Tips for Managing and Organizing Your Research Materials

Andrew Creamer
Data Management & Sharing Specialist
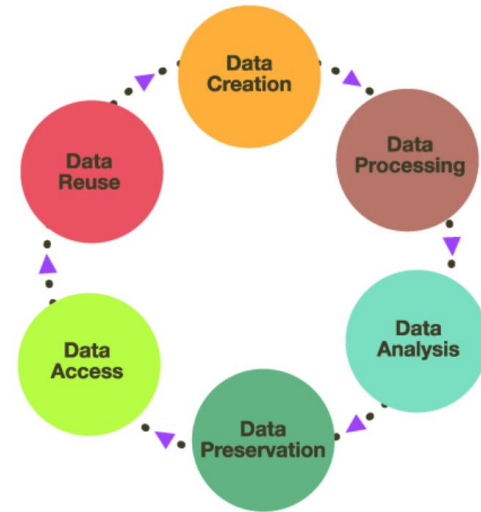*Brown University Library*

DOI: https://doi.org/10.26300/zh76-hk23

Data Stewardship Topics

- File naming
- Folder structure
- File formats
- Metadata for discovery & reuse
- Storage and backup
- Dissemination/sharing
- Rights and permissions
- Long-term preservation

Research data lifecycle

"Living" Data Management Plan (DMP)

Geographic

Maps, GIS, coordinates, locations, landmarks, etc.

Visual

Photographs, postcards, artwork, PNG, JPEG, GIF, glossy photos, print & digital photos, etc.

Audio

CDs, MP3s, WAV, recordings, broadcasts, songs, compositions, scores, etc.

Statistical

Datasets, surveys, census info., demographic, etc.

Video

Film, DVDs, YouTube, Vine, reel-to-reel, 8mm, etc.

Human Subjects

Demographics, ethnography, interviews, journals, etc.

Textual & Hypertextual

Citations, documentation of analog materials, digital scans, OCR text files, documents, spreadsheets, ephemera, publications, appendices, indices, archive of web pages, e.g., html

Authors of primers include archivists and data librarians who attended the 2018-2020 Specialized Data Curation Workshops presented by the Data Curation Network (IMLS RE-85-18-0040-18).

|	
Acrobat PDF Primer	Creators: Peace Ossom-Williamson, Nicole Contaxis, Margaret Lam and Adam Kriesberg Mentor: Jake Carlson
ATLAS.ti Primer	Creator: Margarita Corral Affiliated contributor: Hannah Hadley Mentor: Dave Fearon
Confocal Microscopy Image Primer	Creators: Susan Ivey, Amy Koshoffer, Gretchen Sneff and Huajin Wang Mentor: Lisa Johnston
Consent Forms Primer	Creators: Shanda Hunt, Alicia Hofelich Mohr and Rachel Woodbrook
Databases Primer	Creator: Xuying Xin Mentor: Dave Fearon
Geodatabase Primer	Creators: Andrew Battista, Tom Brittnacher, Zenobie Garrett, Jennifer Moore and Carrie Pirmann Mentor: Mara Blake
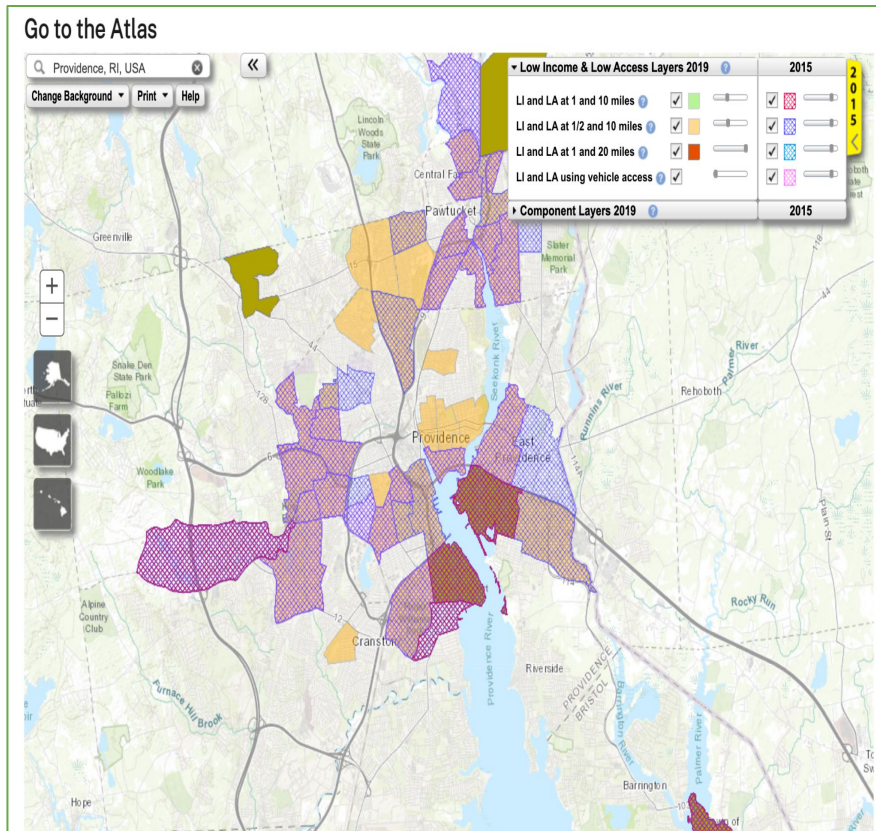
Monica is researching food deserts - how certain low-income neighborhoods have less access to buy fresh foods.

Example Case Study Food Desert Map

Monica's Current File Names

Before meeting with the GIS specialist Monica wants to update her files' names to come up with a consistent way of naming them and adding context that can help others to navigate through her data:

- Streets.shp
- Streets(1).shp
- groceries$.xlsx
- neighborhoods.shp
- census data.shp
- map-final.pdf
- map-final-final-final.pdf

Recommended Practices

- Date in YYYYMMDD format
- Include a version #
- Remove spaces and special characters other than - or _
- Creating a sequence? Add sequence # to file, e.g., if 100s of files start with 001 up to 999. 1000s? Then start with 0001-9999, etc.

