Saving Data

Lessons from Data Refuge for Librarians

Laurie Allen

ALCTS Exchange

May 18, 2017

Thank You

Also huge thanks to our collaborators & partners

Azavea

University of Michigan Libraries

Temple Libraries

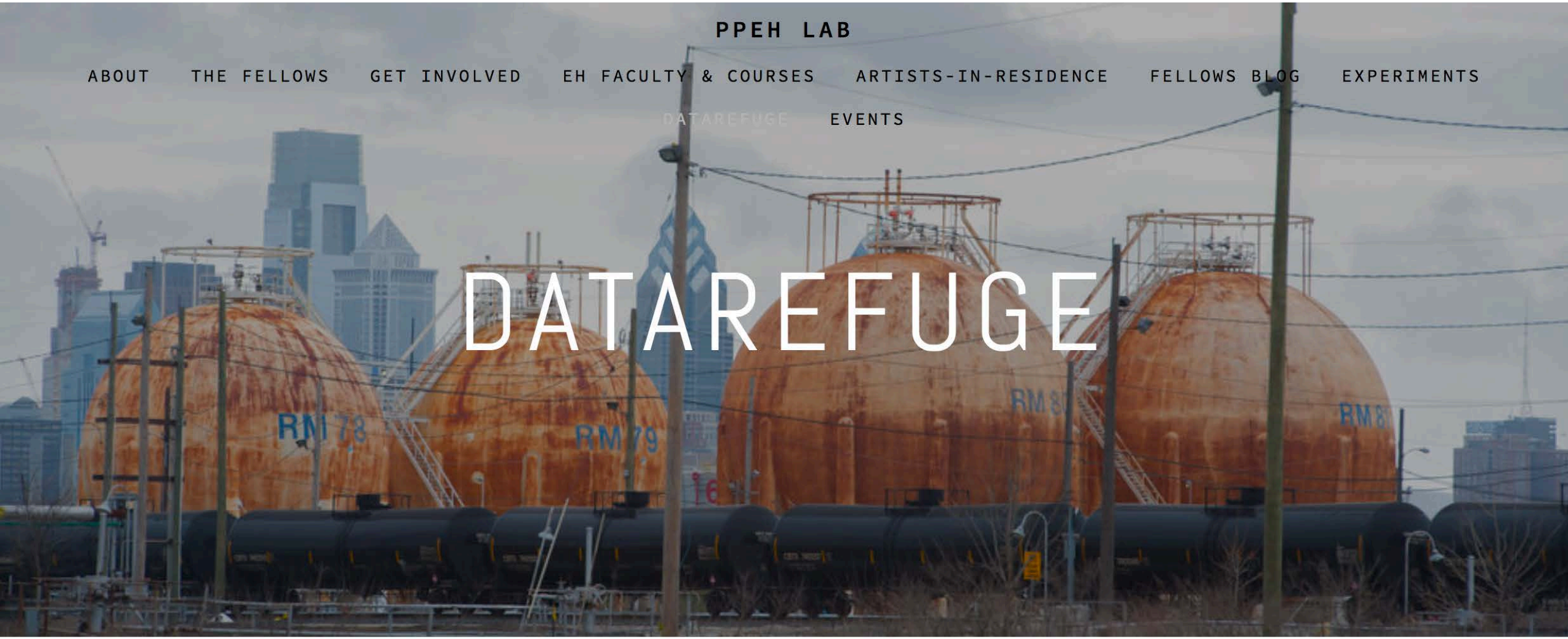
Internet Archive

ProjectARCC

Environmental Data Governance Inititaitve

Union of Concerned Scientists

Climate Mirror



DataRefuge Project | Get Involved

www.datarefuge.org | https://libraries.network

How to Rescue Data | Survey of Valuable Datasets | Storytelling | Future DataRescue Events

FAQs | Latest Updates | Press

Laurie Allen 0

DataRefuge

Datasets Organizations Groups About

Search

Building refuge for federal climate & environmental data

DataRefuge helps to build refuge for federal data and supports climate and environmental research and advocacy. We are committed to fact-based arguments. DataRefuge preserves the facts we need at a time of ongoing climate change.

This site is one part of the project. The vast majority of the government information gathered through this project is available from the Internet Archive through the End of Term project. This data catalog is a place to store data that is difficult or impossible to harvest through web crawlers. Beginning to add difficult data to this websites is one of the tasks of the DataRescuePhilly event on January 13th and 14th.

Questions

- Is there vulnerable and valuable data?
- What is the scope of the problem?
- Who has already been working on this?
- What's already been done?
- How can we make sure we have the most valuable data?
- How can we measure the vulnerability of data?
- How can we engage all kinds of expertise?
- How can we make sure our work is closely tied to our communities?

Questions

- How can we ensure copied data remains citable by scientists?
- Where, how should we store copies?
- How should we describe the copied data?

Questions

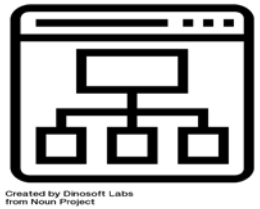
- What is federal environmental and climate data?

