards

See “[A Metadata Schema for Geospatial Resource Discovery Use Cases](#)”



General Challenges with Consortial Metadata



[GeoBlacklight Metadata Application Profile](#)

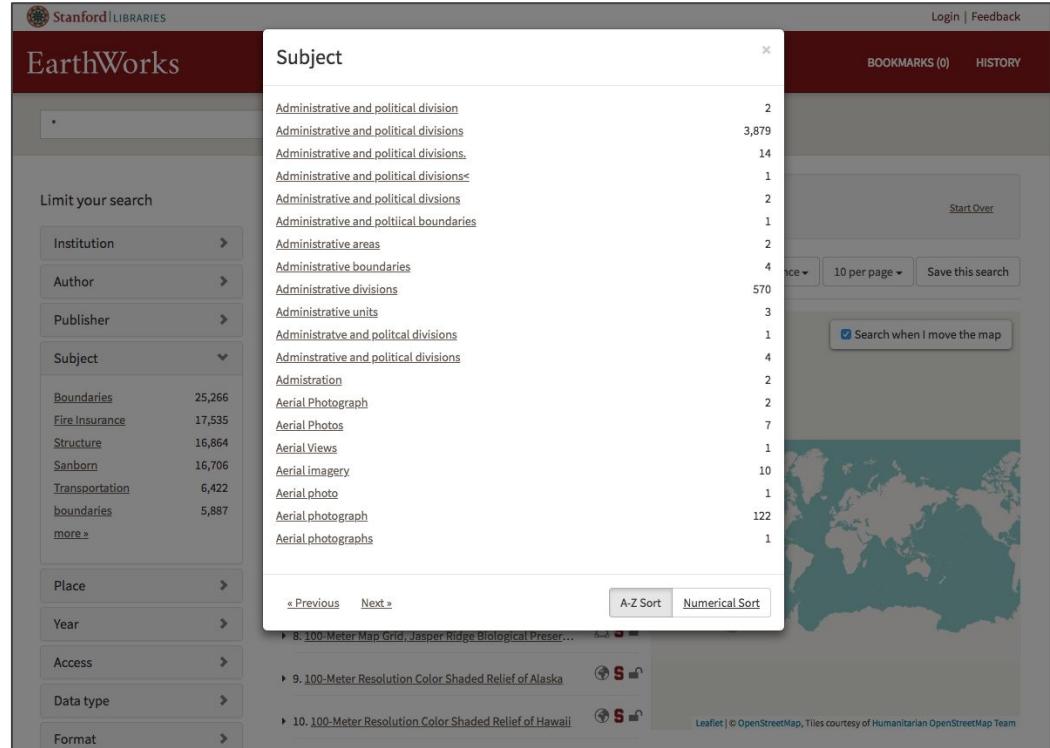
Existing workflows of individuals and institutions

Time and human labor to create and remediate existing records

Lack of continuity or agreement with a common metadata standard

Lack of streamlined tools to create and transform metadata at scale

- Disparate metadata standards
- Varying presentations for GIS file formats
- Improperly formatted or missing required elements
- Broken data access links
- Inconsistent use of controlled vocabularies
- Inaccurate bounding boxes
- Sparse or absent content in the descriptive fields



Common Problems

EarthWorks

BOOKMARKS (0) HISTORY

[Back to Search](#) [Start Over](#)

Azerbaijan Airports, 2008



Author(s) International Steering Committee for Global Mapping (ISCGM)

Description Azerbaijan Airports is a point theme representing airports and airfields in Azerbaijan. This layer is a component of the Global Map, a 1:1,000,000 scale framework dataset of the world. It consists of vector and raster layers of transport, administrative boundaries, drainage, elevation, vegetation, land use and land cover data. The data were prepared from information provided by national mapping and other organizations worldwide.

Place(s) Azerbaijan

Subject(s) transportation

Year 2008

Held by Columbia

⚠ This data is hosted by Columbia, and is currently unavailable to preview and download. Stanford University does not hold this data and cannot provide access to it, at this time. If you have questions about this or other unavailable datasets from Columbia University GIS & Statistical Data Resources please submit your feedback to dssc@library.columbia.edu.