Create a file name convention

- 20210629-pvd-ri-2020-street-map-v01.shp
- 20210629-pvd-ri-2020-grocery-geocodes-v02.xlsx
- 20210629-pvd-ri-2020-census-tracts-map.shp
- 20210629-pvd-ri-2020-food-desert-map-v12.pdf

Folder Structure

My Drive > PVD-Food-Deserts-Summer-Research-Project-2021 > Summer-Research-Project-Files-2021 ▾

Name	Owner	Last modified ↑	File size
📁 2020-PVD-Grocery-Store-Data	me	1:22 PM me	—
📁 2020-PVD-Census-Tract-Data	me	1:24 PM me	—
📁 2020-PVD-Food-Desert-Map-Layers	me	1:25 PM me	—
📁 2020-PVD-Food-Desert-Map-Images	me	1:25 PM me	—

Subfolder: 2020-PVD-Grocery-Store-Data

- File of grocery store addresses
- File of the geocodes (latitude/longitude coordinates) corresponding to the addresses
- Map layer for grocery locations

Subfolder: 2020-PVD-Food-Desert-Map-Layers

Need to keep all the map layers (the files that together comprise the map and need to be kept together):

- File of grocery store locations
- File of Providence city and neighborhood boundaries
- File of Providence streets
- File of census tracts of low income neighborhoods in PVD, RI

Subfolder Title: Versions/Attic

20210621-pvd-ri-2020-food-desert-map-v01

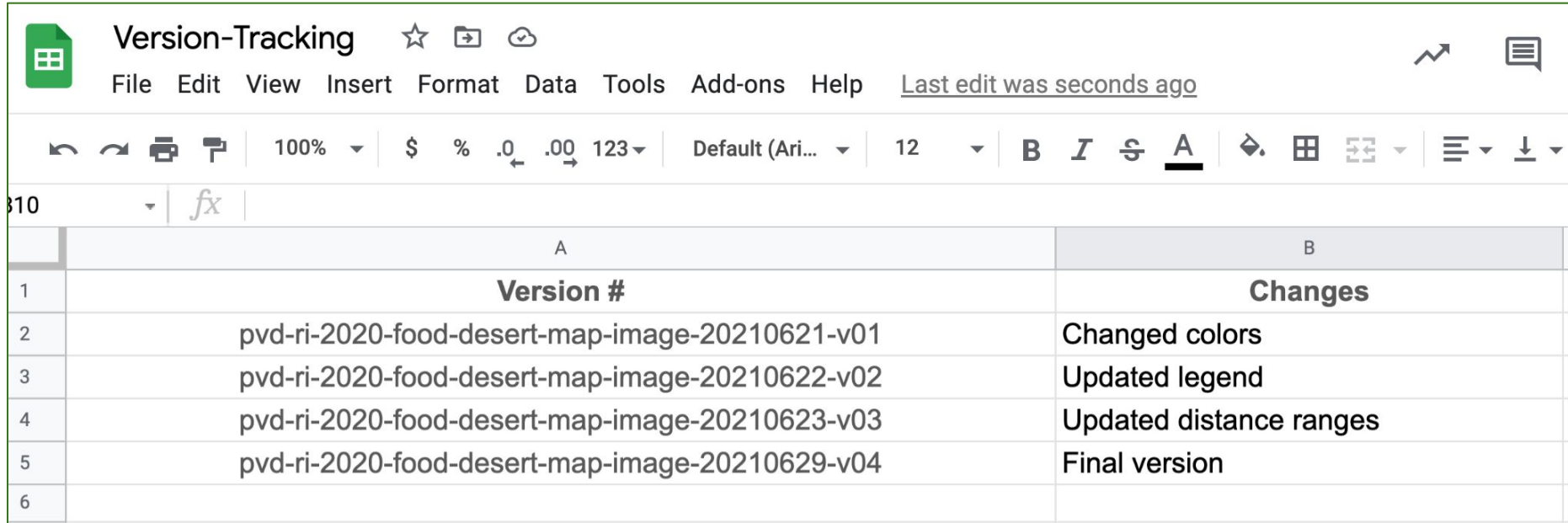20210621-pvd-ri-2020-food-desert-map-v02

20210621-pvd-ri-2020-food-desert-map-v03

20210621-pvd-ri-2020-food-desert-map-v04

My Drive > PVD-Food-Deserts-Summer-Research-... > Summer-Research-Project-... > 2020-PVD-Food-Desert-Map-Ima... ▾

Name	Owner	Last modified ↑	File size
📁 Versions	me	1:26 PM me	—
📁 Final	me	1:26 PM me	—

Add a "Readme" in a folder file directory

	A	B
	Version #	**Changes**
2	pvd-ri-2020-food-desert-map-image-20210621-v01	Changed colors
3	pvd-ri-2020-food-desert-map-image-20210622-v02	Updated legend
4	pvd-ri-2020-food-desert-map-image-20210623-v03	Updated distance ranges
5	pvd-ri-2020-food-desert-map-image-20210629-v04	Final version
6		

Quantitative Surveys

Violet is researching attitudes and perceptions of climate change and has created a quantitative survey and would like to analyze the demographics of respondents and their response data..

- Files of the drafts and final version of **survey protocol and instrument**
- Files of the **spreadsheet of survey participant demographics and responses**
- File of **protocol and code used for the analysis** of responses (e.g., SPSS/R)
- Files of draft and final versions of the **aggregate results** (results of analysis)
- Files of **draft and final versions of graph/figures** reporting the results (images of the results)

Qualitative Interviews

Violet is also researching attitudes and perceptions of climate change and has interviewed residents of a coastal town and would like to analyze the major themes.

- Files of drafts and final versions of **structured interview protocol questions**
- Files of the **audio recordings and transcripts** of participant responses
- File of the **codes used for thematic analysis** of responses (nVivo/Atlas.ti)

Archives File Naming Conventions

001_manhattan_project_bx07_fol012_fieldnote_20170401.txt	thesis_chapter_01_20170401_v2.pdf

Sequence

archive

Box

Folder

Content

Date

Project

Date

Version

Unpacking FAIR: Machine-readable/actionable

"The intent is that these [principles] may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles *put specific emphasis on enhancing the ability of machines* to automatically find and use the data, in addition to supporting its reuse by individuals.

Wilkinson, MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18

Unpacking FAIR: Versatility - from systems and platforms to projects and datasets

"The FAIR Guiding Principles describe distinct considerations for contemporary data publishing environments with respect to supporting both manual and automated deposition, exploration, sharing, and reuse. While there have been a number of recent, often domain-focused publications advocating for specific improvements in practices relating to data management and archival, **FAIR differs in that it describes concise, domain-independent, high-level principles that can be applied to a wide range of scholarly outputs**."

Wilkinson, MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18

Unpacking FAIR: Centers metadata and their role in discovery, use, provenance, and attribution

"Throughout the Principles, we use the phrase '(meta)data' in cases where the Principle should be applied to both metadata and data."