Federal data ≅ webpage

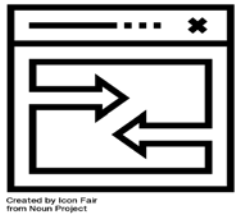
Basic HTML

Embedded Content

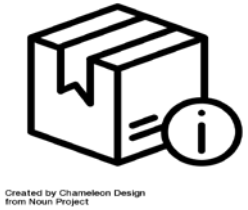
Directories of Files



Query Interface



Research Dataset

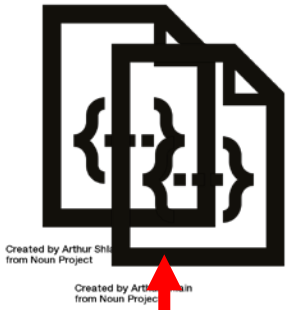


INTERNET ARCHIVE
WayBackMachine

DataRefuge

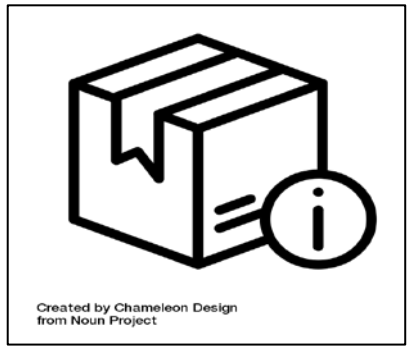
HTML Pages

Data files

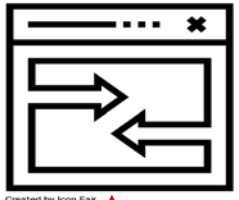


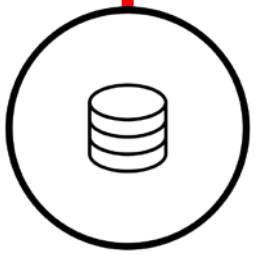
Metadata

Research Dataset

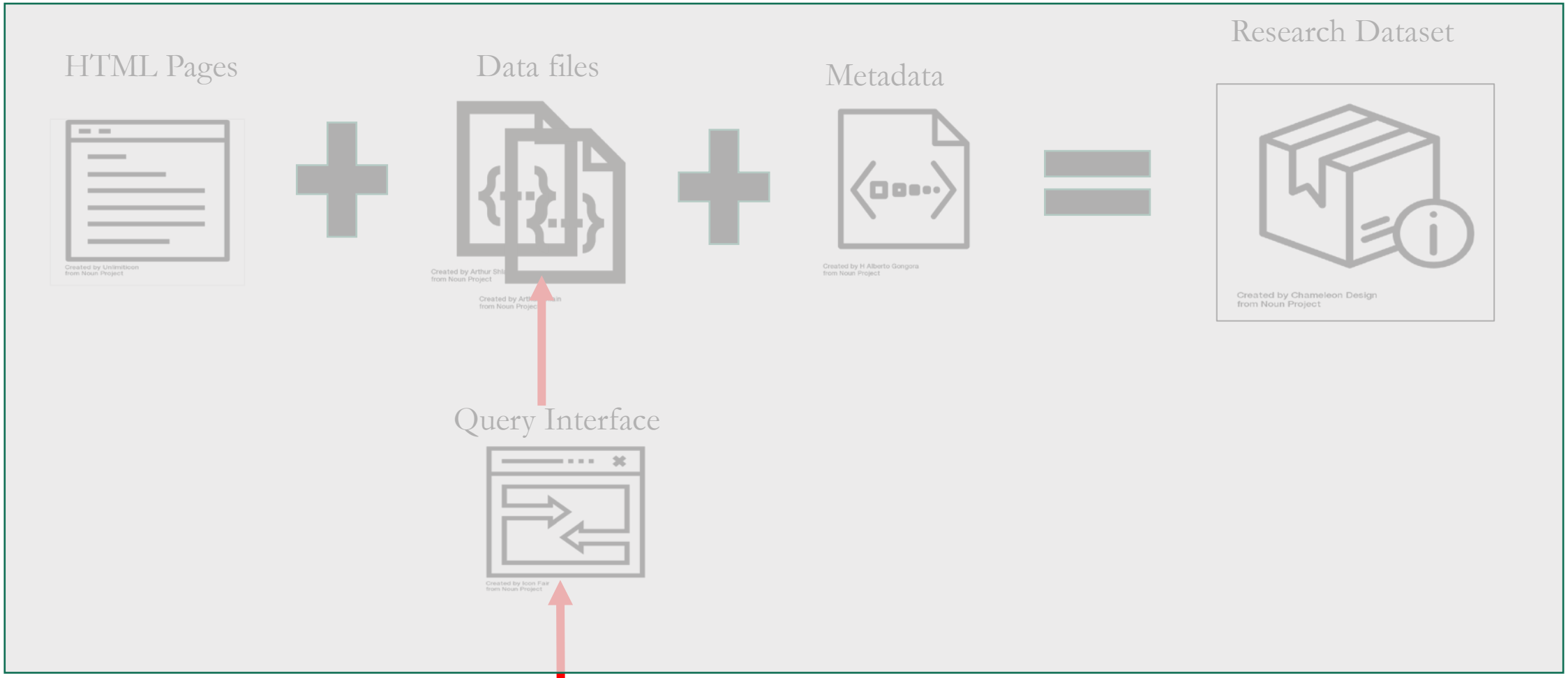


Query Interface



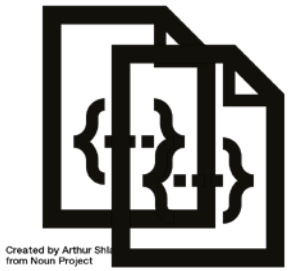


Less than ideal, but better than nothing

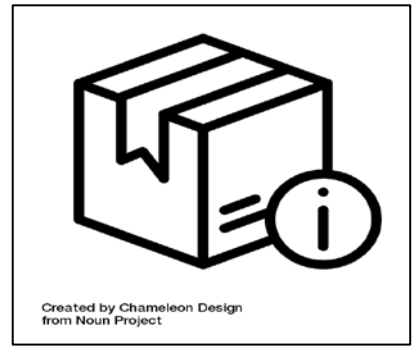


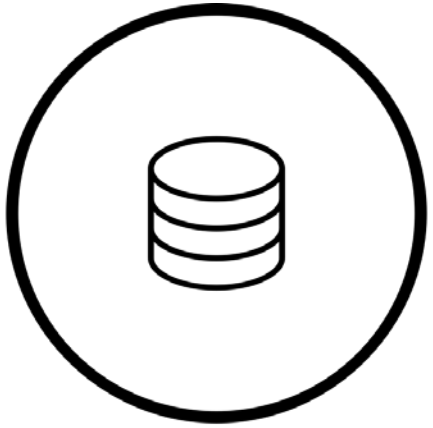
Where does this come from?

Data files



Metadata





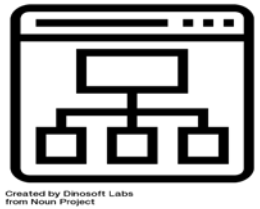
Data Management!