Tools

[Bookmark](#)

[Email](#)

[Web services](#)

[Metadata](#)

Developing Trends: SdrFriend at NYU

```
sdarfriend — bash — 149x32
datasermacbook2:~ andrewbattista$ cd git
datasermacbook2:git andrewbattista$ cd sdrfriend
datasermacbook2:sdrfriend andrewbattista$ bundle exec rake -T
rake build
rake clean
rake clobber
rake fda:addbit[file_path,identifier]
rake fda:bit_batch[bitstream_repository_path,zip_only]
rake fda:delbit[identifier,filename]
rake fda:gbl_to_fda_metadata[csv_input]
rake fda:get[identifier]
rake fda:mint[collection,number,csv_output]
rake fda:translate[handle]
rake files:documentation_containers[destination_path,csv_input]
rake files:download_containers[destination_path,csv_input]
rake files:zip_bitstreams[folder_path]
rake gdal:bounding[shapefile_path]
rake gdal:bounding_many[shapefile_repository_path]
rake geoserver:enable[csv_input]
rake install
rake install:local
rake metadata:alpha_in_place[repository_path]
rake metadata:alphabetize[repository_path]
rake metadata:bithydrate_csv[csv_input,csv_output]
rake metadata:csv_to_json[csv_input,json_output]
rake metadata:dspace_in_place[repository_path]
rake metadata:gencsv[repository_path,csv_output]
rake metadata:split[json_input,destination_path]
rake metadata:template[csv_output]
rake release[remote]
rake spec
datasermacbook2:sdrfriend andrewbattista$
```

Build SdrFriend-0.1.1.gem into the pkg directory
Remove any temporary products
Remove any generated files
Add a single bitstream; <identifier> argument optional IF handle can be extra...
Add multiple bitstreams from a folder; names must conform to handle patterns ...
For an FDA record, delete all bitstreams with a given name
Update FDA metadata with elements from GeoBlacklight CSV
View all current metadata for an FDA item -- identifier can be handle or dspace_id
Create multiple container records / "mint" new handles; collection should be ...
Translate a handle into a dspace internal id
Create documentation container folders for bitstreams, given a CSV listing la...
Create download container folders for bitstreams, given a CSV listing layers
Create zips of all containers residing in a folder
Get solr_geom (in ENVELOPE syntax) for a Shapefile that is in WGS84 / EPSG:4326
Get solr_geom for all Shapefiles, in a directory or subdirectory
Turn on GeoServer layers from a CSV input
Build and install SdrFriend-0.1.1.gem into system gems
Build and install SdrFriend-0.1.1.gem into system gems without network access
Alphabetize keys in records (saves back to original)
Alphabetize keys in records
Add FDA bitstream URLs to CSV
Convert CSV to single JSON record collection file
Add DSpace IDs in records in place (saves back to original)
Generate CSV from set of JSON records
Split-out JSON record collection file into multiple individual geoblacklight....
Create a blank template CSV for collection
Create tag v0.1.1 and build and push SdrFriend-0.1.1.gem to TODO: Set to 'htt...
Run RSpec code examples

SDR Friend is a Ruby gem that we use to interact with the APIs and interfaces of our collection infrastructure

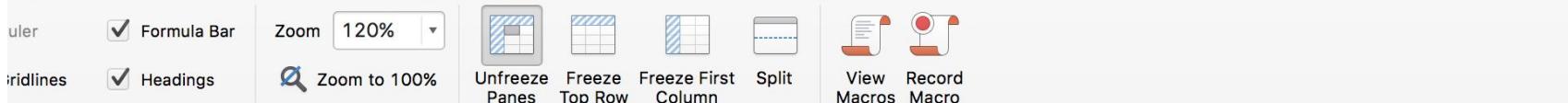


Test Case: Columbia University's GeoBlacklight Metadata

- Harmonizes subjects
- Fixes null fields and reconciles fields not following GeoBlacklight norms
- Corrects typos and removes encoding glitches, unnecessary carriage returns
- Updates authoritative bibliographic information by inserting codebook links

[Summary of pull request](#)

[Summary of metadata being edited](#)



United States. Defense Mapping Agency | East View Cartographic, Inc.