URI: "*A Uniform Resource Identifier (URI) is a unique sequence of characters that identifies a resource used by web technologies*"
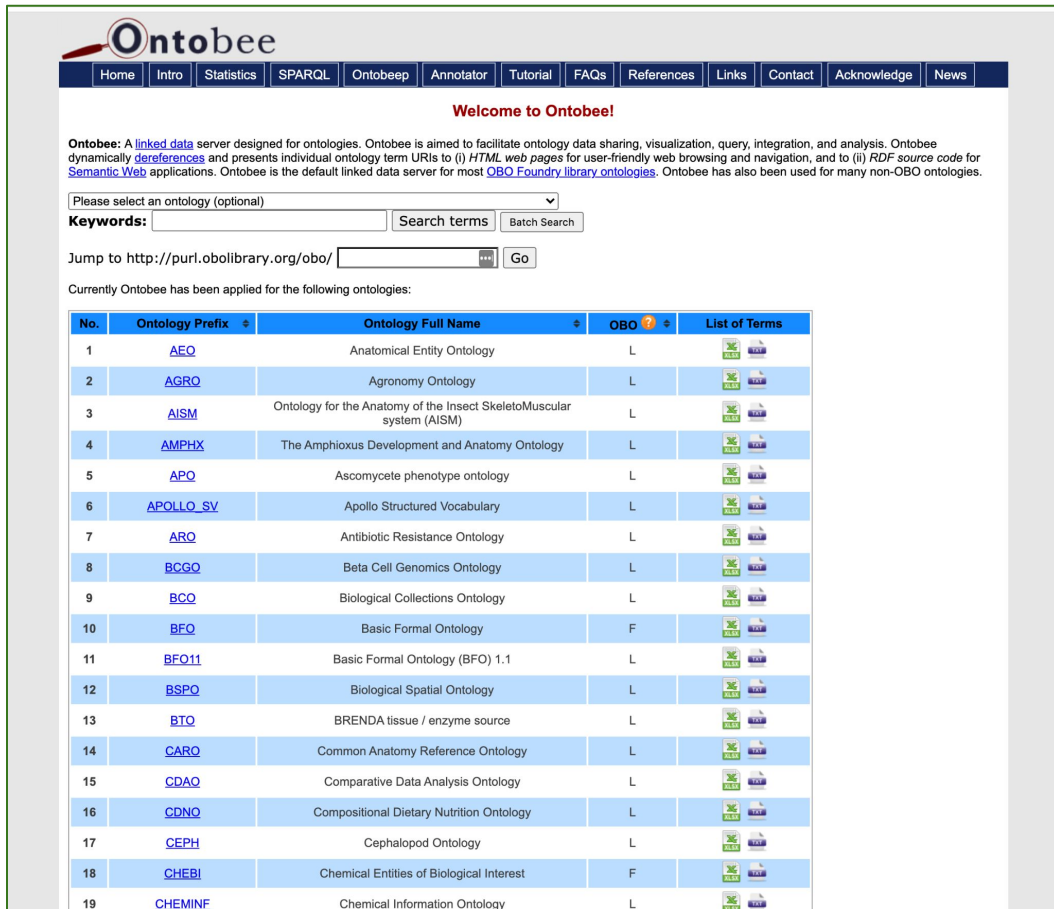
Vocabulary: "a computer-readable file that captures terms, their URIs, and descriptions

Ontology: "a vocabulary with hierarchies, meaningful relations among concepts"

Semantic Model: "meaning of entities and relations in the dataset" -- model in which the ontology and vocabulary class and URIs are embedded and made machine readable

-- GO Fair.n.d. *Fairification Process*
https://www.go-fair.org/fair-principles/fairification-process/

Ontobee

| Home | Intro | Statistics | SPARQL | Ontobeep | Annotator | Tutorial | FAQs | References | Links | Contact | Acknowledge | News |

Welcome to Ontobee!

Ontobee: A linked data server designed for ontologies. Ontobee is aimed to facilitate ontology data sharing, visualization, query, integration, and analysis. Ontobee dynamically dereferences and presents individual ontology term URIs to (i) *HTML web pages* for user-friendly web browsing and navigation, and to (ii) *RDF source code* for Semantic Web applications. Ontobee is the default linked data server for most OBO Foundry library ontologies. Ontobee has also been used for many non-OBO ontologies.

Please select an ontology (optional)
Keywords: [   ] [Search terms] [Batch Search]

Jump to http://purl.obolibrary.org/obo/ [   ] [...] [Go]

Currently Ontobee has been applied for the following ontologies:

No.	Ontology Prefix	Ontology Full Name	OBO	List of Terms
1	AEO	Anatomical Entity Ontology	L	
2	AGRO	Agronomy Ontology	L	
3	AISM	Ontology for the Anatomy of the Insect SkeletoMuscular system (AISM)	L	
4	AMPHX	The Amphioxus Development and Anatomy Ontology	L	
5	APO	Ascomycete phenotype ontology	L	
6	APOLLO_SV	Apollo Structured Vocabulary	L	
7	ARO	Antibiotic Resistance Ontology	L	
8	BCGO	Beta Cell Genomics Ontology	L	
9	BCO	Biological Collections Ontology	L	
10	BFO	Basic Formal Ontology	F	
11	BFO11	Basic Formal Ontology (BFO) 1.1	L	
12	BSPO	Biological Spatial Ontology	L	
13	BTO	BRENDA tissue / enzyme source	L	
14	CARO	Common Anatomy Reference Ontology	L	
15	CDAO	Comparative Data Analysis Ontology	L	
16	CDNO	Compositional Dietary Nutrition Ontology	L	
17	CEPH	Cephalopod Ontology	L	
18	CHEBI	Chemical Entities of Biological Interest	F	
19	CHEMINF	Chemical Information Ontology	L	

Example Using ISA-TAB

"The FAIR and Linked Data Principles specify that all "entities" (i.e., things that are not textual or numerical values) must be named using a resolvable URI. Semantic models often contain multiple terms from existing ontologies and vocabularies."

-- Rodríguez-Iglesias A, Rodríguez-González A, Irvine AG, Sesma A, Urban M, Hammond-Kosack KE, Wilkinson MD. Publishing FAIR Data: An Exemplar Methodology Utilizing PHI-Base. Front Plant Sci. 2016 May 12;7:641. doi: 10.3389/fpls.2016.00641.