Federal data ≅ webpage

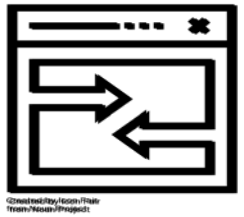
Basic HTML

Embedded Content

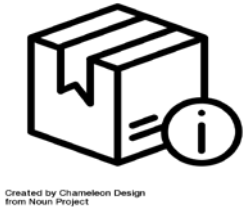
Directories of Files



Query Interface



Research Dataset

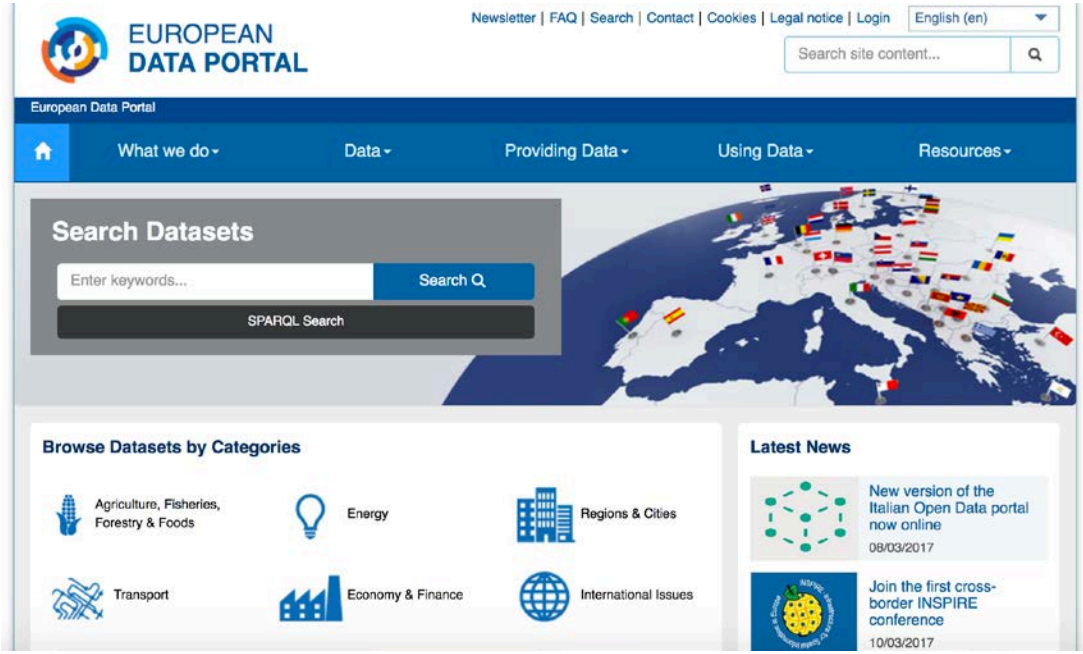


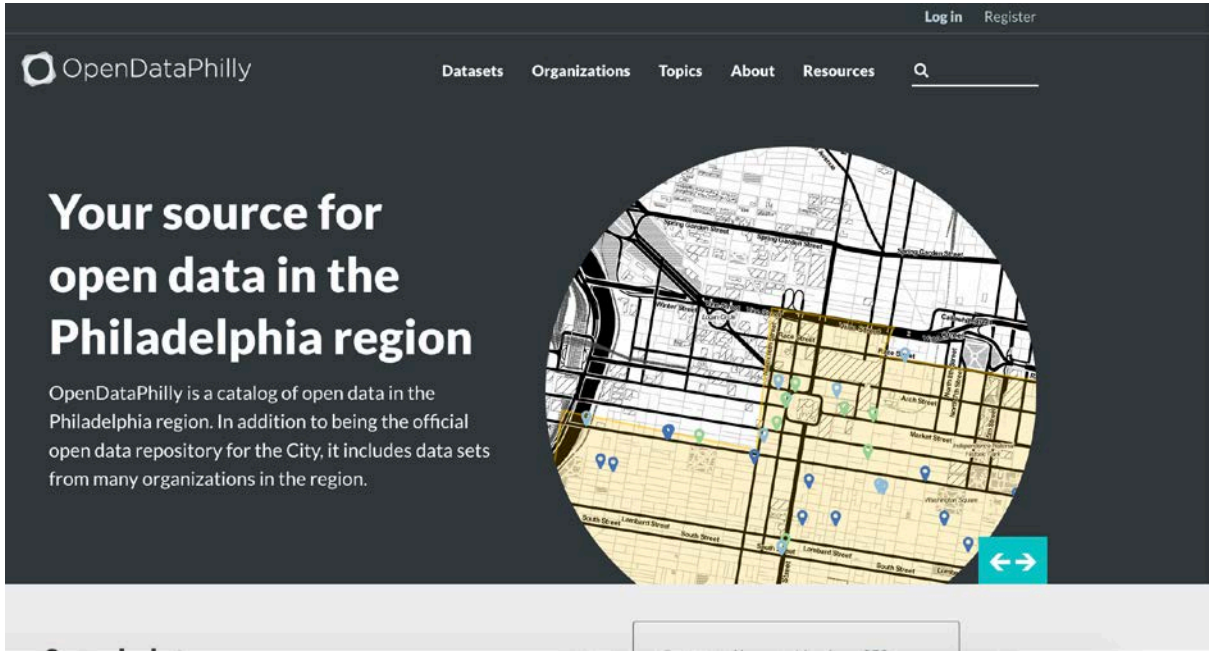
Who's work is this?

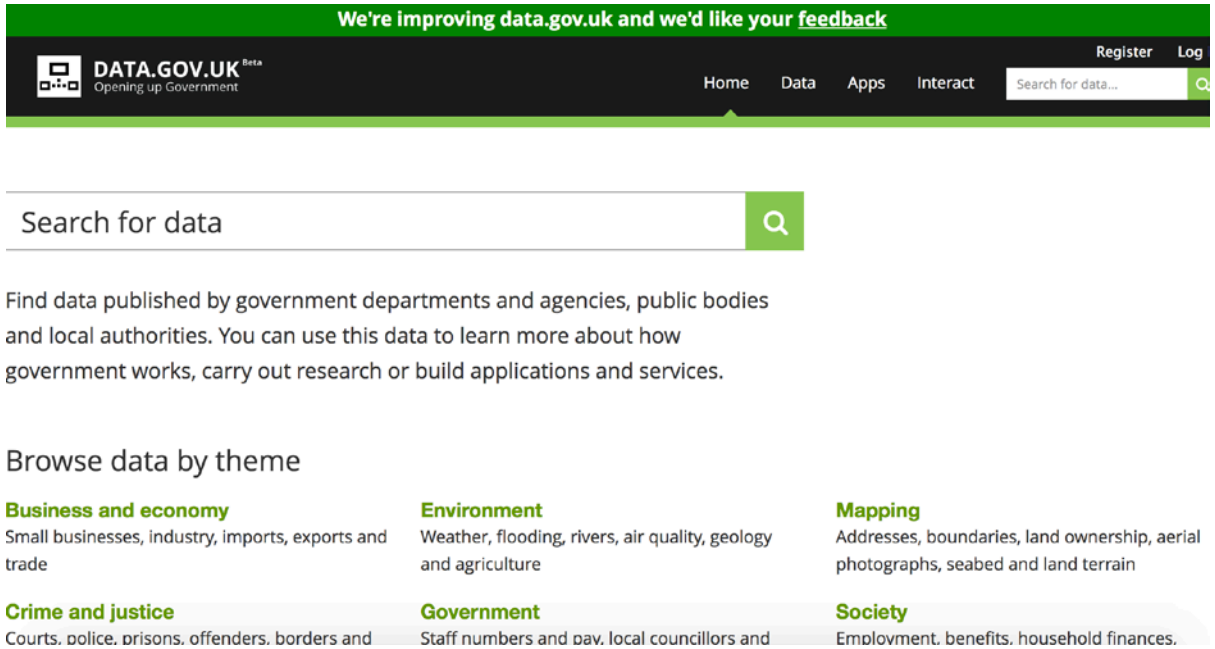
Into the nitty gritty for a moment….