Subject	X	Description
Real property	2,523	Flood risk
Fire insurance maps	2,084	
Fire risk assessment	2,081	es of 104
Transportation	1,428	originally de
Environment	1,288	
Society	1,053	
Boundaries	795	Districts
Imagery and Base Maps	723	mbined i
Planning and Cadastral	697	
Inland Waters	665	
Location	576	
Geoscientific Information	396	er image
Land ownership	384	erenced
Topographic maps	353	
Atlases	344	
Economy	312	
Biota	299	er image
Roads	271	erenced

```
6 + "dc_format_s": "Shapefile",
7 + "dc_identifier_s": "urn:columbia.edu:Columbia.global_mapping_aze_aquecanl",
8 + "dc_language_s": "English",
9 + "dc_publisher_s": "International Steering Committee for Global Mapping",
10 + "dc_rights_s": "Public",
11 "dc_subject_sm": [
25 - "transportation",
26 - "inlandWaters",
27 - "structure"
12 + "Transportation",
13 + "Inland Waters",
14 + "Structure"
28 15 ],
29 16 + "dc_title_s": "Azerbaijan Aqueducts, Canals, Flumes and Penstocks, 2008",
17 "dc_type_s": "Dataset",
18 + "dct_isPartOf_sm": [
19 + "Global Map"
20 + ],
21 + "dct_issued_s": "",
22 + "dct_provenance_s": "Columbia",
23 + "dct_references_s": "{\"http://schema.org/url\": \"https://clio.columbia.edu/?q=urn:columbia.edu:Columbia.global_mapping_aze_aquecanl\"}",
24 + "dct_source_sm": [
25 +
26 + ],
```

```

1 "uuid": "urn:columbia.edu:Columbia.global_mapping_aze_aerofacp",
2 "dc_identifier_s": "urn:columbia.edu:Columbia.global_mapping_aze_aerofacp",
3 "dc_title_s": "Azerbaijan Airports, 2008",
4 "dc_description_s": "Azerbaijan Airports is a point theme representing airports and airfields in Azerbaijan. This layer is a component of the Global Map, a 1:1,000,000 scale framework dataset of the world. It consists of vector and raster layers of transport, administrative boundaries, drainage, elevation, vegetation, land use and land cover data. The data were prepared from information provided by national mapping and other organizations worldwide.",
5 "dc_rights_s": "Public",
6 "dc_provenance_s": "Columbia",
7 "dc_references_s":
8 "(\\"http://www.opengis.net/def/serviceType/ogc/wms\"\":\\"https://geoserver.cul.columbia.edu/geoserver/wms/sde\"\",\\"http://www.opengis.net/def/serviceType/ogc/wfs\"\":\\"https://geoserver.cul.columbia.edu/geoserver/sde/ows\"\",\\"http://schema.org/downloadUrl\"\":\\"http://www.columbia.edu/acis/eds/gis/images/aze_aerofac.zip\"\\",\\\"http://www.w3.org/1999/xhtml\"\":\\"https://geodata.library.columbia.edu/metadata/fcdc/html/global_mapping_aze_aerofac.html\"\\",\\\"http://www.isotc211.org/schemas/2005/gmd\"\":\\"https://geodata.library.columbia.edu/metadata/fcdc/current/global_mapping_aze_aerofac.xml\"\")",
9 "georss_box_s": "39.154125 45.447075 40.742939 50.046078",
10 "layer_id_s": "sde:columbia.global_mapping_aze_aerofacp",
11 "layer_geom_type_s": "Point",
12 "layer_modified_dt": "2009-09-01T00:00:00Z",
13 "layer_slug_s": "sde-columbia-global_mapping_aze_aerofacp",
14 "solr_geom": "ENVELOPE(45.447075, 50.046078, 40.742939, 39.154125)",
15 "solr_year_i": "2008",
16 "dc_creator_sm": [
17   "International Steering Committee for Global Mapping (ISCGM)"
18 ],
19 "dc_format_s": "Shapefile",
20 "dc_language_s": "",
21 "dc_publisher_s": [
22 ],
23 "dc_subject_sm": [
24   "transportation"
25 ],
26 "dc_type_s": "Dataset",
27 "dc_spatial_sm": [
28   "Azerbaijan"
29 ],
30 "dct_temporal_sm": [
31   "2008"
32 ],
33 "dct_issued_s": "",
34 "dct_isPartOf_sm": [
35 ],
36 "nyu_addl_format_sm": [
37   "Shapefile"
38 ],
39 "solr_geom": "ENVELOPE(45.447075, 50.046078, 40.742939, 39.154125)",
40 "solr_year_i": 2008
41 }