ISA Overview

https://www.isacommons.org/
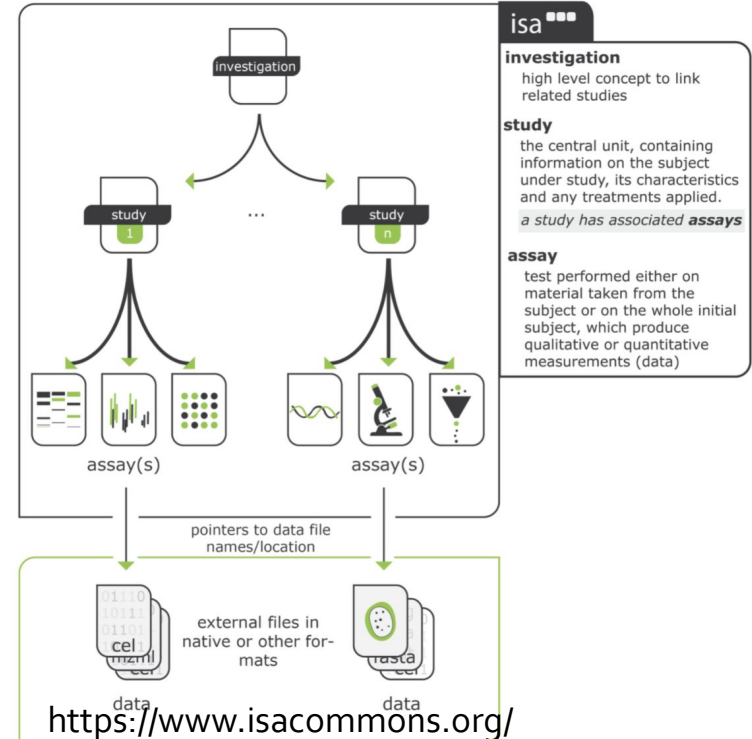
Example Linked ISA-TAB Entity Ontologies

ONTOLOGY SOURCE REFERENCE
Term Source Name   "CHEBI"   "CHMO"   "NCIT"   "OBI" "NCBITAXON"   "BTO"   "EFO" "UO"
Term Source File   "https://www.ebi.ac.uk/metabolights/" "http://data.bioontology.org/ontologies/CHEBI" "http://data.bioontology.org/ontologies/CHMO"   "http://data.bioontology.org/ontologies/NCIT" "http://data.bioontology.org/ontologies/OBI"   "http://data.bioontology.org/ontologies/NCBITAXON" "http://data.bioontology.org/ontologies/BTO"   "http://data.bioontology.org/ontologies/EFO" "http://data.bioontology.org/ontologies/UO"
Term Source Version   "1" "86"   "12"   "34"   "23"   "4"   "26"   "119" "43"
Term Source Description "Metabolights Ontology" "Chemical Entities of Biological Interest Ontology" "Chemical Methods Ontology" "National Cancer Institute Thesaurus" "Ontology for Biomedical Investigations"   "National Center for Biotechnology Information (NCBI) Organismal Classification" "BRENDA Tissue and Enzyme Source Ontology"   "Experimental Factor Ontology" "Units of Measurement Ontology"

Defined entities URIs

Study Assay Measurement
        Type Term
                Accession Number

        "http://purl.obolibrary.org/obo/OBI_
        0000366"

Class: metabolite profiling assay

Term IRI: http://purl.obolibrary.org/obo/OBI_0000366
Definition: An assay that detects and identifies chemical entities resulting from biochemical and cellular metabolism

Annotations

- **definition editor:** Philippe Rocca-Serra
- **alternative term:** metabolomic assay; metabolite assay
- **definition source:** OBI
- **example of usage:** Metabolite profiling of human colon carcinoma - deregulation of TCA cycle and amino acid turnover. Mol Cancer. 2008 Sep 18;7(1):72. PMID: 18799019
- **has curation status:** metadata complete

Class Hierarchy

Thing
+ entity
    + occurrent
        + process
            + planned process
                + assay
                    - Bernoulli trial
                    + imaging assay
                    - radioactivity detection
                    - protein-protein interaction detection assay
                    + transcription profiling assay
                    - DNA sequence feature detection assay
                    - DNA sequence variation detection assay
                    + analyte assay
                    - mass measurement assay
                    + intra cellular electrophysiology recording assay
                    + extracellular electrophysiology recording assay
                    + mass spectrometry assay
                    - copy number variation profiling assay
                    + protein expression profiling assay
                    - NMR spectroscopy assay
                    - hematology assay
                    - survival assessment assay
                    - glucose tolerance test
                    + single-nucleotide-resolution nucleic acid structure mapping assay
                    + handedness assay
                    + binding assay
                    + efficacy of epitope intervention experiment
                    more...
                    - metabolite profiling assay

Defined vocabulary attribute and value URI

Study Design Descriptor
    Benzoxazine
        "http://purl.obolibrary.org/obo/CHEBI_46969"

Findable: Enabling Discovery

F1. (Meta)data are assigned a *globally unique and persistent identifier*

F2. Data are described with **rich metadata**

F3. Metadata clearly and explicitly *include the identifier of the data* they describe

F4. (Meta)data are registered or indexed in a *searchable resource*

Accessible

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

Reusable

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

Example Case study metadata

- Use a standard (if one exists) - (e.g., ISO 19115)
- Ontology/vocabularies used
- Details on provenance -e.g., census data used
- Income ranges used USD/per capita or per household
- Year of data represented
- Date map created
- Distance measurement (miles/km)

- Coordinate standard used for longitude and latitude/geocodes
- Geonames for controlled list of names of cities, states
- ISO codes for location (country codes)
- Creator ORCiD
- Rights and license information
- Subject headings
- Grant ID
- Digital Object Identifier (DOI)

Help with data discovery and attribution

Dataset Persistent ID ❔	doi:10.7910/DVN/QHKPAI
Publication Date ❔	2020-08-03
Title ❔	San Diego business location data, 1958, from Polk's San Diego City Directory
Author ❔	Marlow, Thomas (Brown University) - ORCID: 0000-0003-3989-6775 Frickel, Scott (Brown University) - ORCID: 0000-0002-7368-885X
Contact ❔	Use email button above to contact. Brown Library Research Data Management Services (Brown University)
Description ❔	This dataset contains parsed, extracted, and geocoded historical business and manufacturing data from Polk's San Diego City Directory using the open source directoreadr software. Images used for data extraction can be found at https://www.sandiego.gov/digitalarchives/collections/specialcollections/citydirectories. (2020-06-01)
Subject ❔	Earth and Environmental Sciences
Keyword ❔	Environmental health (MeSH) https://id.nlm.nih.gov/mesh/D004782.html Cities (MeSH) https://id.nlm.nih.gov/mesh/D002947.html Industry (MeSH) https://id.nlm.nih.gov/mesh/D007221.html Manufacturing industry (MeSH) http://id.nlm.nih.gov/mesh/D066192 California--San Diego (FAST) http://id.worldcat.org/fast/1205232 California--San Diego County (FAST) http://id.worldcat.org/fast/1204290
Related Publication ❔	Polk's San Diego (San Diego County, Calif.) City Directory 1958 Including La Jolla .R. L. POLK & COo, Publishers 120 East 8th Street, Los Angeles 14, Calif. purl: 1136988096 http://www.worldcat.org/oclc/1136988096
Language ❔	English
Grant Information ❔	Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH): P42 ES013660-14S1