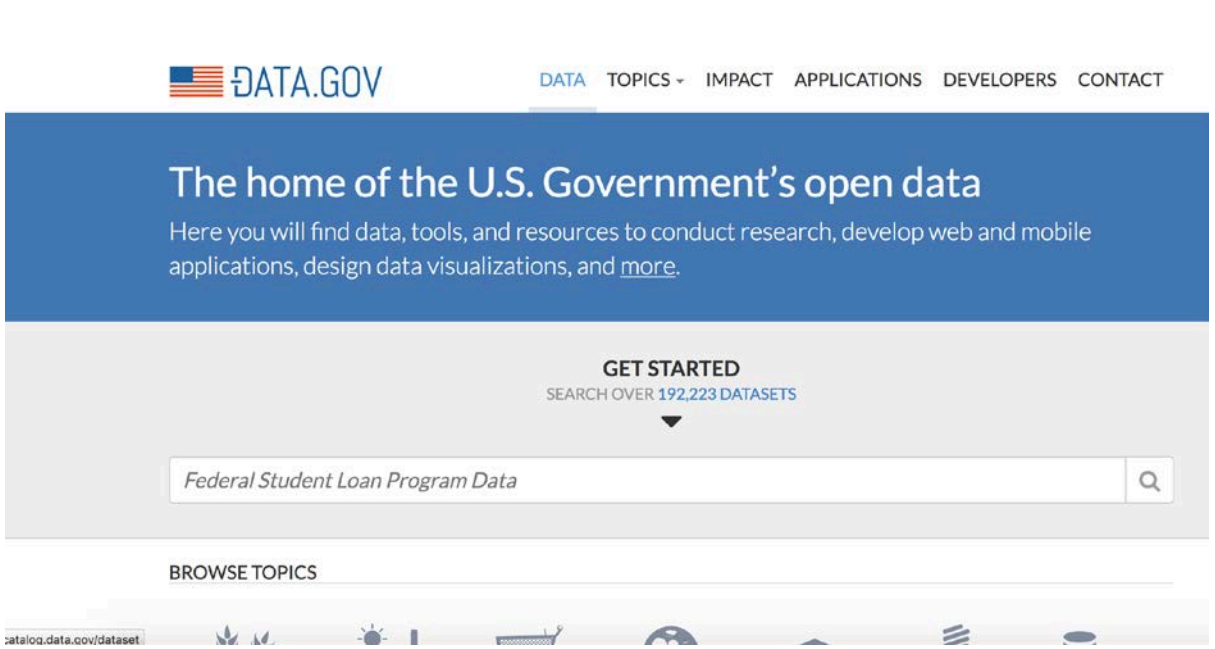
This mission is achieved through:

- Organizing processes that enable desired information to be identified and located.
- Expert assistance rendered by trained professionals in a network of libraries.
- Collections of publications at a network of libraries.
- Archived online information dissemination products from GPO Access, Federal agency Web sites, and partner Web sites.


EUROPEAN DATA PORTAL
Newsletter | FAQ | Search | Contact | Cookies | Legal notice | Login
English (en)
Search site content...
European Data Portal
What we do
Data
Providing Data
Using Data
Resources
Search Datasets
Enter keywords...
Search
SPARQL Search
Browse Datasets by Categories
Agriculture, Fisheries, Forestry & Foods
Energy
Regions & Cities
Transport
Economy & Finance
International Issues
Latest News
New version of the Italian Open Data portal now online
08/03/2017
Join the first cross-border INSPIRE conference
10/03/2017


Log in
Register
OpenDataPhilly
Datasets
Organizations
Topics
About
Resources
Your source for open data in the Philadelphia region
OpenDataPhilly is a catalog of open data in the Philadelphia region. In addition to being the official open data repository for the City, it includes data sets from many organizations in the region.


We're improving data.gov.uk and we'd like your feedback
DATA.GOV.UK Beta
Opening up Government
Register
Log
Home
Data
Apps
Interact
Search for data...
Search for data
Find data published by government departments and agencies, public bodies and local authorities. You can use this data to learn more about how government works, carry out research or build applications and services.
Browse data by theme
Business and economy
Small businesses, industry, imports, exports and trade
Environment
Weather, flooding, rivers, air quality, geology and agriculture
Mapping
Addresses, boundaries, land ownership, aerial photographs, seabed and land terrain
Crime and justice
Courts, police, prisons, offenders, borders and
Government
Staff numbers and pay, local councillors and
Society
Employment, benefits, household finances,


DATA.GOV
DATA
TOPICS
IMPACT
APPLICATIONS
DEVELOPERS
CONTACT
The home of the U.S. Government's open data
Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.
GET STARTED
SEARCH OVER 192,223 DATASETS
Federal Student Loan Program Data
BROWSE TOPICS

Federal

U.S. Department of Transportation

Federal Highway Administration

Federal Highway Administration

Publisher

Federal Highway Administration

Contact

Dan Morgan

Share on Social Sites

Google+

Twitter

Facebook

Significant Guidance Issued by the Federal Highway Administration

Metadata Updated: May 25, 2016

A list of Significant Guidance documents, which include guidance document disseminated to regulated entities or the general public that may reasonably be anticipated to lead to an annual effect on the economy of $100 million or more or adversely affect in a material way the economy, a sector of the economy, productivity, competition, jobs, the environment, public health or safety, or State, local, or tribal governments or communities; create a serious inconsistency or otherwise interfere with an action taken or planned by another agency; materially alter the budgetary impact of entitlements, grants, user fees, or loan programs or the rights and obligations of recipients thereof; or raise novel legal or policy issues arising out of legal mandates, the President's priorities, or the principles set forth in Executive Order 12866, as further amended.

Access & Use Information

Public: This dataset is intended for public access and use.