```

Before

```

1 "dc_creator_sm": [
2 ],
3 "dc_description_s": "Azerbaijan Airports is a point theme map representing airports and airfields in Azerbaijan. This layer is a component of the Global Map, a 1:1,000,000 scale framework dataset of the world. It consists of vector and raster layers of transport, administrative boundaries, drainage, elevation, vegetation, land use and land cover data. The data were prepared from information provided by national mapping and other organizations worldwide. The Global Map data had been managed by the Geospatial Information Authority of Japan, but that organization was terminated in August, 2016. The data has been transferred to the Geospatial Information Section in the United Nations. For the most recent documentation, refer to the Global Map Specifications at https://github.com/globalmaps/specifications/.",
4 "dc_format_s": "Shapefile",
5 "dc_identifier_s": "urn:columbia.edu:Columbia.global_mapping_aze_aerofacp",
6 "dc_language_s": "English",
7 "dc_publisher_s": "International Steering Committee for Global Mapping",
8 "dc_rights_s": "Public",
9 "dc_subject_sm": [
10   "Transportation",
11   "Airports"
12 ],
13 "dc_title_s": "Azerbaijan Airports, 2008",
14 "dc_type_s": "Dataset",
15 "dct_isPartOf_sm": [
16   "Global Map"
17 ],
18 "dct_issued_s": "",
19 "dct_provenance_s": "Columbia",
20 "dct_references_s":
21 "(\\"http://schema.org/url\"\":\\"https://clio.columbia.edu/?q=urn:columbia.edu:Columbia.global_mapping_aze_aerofacp\"\",\\"http://schema.org/downloadUrl\"\":\\"http://www.columbia.edu/acis/eds/gis/images/aze_aerofac.zip\"\\",\\\"http://www.opengis.net/def/serviceType/ogc/wfs\"\":\\"https://geoserver.cul.columbia.edu/geoserver/sde/ows\"\\",\\\"http://www.opengis.net/def/serviceType/ogc/wms\"\":\\"https://geoserver.cul.columbia.edu/geoserver/wms/sde\"\",\\\"http://lccn.loc.gov/sh85035852\"\":\\"https://github.com/globalmaps/specifications/blob/master/gmspec-2.2.pdf\"\")",
22 "dct_source_sm": [
23 ],
24 "dct_spatial_sm": [
25   "Azerbaijan"
26 ],
27 "dct_temporal_sm": [
28   "2008"
29 ],
30 "geolacklight_version": "1.0",
31 "layer_geom_type_s": "Point",
32 "layer_id_s": "sde:columbia.global_mapping_aze_aerofacp",
33 "layer_modified_dt": "2018-04-23T0016Z",
34 "layer_slug_s": "sde-columbia-global_mapping_aze_aerofacp",
35 "nyu_addl_format_sm": [
36   "Shapefile"
37 ],
38 "solr_geom": "ENVELOPE(45.447075, 50.046078, 40.742939, 39.154125)",
39 "solr_year_i": 2008
40 }

```

After

Metadata would have to be created in reverse (i.e., GeoBlacklight to FGDC or ISO 19139) when fixed

Metadata would have to be fixed in multiple places (i.e., institutional repository, discovery application)

The practices of one institution might not mirror the practices of another



Potential Barriers to Collaborative Edits

Metadata Scoring Rubric and article on GIS metadata for consortial use -
<https://osf.io/7auts/>

Blog post series on developing a geospatial data workflow at NYU -
<https://andrewbattista.github.io/>

Repository for SdrFriend -
<https://github.com/sqbalogh/sdrfriend/> (Private for now - request access)



Additional Resources

Digital Colonialism?

Taylor Hixson

taylor.hixson@nyu.edu
Geospatial Services Librarian
New York University Abu Dhabi Library

Background



NYU Spatial Data Repository

Submit History Login

Find maps and data through NYU

Get started here

Search... Search

Browse by category

Institution: Princeton, Harvard, NYU, Stanford, Baruch CUNY, More

Data type: Image, Polygon, Raster, Line, Point, Mixed, More

Placename

Subject

Browse by location

Search here

A screenshot of the NYU Spatial Data Repository website. The header is purple with the NYU logo and "Spatial Data Repository". Below the header, there's a search bar with "Find maps and data through NYU" and "Get started here" buttons. The main content area has sections for "Browse by category" (Institution, Data type), "Browse by location" (a world map), and "Placename" and "Subject" filters. The "Institution" section lists Princeton, Harvard, NYU, Stanford, Baruch CUNY, and more. The "Data type" section lists Image, Polygon, Raster, Line, Point, Mixed, and More. The "Placename" and "Subject" sections have dropdown menus. The "Browse by location" section shows a world map with a search bar.