"Metadata and data should be easy to find for both humans and computers"

{"@context":"http://schema.org","@type":"Dataset","@id":"https://doi.org/10.7910/DVN/QHKPAI","identifier":"https://doi.org/10.7910/DVN/QHKPAI","name":"San Diego business location data, 1958, from Polk's San Diego City Directory","creator":[{"name":"Marlow, Thomas","affiliation":"Brown University","@id":"https://orcid.org/0000-0003-3989-6775","identifier":"https://orcid.org/0000-0003-3989-6775"},{"name":"Frickel, Scott","affiliation":"Brown University","@id":"https://orcid.org/0000-0002-7368-885X","identifier":"https://orcid.org/0000-0002-7368-885X"}],"author": [{"name":"Marlow, Thomas","affiliation":"Brown University","@id":"https://orcid.org/0000-0003-3989-6775","identifier":"https://orcid.org/0000-0003-3989-6775"},{"name":"Frickel, Scott","affiliation":"Brown University","@id":"https://orcid.org/0000-0002-7368-885X","identifier":"https://orcid.org/0000-0002-7368-885X"}],"datePublished":"2020-08-03","dateModified":"2020-08-14","version":"1","description":["This dataset contains parsed, extracted, and geocoded historical business and manufacturing data from Polk's San Diego City Directory using the open source directoreadr software. Images used for data extraction can be found at https://www.sandiego.gov/digitalarchives/collections/specialcollections/citydirectories."],"keywords":["Earth and Environmental Sciences","Environmental health","Cities","Industry","Manufacturing industry","California--San Diego","California--San Diego County"],"citation": [{"@type":"CreativeWork","text":"Polk's San Diego (San Diego County, Calif.) City Directory 1958 Including La Jolla .R. L. POLK & COo, Publishers 120 East 8th Street, Los Angeles 14, Calif.","@id":"http://www.worldcat.org/oclc/1136988096","identifier":"http://www.worldcat.org/oclc/1136988096"}],"license": {"@type":"Dataset","text":"CC0","url":"https://creativecommons.org/publicdomain/zero/1.0/"},"includedInDataCatalog":{"@type":"DataCatalog","name":"Harvard Dataverse","url":"https://dataverse.harvard.edu"},"publisher":{"@type":"Organization","name":"Harvard Dataverse"},"provider":{"@type":"Organization","name":"Harvard Dataverse"},"funder": [{"@type":"Organization","name":"Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH)"}],"distribution": [{"@type":"DataDownload","name":"FOutput_geocoded.tab","fileFormat":"text/tab-separated-values","contentSize":2790348,"@id":"https://doi.org/10.7910/DVN/QHKPAI/MMLJ0M","identifier":"https://doi.org/10.7910/DVN/QHKPAI/MMLJ0M","contentUrl":"https://dataverse.harvard.edu/api/access/datafile/4002447"}]}

Weaponizing FAIR

"The FAIR Principles are aspirational, in that they do not strictly define how to achieve a state of "FAIRness"; rather they describe a continuum of features, attributes, and behaviors that move a digital resource closer to that goal. Despite their rapid community uptake, the question of how the FAIR Principles should be implemented has been prone to diverse interpretation. Some resource providers claim to be "already FAIR" or "to enable FAIRness" – statements that currently cannot be objectively evaluated. These manifold interpretations of the FAIR Principles are counterproductive, posing a risk of fragmentation and confusion in a manner antithetical to their intended purpose (or worse, being ignored entirely due to a lack of formal clarity). This state of affairs is somewhat ironic in that, because FAIR speaks to machine-actionable operations, FAIR digital objects should therefore be amenable to unambiguous and indeed, completely automated forms of validation and evaluation."

Wilkinson, M.D., Dumontier, M., Sansone, SA. et al. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sci Data* 6, 174 (2019). https://doi.org/10.1038/s41597-019-0184-5

FAIR Metrics

FAIRsharing.org

FAIR Maturity Indicators

Search tests and collec SEARCH

FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.

We are back online ! Thank you for your patience ! If you notice any unexpected failures in the tests, please report them to mark.wilkinson@upm.es

Import MI Tests

Import Maturity Indicators Tests as YAML smartAPI interface annotation

Get started

Create collections

Assemble Maturity Indicators Tests into community centered collections

Get started

Evaluate resources

Evaluate resources FAIRness against Collections of Maturity Indicator Tests

Get started

Philosophy of FAIR testing

Evaluating FAIRness is a controversial issue! We (the Maturity Indicator Authoring Group) feel that these concerns can be eased through bringing more clarity regarding the "philosophy" behind FAIRness testing.

FAIR Cookbook and Recipes

https://fairplus.github.io/the-fair-cookbook/

Repository Finder

Find a repository to upload your data.

Repository Finder, a pilot project of the Enabling FAIR Data Project led by the American Geophysical Union (AGU) in partnership with DataCite and the Earth, space and environment sciences community, can help you find an appropriate repository to deposit your research data. The tool is hosted by DataCite and queries the re3data registry of research data repositories.

As part of the FAIRsFAIR project, which aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle, the Repository Finder is extended to query for repositories relevant to FAIRsFAIR Project.

Search re3data for a repository to upload your data

| Type to search... | Search |

or

See the repositories in re3data that meet the criteria of the Enabling FAIR Data Project.

See the repositories in re3data that meet the criteria of the FAIRsFAIR Project.

Spreadsheet tips

- Have all column headers in a single row across the top
- One data point/variable per column and per cell (e.g., you wouldn't mix celsius and fahrenheit data or mix column headers "temperature and humidity"…)
- Avoid empty cells or use a consistent fixed code for defining any missing values or null values
- Consider using data validation to prevent data entry errors common with free text
- Create another column field if you need to add notes/annotations- do not mix data and annotations in the cells
- Use NA for empty cells
- Beware Excel and dates...