License: See this page for license information.

Collection

This dataset is a collection of other datasets.

Search datasets within this collection

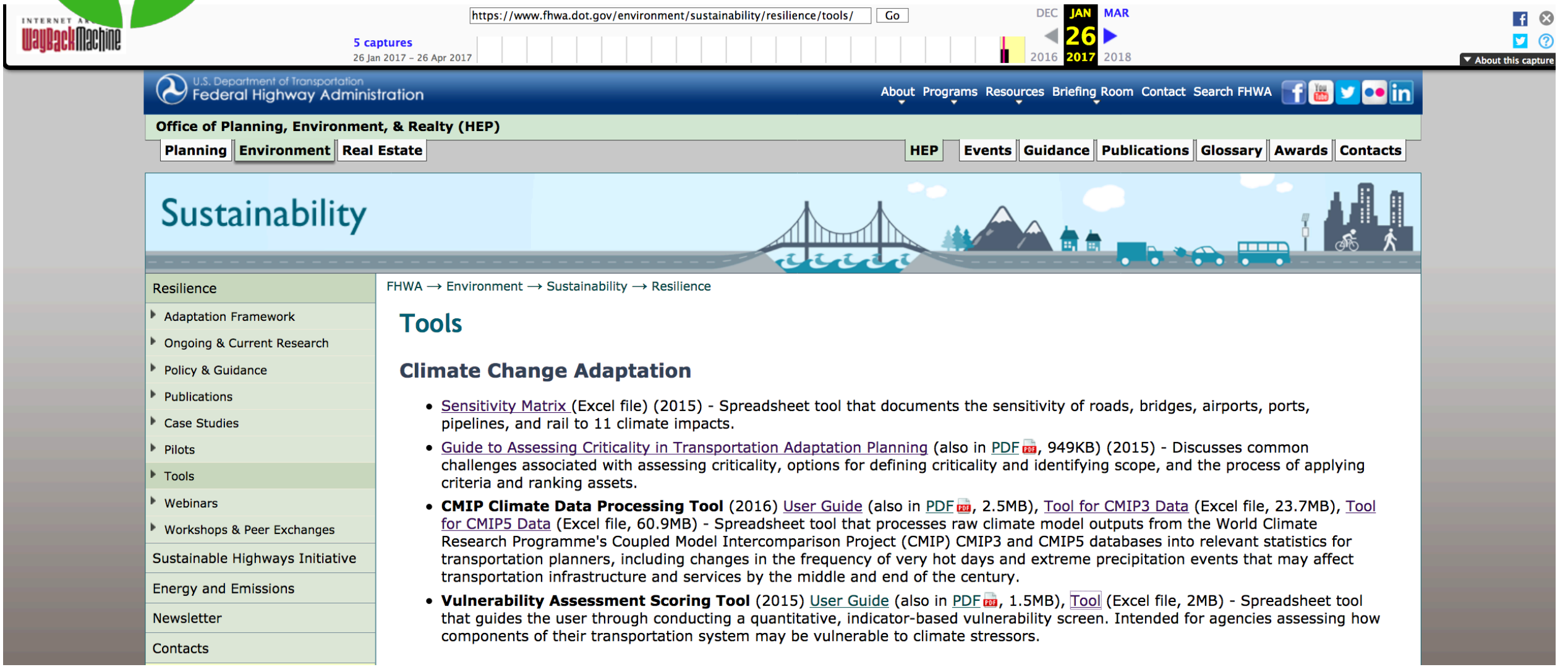
Downloads & Resources

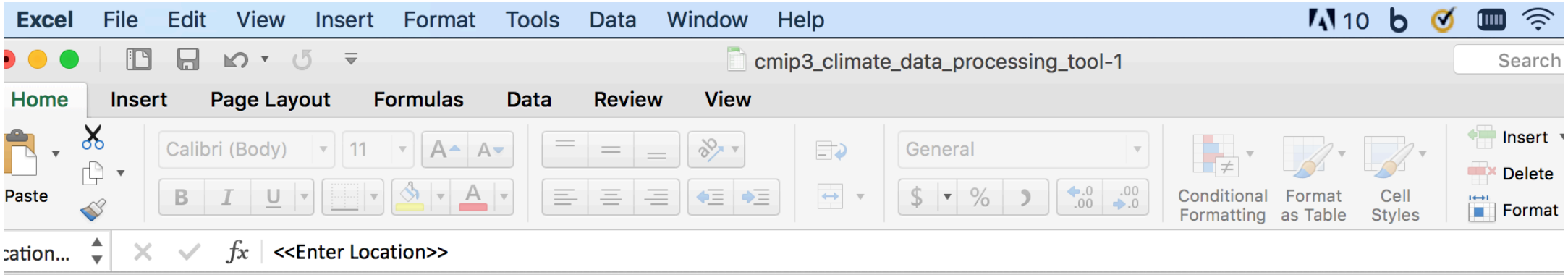
Web Page

fhwaguidance.cfm

Visit page

An example…


https://www.fhwa.dot.gov/environment/sustainability/resilience/tools/
Go
5 captures
26 Jan 2017 – 26 Apr 2017
DEC
JAN 26 2017
MAR
2016
2018
About this capture
U.S. Department of Transportation
Federal Highway Administration
About Programs Resources Briefing Room Contact Search FHWA
Office of Planning, Environment, & Realty (HEP)
Planning Environment Real Estate
HEP Events Guidance Publications Glossary Awards Contacts
Sustainability
Resilience
Adaptation Framework
Ongoing & Current Research
Policy & Guidance
Publications
Case Studies
Pilots
Tools
Webinars
Workshops & Peer Exchanges
Sustainable Highways Initiative
Energy and Emissions
Newsletter
Contacts
FHWA → Environment → Sustainability → Resilience
Tools
Climate Change Adaptation
Sensitivity Matrix (Excel file) (2015) - Spreadsheet tool that documents the sensitivity of roads, bridges, airports, ports, pipelines, and rail to 11 climate impacts.
Guide to Assessing Criticality in Transportation Adaptation Planning (also in PDF, 949KB) (2015) - Discusses common challenges associated with assessing criticality, options for defining criticality and identifying scope, and the process of applying criteria and ranking assets.
CMIP Climate Data Processing Tool (2016) User Guide (also in PDF, 2.5MB), Tool for CMIP3 Data (Excel file, 23.7MB), Tool for CMIP5 Data (Excel file, 60.9MB) - Spreadsheet tool that processes raw climate model outputs from the World Climate Research Programme's Coupled Model Intercomparison Project (CMIP) CMIP3 and CMIP5 databases into relevant statistics for transportation planners, including changes in the frequency of very hot days and extreme precipitation events that may affect transportation infrastructure and services by the middle and end of the century.
Vulnerability Assessment Scoring Tool (2015) User Guide (also in PDF, 1.5MB), Tool (Excel file, 2MB) - Spreadsheet tool that guides the user through conducting a quantitative, indicator-based vulnerability screen. Intended for agencies assessing how components of their transportation system may be vulnerable to climate stressors.