Background



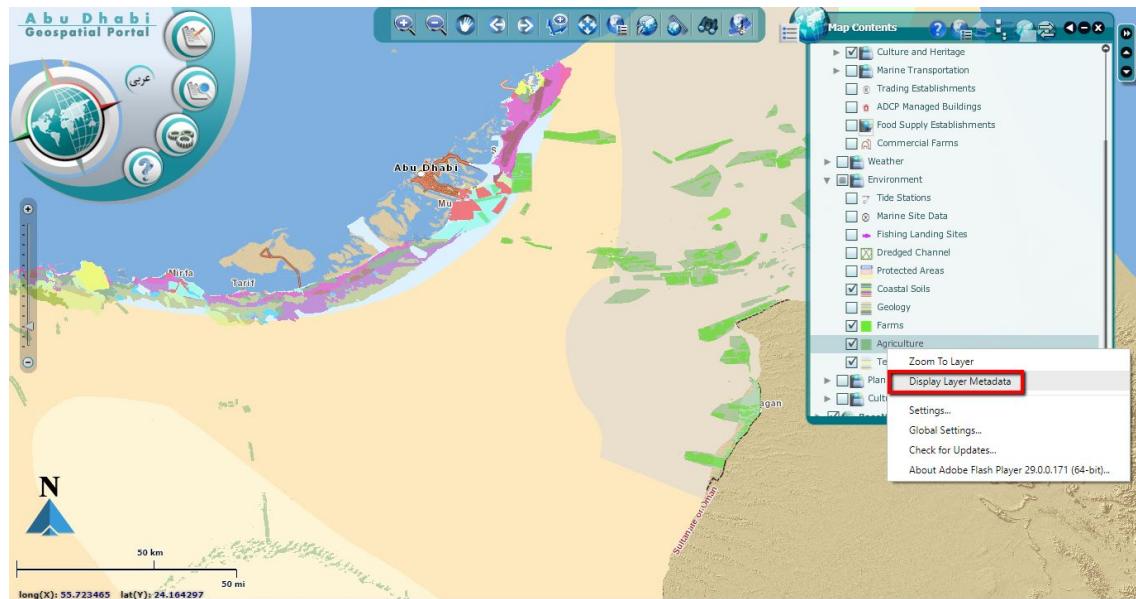
Bayanat.ae

United Arab Emirates

Welcome to the
**Abu Dhabi Spatial Data
Infrastructure Program**

Geography is where we live, work, play and learn...
based information gives us the means to better understand
around us...

حكومة أبوظبي الرقمية
ABU DHABI DIGITAL GOVERNMENT



Background

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0" xmlns:georss="http://www.georss.org/georss" xmlns:media="http://search.yahoo.com/mrss" xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/"
xmlns:atom="http://www.w3.org/2005/Atom">
<channel>
<title><![CDATA[Geoportal GEORSS.]]></title>
<description><![CDATA[Most recently updated metadata documents.]]></description>
<link><![CDATA[http://geoportal.abudhabi.ae/geoportal]]></link>
<docs><![CDATA[http://www.rssboard.org/rss-specification]]></docs>
<category><![CDATA[GeoRSS]]></category>
<copyright><![CDATA[Copyright (C) Environmental Systems Research Institute, 2008.]]></copyright>
<generator><![CDATA[Geoportal Extension 9.3.1]]></generator>
<managingEditor><![CDATA[adsdi.servicedesk@adsic.abudhabi.ae]]></managingEditor>
<webMaster><![CDATA[adsdi.servicedesk@adsic.abudhabi.ae]]></webMaster>
<item>
<link><![CDATA[http://geoportal.abudhabi.ae/geoportal/rest/document?id=%7BAB707304-F382-4E93-AADC-1FC43A8E07D2%7D]]></link>
<guid isPermaLink="true">http://geoportal.abudhabi.ae/geoportal/rest/document?id=%7BAB707304-F382-4E93-AADC-1FC43A8E07D2%7D</guid>
<title><![CDATA[Commercial Farms ADFCA]]></title>
<pubDate><![CDATA[Tue, 12 Jul 11 09:40:16 GST]]></pubDate>
<source url="http://geoportal.abudhabi.ae/geoportal/rest/find/document?max=500&searchText=Farms">Geoportal GeoRSS.</source>
<description><![CDATA[
<div class="snippet">
<div class="abstract" style="overflow: auto;">