Metadata for Use

Details about the data in the project and methods used to collect and analyze data in the project

- Protocols
- Readme
- Codebook and data dictionary

Codebook and Data Dictionary

Section	Variable Name	Variable Label	Value	Value description	Total sample (N=2228)			Main analytical sample (complete dyads) (N=1872)			Note (info on recoding decision rules and/or version changes 1.3 and 1.4) Comments regarding version changes in red.
					N	Percent	Cum N	N	Percent	Cum N	
	cpait_CaseNo	Household identifier	String	Range from 99100100 to 99115281	2228	100.00%	2228	1872	100.00%	1872	
	cpait_AlzId	Alzheimer ID patient	String	Range from 100100 to 115281	2228	100.00%	2228	1872	100.00%	1872	
	cpait_RESULTCATEGORY	Resultcategory (complete or missing)	String	Complete	2228	100.00%	2228	1872	100.00%	1872	
	REC_cpait_partnerint	Recode: patient has corresponding care partner interview	0	No care partner interview	356	15.98%	356	0	0.00%	0	Indicates whether there is a matcing care partner interview for the patient record.
			1	Care partner interview available	1872	84.02%	2228	1872	100.00%	1872	
Logistics	cpait_rmpend_date	End date of relationship mapping variable	continuous	range: 21017-21273	2228	100.00%	2228	1872	100.00%	1872	
	cpait_resultcode_date	Date of most recent/last interview outcome	continuous	range: 21017-21274	2228	100.00%	2228	1872	100.00%	1872	
	REC_cpait_Qversion	Recode: indicates which questionnaire version was administered	1	Old (1.3, before 10/08)	1013	45.47%	1013	869	46.42%	869	REC_cpait_Qversion indicates whether the participant received the old questionnaire, the new questionnaire or both. It's coded as follows: 1 (old/1.3) if the END date was before 10/08 (interview finished before the change date) coded 2 (new/1.4) if the START date was on or after 10/08 (interview started on or after the change date) coded 3 (restart) if none of the above applied (interview did not finish before 10/08 AND did not start after, so interview has started before 10/08 but finished after 10/08). cpait_rmpend_date is used for the start date (date that relationship mapping section was finished) and cpait_resultcode_date is used for the end date (date of most recent/last interview outcome). REC_cpait_Qversion_specific uses the first variable as a basis but specifies (as much as possible) for all the restarts where in the questionnaire they restarted, these restart points were determined by looking at missing data patterns. It will be helpful to use these variables to see which questionnaire was administered to patients, specifically when you look at variables which are present both in 1.3 and 1.4 but for which the wording or prompts changed (which may influence how participants answered the question).
			2	New (1.4, on/after 10/08)	1186	53.23%	2199	979	52.30%	1848	
			3	Both: restart that spans 10/08	29	1.30%	2228	24	1.28%	1872	
	REC_cpait_Qversion_specific	Recode: indicates questionnaire version administered, specifies restarts	1	Old (1.3, before 10/08)	1013	45.47%	1013	869	46.42%	869	
			2	New (1.4, on/after 10/08)	1186	53.23%	2199	979	52.30%	1848	
			3	Restart: start of qualitative story	18	0.81%	2217	15	0.80%	1863	
			4	Restart: during qualitative story	2	0.09%	2219	0	0.00%	1863	
			5	Restart: start/during planning for the future	3	0.13%	2222	3	0.16%	1866	
			6	Restart: before ECOG, after planning for the future	1	0.04%	2223	1	0.05%	1867	
			7	Restart: during ECOG	1	0.04%	2224	1	0.05%	1868	
			8	Restart: start of hypothetical risky treatment	2	0.09%	2226	2	0.11%	1870	
			9	Restart: during hypothetical risky treatment	1	0.04%	2227	1	0.05%	1871	
			10	Restart: after hypothetical risky treatment, means no different from 1.4	1	0.04%	2228	1	0.05%	1872	
	cpait_Relationship	Relationship to care partner	1	Husband	623	27.96%	623	545	29.11%	545	
			2	Wife	1165	52.29%	1788	1058	56.52%	1603	
			3	Significant other, partner	57	2.56%	1845	44	2.35%	1647	
			4	Father	0	0.00%	1845	0	0.00%	1647	
			5	Mother	0	0.00%	1845	0	0.00%	1647	
			6	Grandfather	0	0.00%	1845	0	0.00%	1647	
			7	Grandmother	0	0.00%	1845	0	0.00%	1647	
			8	Aunt	0	0.00%	1845	0	0.00%	1647	
			9	Uncle	1	0.04%	1846	0	0.00%	1647	
			10	Brother	6	0.27%	1852	2	0.11%	1649	
			11	Sister	17	0.76%	1869	11	0.59%	1660	
			12	Son	64	2.87%	1933	34	1.82%	1694	
			13	Daughter	203	9.11%	2136	126	6.73%	1820	
			14	Son-in-law	0	0.00%	2136	0	0.00%	1820	
			15	Daughter-in-law	12	0.54%	2148	7	0.37%	1827	
ip mapping			16	Father-in-law	0	0.00%	2148	0	0.00%	1827	
			17	Mother-in-law	0	0.00%	2148	0	0.00%	1827	
			18	Friend	40	1.80%	2188	25	1.34%	1852	
			97	Other	38	1.71%	2226	19	1.01%	1871	
			.d	DK	1	0.04%	2227	0	0.00%	1871	
			.r	Refused	1	0.04%	2228	1	0.05%	1872	
	cpait_RelationshipOS	Relationship to care partner, other specify	NULL	MISSING	2190	98.29%	2190	1853	98.99%	1853	
			String	Range from Caregiver to stepfather	38	1.71%	2228	19	1.01%	1872	
			1	Spouse / significant other	1850	83.03%	1850	1651	88.19%	1651	
			2	Children (incl in-law & step)	281	12.61%	2131	168	8.97%	1819	

Analysis Code

Feedback   Discover...   I Want to...   Login →

Code for "Trends in Non-Invasive and Invasive Mechanical Ventilation among Medicare Beneficiaries at the End of Life"

Overview   Full Metadata →

Year:
2020
Contributor:
Sullivan, Donald S. (Author)
Kim, Hyosin (Author)
Gozalo, Pedro (Author)
Bunker, Jennifer (Author)
Teno, Joan M. (Author)
Subject:
non-invasive mechanical ventilation
NIMV
Invasive Mechanical Ventilation
IMV
DOI
https://doi.org/10.26300/pxh2-3790

Files
MODS   👁 ⬇
ZIP   👁 ⬇

Views
Archive Contents   👁 ⬀
Archive Default   👁 ⬀

Description

Abstract:
The objective of this study was to assess document the trends use of IMV and NIMV among persons with hospitalizations in the last 30 days of life. A secondary goal is to examine how if differs among persons with dementia, Cancer, COPD, and CHF.