U.S. DOT Coupled Model Intercomparison Project (CMIP) Climate Data Processing Tool

CMIP3 Processing Tool

Directions

1. Follow all steps in the **User's Guide** to request and save all data. Note that it is very important that all data have been saved in the correct folders.

2. Answer the following questions about the data you downloaded.

• *Confirm that you have downloaded data from CMIP3. (If you downloaded data from CMIP5, use the CMIP5 version of this tool.)*	CMIP3
• *Describe the location you selected (for your reference only)*	*<<Enter Location>>*
• *How many climate models did you select? (i.e., how many boxes did you check in Step 2.6?)*	1
• *How many grid cells did you download?*	1
• *In addition to the projected values for each variable, the tool will provide the range in values given a certain confidence interval. What confidence interval would you like to see?*	95%
• *Describe the emissions scenario(s) you chose (for your reference only)*	*<<Enter Scenario>>*

Output Variables

What types of output variables do you want to generate?

If you are only interested in precipitation-related outputs, for example, uncheck the "Temperature Variables" box. This will save processing time.

- [x] Temperature Variables
- [x] Precipitation Variables

3. Verify that you have followed the instructions and saved data in the correct locations.

data.json


```
{
  "accessLevel": "public",
  "accrualPeriodicity": "R/P1M",
  "bureauCode": [
    "005:55"
  ],
  "contactPoint": {
    "@type": "vcard:Contact",
    "fn": "Michael Sutton",
    "hasEmail": "mailto:michael.sutton@stl.usda.gov"
  },
  "description": "Data provides current information regarding single f
  previously owned properties are for sale by public auction or other
  "distribution": [
    {
      "@type": "dcat:Distribution",
      "description": "Resale Properties data file",
      "downloadURL": "https://www.sc.egov.usda.gov/data/files/SFHFCLDa
      "mediaType": "text/csv",
      "title": "Resale Properties - Foreclosure"
    },
    {
      "@type": "dcat:Distribution",
      "downloadURL": "https://homesales.gov/homesales/mainAction.do",
      "mediaType": "text/html",
      "title": "Properties for Sale"
    }
  ],
  "identifier": "USDA-RD-0004"
```


DATA CATALOG

/ Datasets | Organizations

/ / Department of Agriculture / Rural Development, Department ... | Submit Data Story | Report Data Issue

USDA Rural Development

Rural Development

Topics

Agriculture

Publisher

Rural Development, Department of Agriculture

Contact

Michael Sutton

Share on Social Sites

Google+

Twitter

Facebook

USDA Rural Development Resale Properties - Foreclosure

Metadata Updated: February 7, 2017

Data provides current information regarding single family homes and ranches for sale by the U.S. Federal Government. These previously owned properties are for sale by public auction or other method depending on the property.

Access & Use Information

Public: This dataset is intended for public access and use.
License: Creative Commons CCZero

Downloads & Resources

Resale Properties - Foreclosure 25 views
Resale Properties data file
Open With | Download