The dataset represents Commercial Farm locations for the entire emirate of Abu Dhabi as point features that can be identified by Name and Type, as supplied by ADFCA.
</div>
<div class="links">
<A href="http://geoportal.abudhabi.ae/geoportal/catalog/search/viewMetadataDetails.page?uuid=%7BAB707304-F382-4E93-AADC-1FC43A8E07D2%7D" target="_blank">Details</A>
<A href="http://geoportal.abudhabi.ae/geoportal/rest/document?id=%7BAB707304-F382-4E93-AADC-1FC43A8E07D2%7D" target="_blank">Metadata</A>
</div>
</div>
]]></description>
<category><![CDATA[offlineData]]></category>
<media:content url="http://geoportal.abudhabi.ae/directories/images/metadata/AB707304-F382-4E93-AADC-1FC43A8E07D2.png"/>
<media:text><![CDATA[Commercial Farms ADFCA]]></media:text>
<georss:polygon>24.5074 54.3289 24.5074 55.8294 23.6257 55.8294 23.6257 54.3289 24.5074 54.3289</georss:polygon>
</item>
<item>
<link><![CDATA[http://geoportal.abudhabi.ae/geoportal/rest/document?id=%7B16C23A1F-0A5B-419D-A902-9BB65327C0AC%7D]]></link>
<guid isPermaLink="true">http://geoportal.abudhabi.ae/geoportal/rest/document?id=%7B16C23A1F-0A5B-419D-A902-9BB65327C0AC%7D</guid>
<title><![CDATA[Izbas (Priate Farms) ADFCA]]></title>
<pubDate><![CDATA[Tue, 12 Jul 11 09:39:33 GST]]></pubDate>
<source url="http://geoportal.abudhabi.ae/geoportal/rest/find/document?max=500&searchText=Farms">Geoportal GeoRSS.</source>
<description><![CDATA[
... --- "...--"
]]></description>
```

Challenge:

U.S. authorities to describe
non-U.S. concepts and data

1. Subjects
2. Geographic areas

Are camels livestock? LCSH says no. Emiratis say yes!

Select a Link Below to Continue...

[Authority Record](#)

[Scope Note](#)

Narrower Term: [Cattle.](#)

Narrower Term: [Donkeys](#)

Narrower Term: [Draft animals.](#)

Narrower Term: [Female livestock](#)

Narrower Term: [Feral livestock](#)

Narrower Term: [Goats.](#)

Narrower Term: [Hinnies](#)