Notes:
This research was funded by the National Institute on Aging of the National Institutes of Health under award number R56AG063748, 2P01AG027296-11

Content   Views ▾ | Files ▾

Code for "Trends in Non-Invasive and Invasive Mechanical Ventilation among Medicare Beneficiaries at the End of Life"

Links
Download

Contents
- code/Code_Sharing_JAMA_Sullivan_Et_Al_2020.docx
- code/Code_Sharing_JAMA_Sullivan_Et_Al_2020.pdf

Original Statistical Analysis:

Notes:
Stata code used for analyses of file created by programmer.

```
* set up variables / analytic files

* create outcome variables,  2000-2017 / mv_noninv & mv_inv: no data b/t 2000 and 2009
d mvnoninv_30 mvinv_30 mv_noninv mv_inv
tab hkyear mvnoninv_30, m
tab hkyear mvinv_30, m
tab mvnoninv_30 mvinv_30, m


* create a "vent" variable
gen vent = 0
replace vent = 1 if mvnoninv_30==1 & mvinv_30==0
replace vent = 2 if mvnoninv_30==1 & mvinv_30==1
replace vent = 3 if mvnoninv_30==0 & mvinv_30==1


label define ventstat 0 "No MV" 1 "NIMV" 2 "Both" 3 "IMV"
label values vent ventstat
```

Example Data Dictionary

	A	B	
1	Count	Senate Ballot File Variable Name	Senate Ballot Variable Description
2	1	YEAR	The year in which the state legislature voted on the Senate election
3	2	CHAMBER	State legislatures were bicameral with two chambers; House stands for the state House of Repr
4	3	SEAT	Senate terms are six years in length and for this time period started on March 4th and ended six
5	4	ELECTTYPE	Senate elections for expiring terms are lableled general (g) and elections held to fill unexpected
6	5	SEPARATE/JOINT	joint indicates the vote was taken joint session. Voters in joint session were typically but not alwa
7	6	JOINTBALLOT#	For this data, ballot numbering begins with the first joint session ballot labeled 1.
8	7	SENCANDIDATE	Name of candidate for U.S. Senate that state legislator cast a ballot for; NV, absent, not voting,
9	8	ST LEGISLATURE VOTER	Member of state legislature recorded in state or house journal as voting or absent.
10	9	PARTY	Populist
11	10	DISTRICT	Specific district of state legislator; state legislatures apportioned differently across states either
12	11	COUNTY	Specific district of state legislator; state legislatures apportioned differently across states either
13			
14			
15			
16		State Completed	Notes
17		California	no county data
18		Delaware	no district data

Data Storage and Backup Best Practice

3-2-1 Rule

- Local copy of data files
- 2nd local copy on *separate* drive
- Remote/cloud copy of data

*Check data risk classification levels for human subjects data

https://ithelp.brown.edu/kb/articles/decide-how-to-store-and-share-files#comparison

---

Decide How to Store and Share Files

2019-07-01 | 💬 0 Comments | in Campus File Storage

Subscribe 🔊   Download PDF 📄

Several services at Brown allow you to share and store files. This guide will let you compare the options and decide which one(s) are right for you.

Table of Content

- Easy to Use
  - Google Drive
  - Dropbox
  - BrownBox
- Especially for Research
  - Campus File Storage / Research
  - Research Data Storage (RData)
  - Stronghold
- General File Storage
  - Campus File Storage / Department File Services
  - Campus File Storage / Home Folder
- Comparison table

Rights and Permissions

- Raw data not generally considered eligible for copyright

- Creative compilations of data are considered eligible for copyright (databases, visualizations, figures)

- Creator owns/share rights with co-creators (up and until one were to ever sign rights away)

- Creators can thus assign a license (spelling out set of permissions) for users so that users do not need to contact the creator for those permissions (e.g., especially helpful for uses that one generally may consider falling under Fair Use Doctrine like reproducing a work or disseminating copies) or reusing/repurposing data to create a derivative work

Licensing your work for access and use by others

CREATIVE COMMONS LICENSES

	COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
PUBLIC DOMAIN	✓	✗	✓	✓	✓
CC BY	✓	✓	✓	✓	✓
CC BY-SA	✓	✓	✓	✓	✗
CC BY-ND	✓	✓	✓	✗	✗
CC BY-NC	✓	✓	✗	✓	✓
CC BY-NC-SA	✓	✓	✗	✓	✗
CC BY-NC-ND	✓	✓	✗	✗	✗

✓ You can redistribute (copy, publish, display, communicate, etc.)

✓ You have to attribute the original work

✓ You can use the work commercially

✓ You can modify and adapt the original work

✓ You can choose license type for your adaptations of the work.

URIs for rights and permissions

```
<mods:accessCondition xmlns:xlink="http://www.w3.org/1999/xlink"  type="rights statement"
xlink:href="http://rightsstatements.org/vocab/InC/1.0/">In Copyright</mods:accessCondition>

<mods:accessCondition xmlns:xlink="http://www.w3.org/1999/xlink" type="use and reproduction"
xlink:href="http://www.gnu.org/licenses/gpl-3.0.en.html">GNU General Public License
3.0</mods:accessCondition>
----------------------------------------------------------------------------------------------------------------------------------
<mods:accessCondition type="rightsstatement"
xlink:href="https://rightsstatements.org/page/NoC-US/1.0/?language=en">No Copyright
</mods:accessCondition>

<mods:accessCondition xmlns:xlink="http://www.w3.org/1999/xlink" type="use and reproduction"
xlink:href="https://creativecommons.org/publicdomain/zero/1.0/">CC0 Public
Domain</mods:accessCondition>
```

repository.library.brown.edu

BROWN UNIVERSITY LIBRARY

Brown Digital Repository   Feedback   Discover...   I Want to...   Login →

Featured collections...

Data for Publications

Herbarium

Opening the Archives: Documenting U.S.-Brazil Relations, 1960s-80s

Prints, Drawings and Watercolors from the Anne S.K. Brown Military Collection

Summer Research Symposium

Vietnam Veterans Archive

Discover...

▶ **Uploaded Content**

Content added by scholars through self-service upload

▶ **Theses & Dissertations**

Brown University dissertations since 2008

▶ **Deposited Collections**

Scholar-created research collections, research data, posters, proceedings, and

I want to...

▶ **Upload something**

Uploaded items can be kept private, shared at Brown, or shared with the public.

▶ **Digitize something**

Digital Production Services can assist with digitization of materials in support of scholarship, research, and teaching.

▶ **Use an API**

The BDR offers a complete set of Application Programming Interfaces (APIs) for using content in your own applications.