Properties for Sale 33 views
mainAction.do
Visit page

Landing Page
Visit page

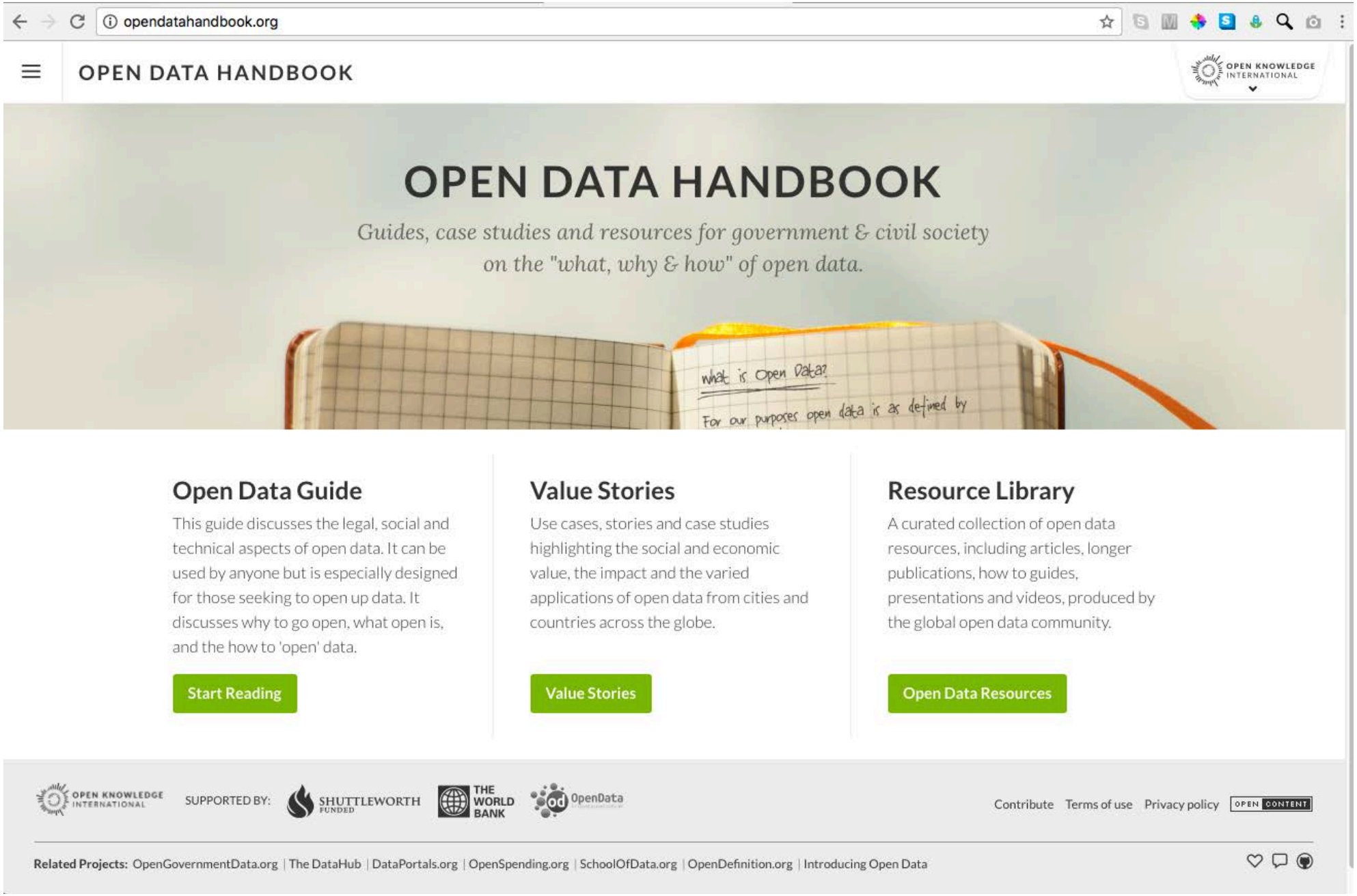
Dates

Metadata Created Date	May 17, 2014
Metadata Updated Date	February 7, 2017
Data Update Frequency	R/P1M

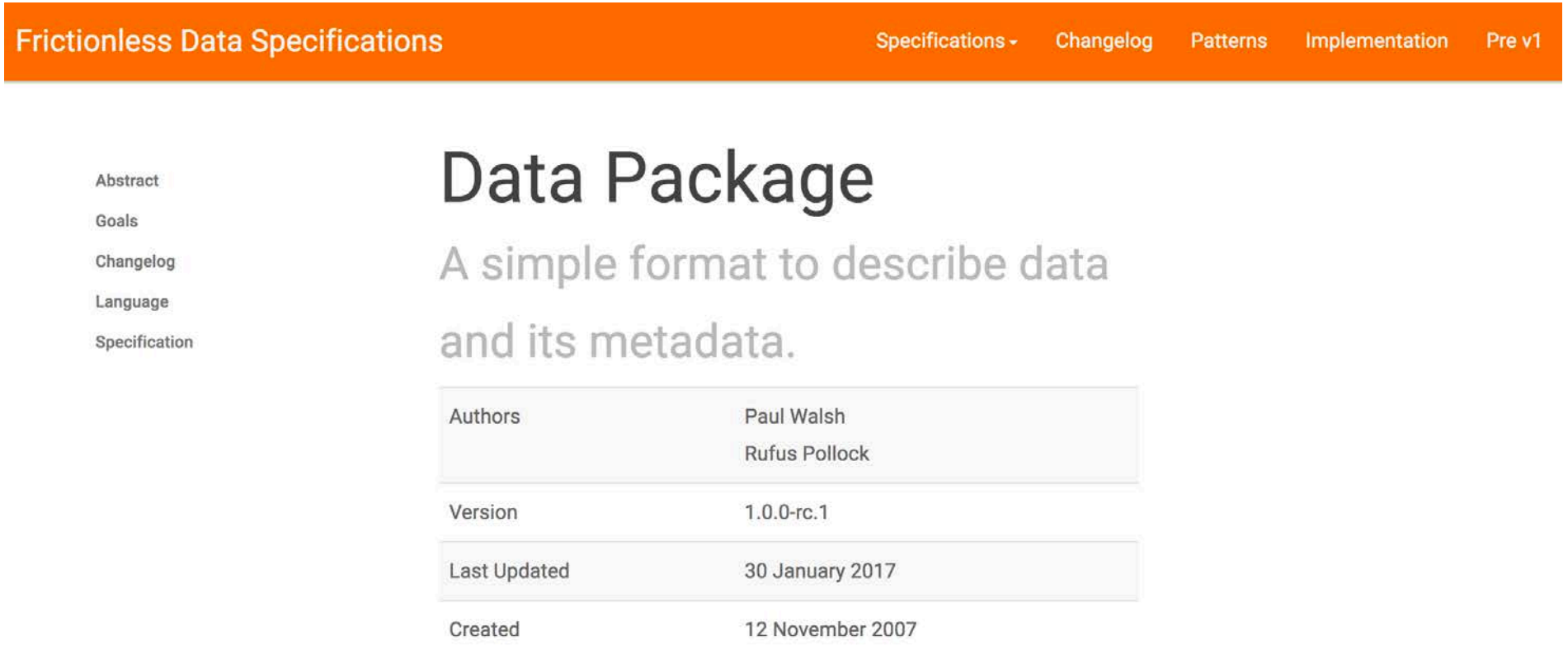
Metadata Source

Data.json Metadata
Download Metadata

Harvested from USDA JSON


opendatahandbook.org
OPEN DATA HANDBOOK
OPEN KNOWLEDGE INTERNATIONAL
OPEN DATA HANDBOOK
Guides, case studies and resources for government & civil society on the "what, why & how" of open data.
what is Open Data?
For our purposes open data is as defined by
Open Data Guide
This guide discusses the legal, social and technical aspects of open data. It can be used by anyone but is especially designed for those seeking to open up data. It discusses why to go open, what open is, and the how to 'open' data.
Start Reading
Value Stories
Use cases, stories and case studies highlighting the social and economic value, the impact and the varied applications of open data from cities and countries across the globe.
Value Stories
Resource Library
A curated collection of open data resources, including articles, longer publications, how to guides, presentations and videos, produced by the global open data community.
Open Data Resources
OPEN KNOWLEDGE INTERNATIONAL
SUPPORTED BY:
SHUTTLEWORTH FUNDED
THE WORLD BANK
OpenData
Contribute
Terms of use
Privacy policy
OPEN CONTENT
Related Projects: OpenGovernmentData.org | The DataHub | DataPortals.org | OpenSpending.org | SchoolOfData.org | OpenDefinition.org | Introducing Open Data