Narrower Term: [Horses](#)

Narrower Term: [Male livestock](#)

Narrower Term: [Miniature livestock.](#)

Narrower Term: [Minilivestock](#)

Narrower Term: [Mithun](#)

Narrower Term: [Mules.](#)

Narrower Term: [Photography of livestock.](#)

Narrower Term: [Poultry](#)

Narrower Term: [Sheep.](#)

Narrower Term: [Swine.](#)

See Also: [Food animals.](#)

See Also: [Herders.](#)

See Also: [Rangelands.](#)

SOURCE OF HEADINGS: Library of Congress Online Catalog

INFORMATION FOR: Camels.

Please note: Broader Terms are not currently available

Select a Link Below to Continue...

[Authority Record](#)

Narrower Term: [Bactrian camel](#)

LC control no.: sh 85019119

LCCN Permalink: <https://lccn.loc.gov/sh85019119>

HEADING: Camels

000 00579cz a2200241n 450

001 4671728

005 20120322080218.0

008 860211i anannbabn |a ana

010 __ |a sh 85019119

035 __ |a (DLC)sh 85019119

035 __ |a (DLC)18413

040 __ |a DLC |c DLC |d DLC

053 _0 |a QL737.U54 |c Zoology

053 _0 |a SF401.C2 |c Culture

150 __ |a Camels

450 __ |a Arabian camel

450 __ |a Camelus

450 __ |a Camelus dromedarius

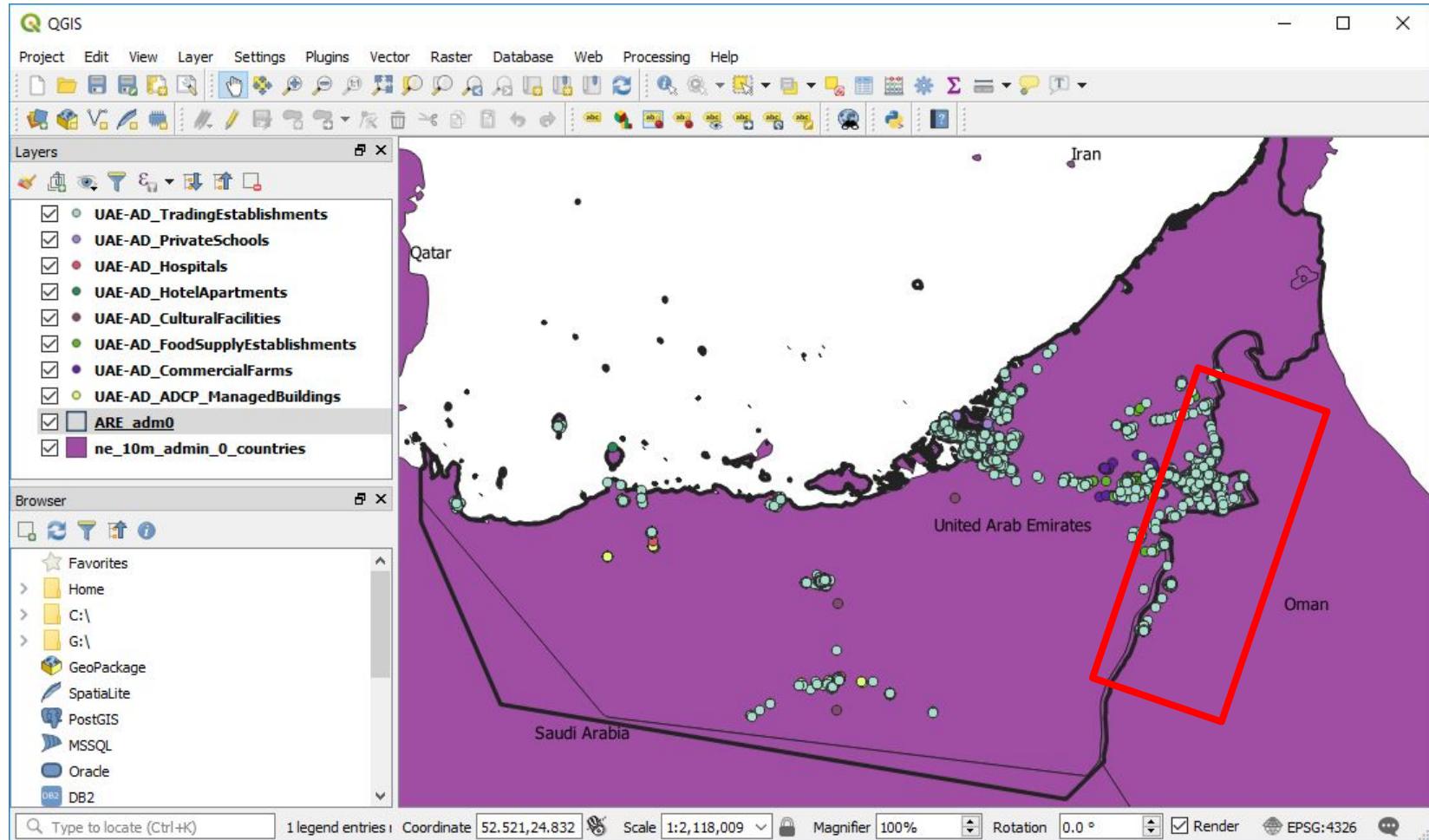
450 __ |w nne |a Dromedary

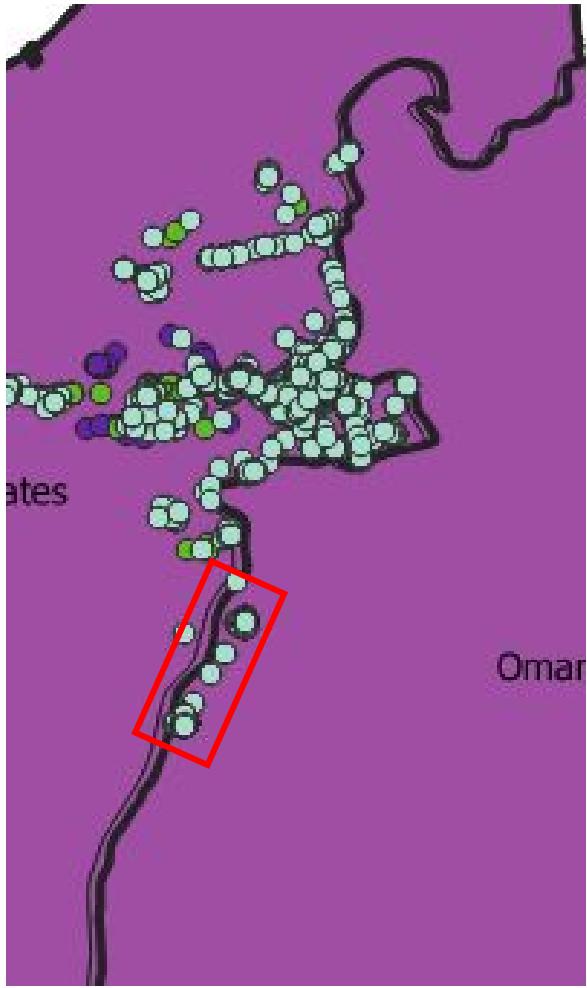
450 __ |a One-humped camel

550 __ |w g |a Camelidae

906 __ |t 9217 |u fk04 |v 0

953 __ |a xx00 |b fk05





UAE Bounding Box with OSM

BoundingBox [Georeference](#)

+ Find a place with OpenStreetMap ... Find a place with Google ...

Copy & Paste

MARC ▾

034: <code>\$\$dE0512400\$\$eE0563600\$\$fN0260900\$\$gN0223700</code>
255: <code>\$\$c(E 51°24'00"--E 56°36'00" / N 26°09'00"--N 22°37'00")</code>

Copyright © 2017 Klokant Technologies

UAE Bounding Box with Google

BoundingBox Georeference

+ Find a place with OpenStreetMap ... Find a place with Google ...

adhd
الدبي
Al-Kharj
الخرج

Dammam
الدمام

Bahrain
البحرين

Al Hufuf
العفوف

Qatar
قطر

Ras al-Khaimah
رأس الخيمة

Dubai