▶ **Assign a DOI**

Digital Object Identifiers (DOIs) help publishers and scholars cite content stored in the BDR.

▶ **Consult on a project**

The Center for Digital Scholarship works with scholars on digital projects across disciplines.

▶ **Read FAQ**

Frequently Asked Questions may already answer your BDR-related question.

Longitudinal Manufacturing Database for Rhode Island Establishments, 1953 - 2016

Overview   **Full Metadata →**

Content

Views ▾ | Files ▾

Title
Longitudinal Manufacturing Database for Rhode Island Establishments, 1953 - 2016

Contributors
Marlow, Thomas (Creator)
Frickel, Scott (Creator)

Doi
10.26300/wfpn-k614

Date Created
2018

Abstract
This file contains the location of manufacturing facilities in Rhode Island between the years 1953 and 2012. This data was collected from digitized manufacturing directories using the open source software Georeg. For details of this data collection process see the Berenbaum et al. (bdr:841108 and bdr:841109). Berenbaum, D., D. Deighan, T. Marlow, A. Lee, S. Frickel, and M. Howison. "Mining Spatio-Temporal Data on Industrialization from Historical Registries." Journal of Environmental Informatics. DOI: 10.3808/jei.201700381

Longitudinal Manufacturing Database for Rhode Island Establishments, 1953 - 2016

Links

Download

README

```
TITLE: Longitudinal Manufacturing Database for Rhode Island Establishments, 1953 - 2016

DESCRIPTION:
This file contains the location of manufacturing facilities in Rhode Island between the years 1953 and 2012

Berenbaum, D., D. Deighan, T. Marlow, A. Lee, S. Frickel, and M. Howison. In Press. "Mining Spatio-Temporal

FILENAME: "RI_mfgs-full_data.csv"

Significant cleaning of the data occurred with the software R and Python after data collection. Most signifi

FIELDS:
"id" – Integer; Unique identifier for establishments. IDs were generated for every row in the data.

"combined_id" – Integer; Identifier for unique combinations of company name and street name. IDs were crea

"street_id" – Integer; Identifier for unique street names. IDs were created using the "dedupe" library in F

"coname_id" – Integer; Identifier for unique company names. IDs were created using the "dedupe" library in

"year" – Integer; Year data on a facility was collected.

"coname" – Character; String containing the name of the company.

"street" – Character; String containing the collected street address. This is NOT necessarily the exact add

"city" – Character; String containing the city location of the manufacturing facility.

"SIC" – Integer; Two digit standard industrial classification (SIC) code. The field "sic_name" translates t
```

Selection and Appraisal

What data will be shared?
What data needs to be preserved and for how long?
Who will have access?
When will they have access?
Where will the data be accessible?
How will they access the data?
Any restrictions?

Preserving Project

Santiago. Joaquín Insausti y Fernández Galeano, Javier (2020). Archivos Digitales Queer: Cartografías Digitales de las Redes Transnacionales LGBTQ en Latinoamérica a través del Archivo de Robert Roth. *Moléculas Malucas*.
https://www.moleculasmalucas.com/post/archivos-digitales-queer

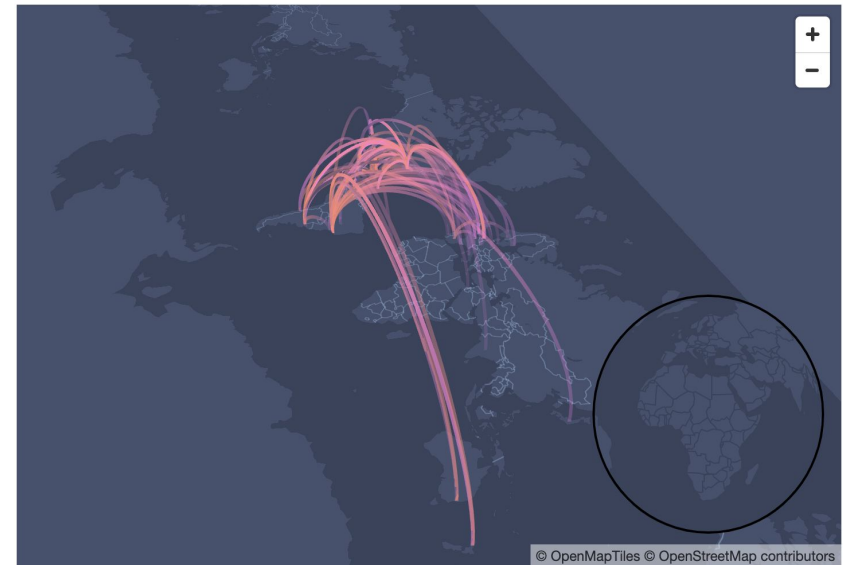**Redes Transnacionales LGBTQ en Latinoamérica**
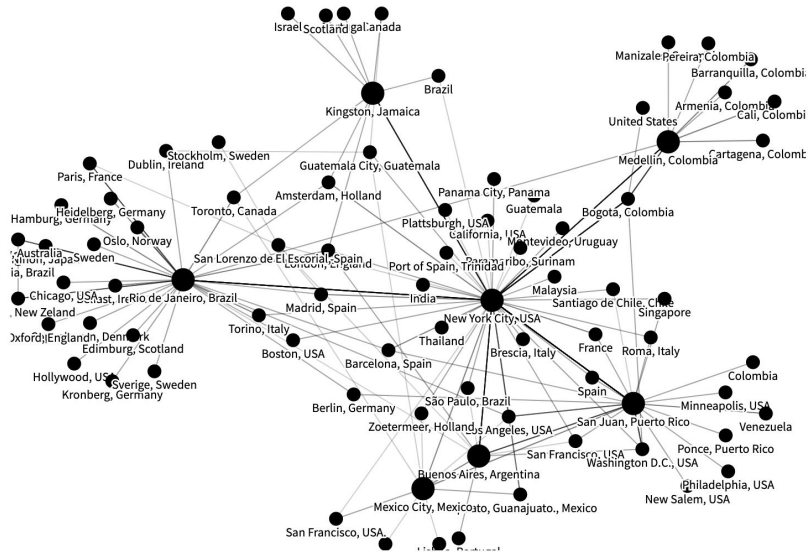
Source: Elaboracíon propia en base a la Robert Roth collection at Cornell University Library.

✳ A Flourish network chart

Redes transnacionales queer en Latinoamérica.
Cartas obrantes en el archivo Roth, segun destinatario y remitente.

© OpenMapTiles © OpenStreetMap contributors

Source: Roberth Roth Collection at Cornell University Library.

Thank you

Contact Me

Andrew Creamer

andrew_creamer@brown.edu