Frictionless Data

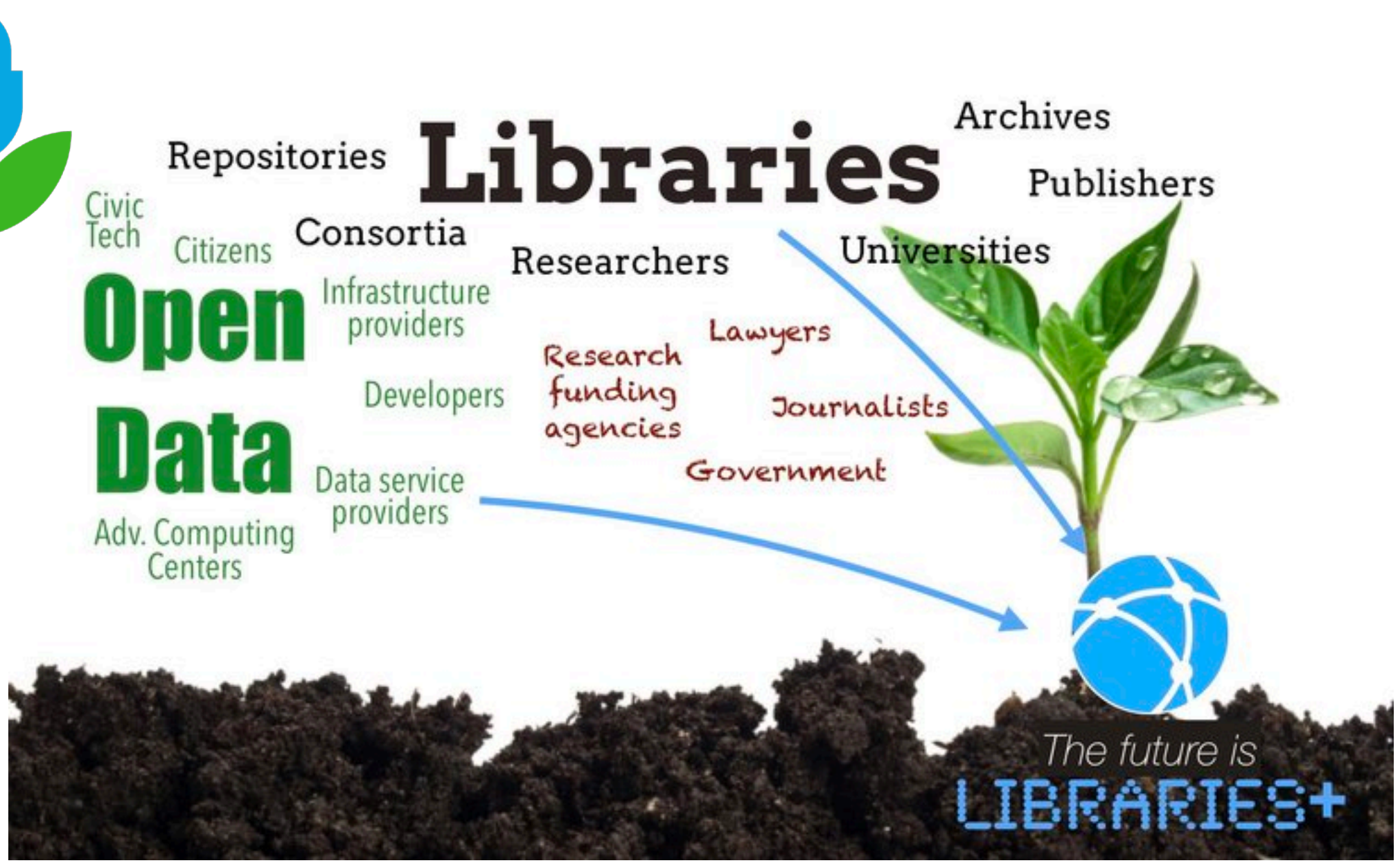


http://specs.frictionlessdata.io/data-package/

Libraries + Network

One path on the long trail

From bucket brigade to sustainable network


Repositories
Libraries
Archives
Publishers
Civic Tech
Citizens
Consortia
Researchers
Universities
Open
Infrastructure providers
Lawyers
Research funding agencies
Developers
Journalists
Data
Government
Data service providers
Adv. Computing Centers
The future is
LIBRARIES+

A popular Government, without popular information, or a means of acquiring it, is but a prologue to a farce or a tragedy; or perhaps both… a people who mean to be their own governors must arm themselves with the power which knowledge gives.

James Madison, Letter to W.T. Barry August 4, 1822

Experiment & Pilot!

thank you

Extra slides after this

value & vulnerability

How to measure?

Legal vulnerability

- What laws are in place to require the collection, safety, and redistribution of these data?
- What enforcement mechanisms back up those laws?

Technical vulnerability

- Is this data stewarded as part of a **partnership** with an institution, organization or government that will likely keep it going for a while?
- Is this **data already duplicated** somewhere?
  - If so, how vulnerable is the backup location?
- Is the data in a proprietary, difficult to duplicate **format**

Political vulnerability

- Do we have good reason to expect that **this dataset** is particularly targeted by hostile political forces?
- Do we have good reason to expect that **the agency and/or unit** who produce and maintain this data are particularly targeted by hostile political forces, or that their funding stream is unstable?
- Are the **primary users** associated with this dataset targets of hostile political forces?

Uses & uniqueness of the data

- Is this dataset the **only source** of data on the topic it covers?
- Is it **especially vital** to a particularly vital activity?

Data Rescue Events

Event Paths

- Seed and Sort webpages
- Data Refuge
  - Harvest
  - Check
  - Bag
  - Describe
- Storytelling
- Long trail

Laurie Allen 0

DataRefuge

Datasets Organizations Groups About

Search

Building refuge for federal climate & environmental data

DataRefuge helps to build refuge for federal data and supports climate and environmental research and advocacy. We are committed to fact-based arguments. DataRefuge preserves the facts we need at a time of ongoing climate change.

This site is one part of the project. The vast majority of the government information gathered through this project is available from the Internet Archive through the End of Term project. This data catalog is a place to store data that is difficult or impossible to harvest through web crawlers. Beginning to add difficult data to this websites is one of the tasks of the DataRescuePhilly event on January 13th and 14th.

Where is the data?

LIBRARIES+ NETWORK

Libraries+ Network Meeting
May 8-9, 2017 - Washington, D.C.

On May 8 and 9, representatives from research libraries, the open data community, government agencies, researchers, and other key stakeholders will be convening in Washington DC at New America to discuss and plan for the long-term preservation and access of federally produced data.

1860
1934

Libraries + Network https://libraries.network

- Association of Research Libraries
- California Digital Library
- Center for Open Science
- data.gov
- Dataverse
- Digital Libraries Federation
- HathiTrust
- ICPSR
- Internet Archive
- Library of Congress
- Mozilla Foundation
- National Archives and Records Administration
- National Oceanic and Atmospheric Association
- New America Foundation
- Pennsylvania State University
- Social Science Research Council
- Stanford Libraries
- Sunlight Foundation
- Temple University Libraries
- Union of Concerned Scientists
- University of Maryland Libraries
- University of Michigan Libraries
- University of Minnesota
- University of North Texas
- University of Pennsylvania