دبي

Fujairah
الفجيرة

Abu Dhabi
أبوظبي

United Arab Emirates

Al Ain
العين

Sohar
صهار

Iri
يري

Nizwa
نزوءة

Copy & Paste

034: \$dE0512800\$\$eE0562600\$\$fN0260400\$\$gN0223700

MARC

255: \$\$c(E 51°28'00"--E 56°26'00"/N 26°04'00"--N 22°37'00")

Copyright © 2017 [Klokán Technologies](#) -

--	--	--	--	--	--	--	--	--	--

Copy & Paste

MARC ▼

034: \$dE05128000\$eE0562600\$\$fN0260400\$\$gN0223700
255: \$\$c[E 51°28'00"~-E 56°26'00"/N 26°04'00"~-N 22°37'00")

--	--	--	--	--	--	--	--	--	--

Copy & Paste

MARC ▼

034: \$dE05124000\$eE0563600\$\$fN0260900\$\$gN0223700
255: \$\$c[E 51°24'00"~-E 56°36'00"/N 26°09'00"~-N 22°37'00")

Opportunity: Supervised self-deposit

NYUAD Spatial Data Repository Deposit Form

Use this form to provide descriptive information about your geospatial data. All data submitted through this form will be given consideration for submission to the NYU Spatial Data Repository (geo.nyu.edu).

For examples of how to format the required information, see this spreadsheet:
tinyurl.com/ydfxpob

For questions about the requirements, email NYU Abu Dhabi Geospatial Data Services Librarian Taylor Hixson (taylor.hixson@nyu.edu)

Writing geospatial data titles

The title should provide a few words about the data content and location. Do not include any temporal information here.

Title *

Your answer

Writing geospatial data descriptions

Vector layer descriptions: "This _____ shapefile layer represents...." Fill in the blank with the vector layer type: point, line, polygon, or mixed.

Raster layer descriptions: "This georeferenced raster layer represents...."

In descriptions, include information about the location, content, time period, and anything related to data collection and creation methods.

Description *

Your answer

taylor.hixson@nyu.edu

Further readings related to digital colonialism:
tinyurl.com/HixsonIASSIST18

Thank you!

Questions?