

Informationist Perspectives on Data Reuse: Lessons Learned from
Developing a Spatial Approach for Toxic Transferal from Industrial and
Vacant Land Uses to Green Infrastructure

Andrew Creamer, MSLIS, MAEd
Scientific Data Management Specialist
Brown University Library
Brown Superfund Research Program (SRP)

Galen Newman, PhD

PI, Community Engagement Core
Associate Professor, Landscape
Architecture and Urban Planning
Texas A&M University
gnewman@arch.tamu.edu

Jennifer A. Horney

Co-PI, Community Engagement
Core
Professor, Epidemiology
University of Delaware
horney@udel.edu

Scott Frickel, PhD

PI, Community Engagement
Core
Professor, Sociology
Brown University
scott_frickel@brown.edu

Andrew Creamer

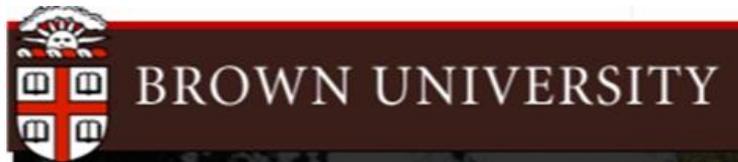
Science Data Specialist
Brown University
andrew_creamer@brown.edu

Keith Pezzoli, PhD

PI, Community Engagement
and Research Translation
Professor, Urban Studies and
Planning
UC San Diego
kpezzoli@sdsc.edu

Ilya Zaslavsky, PhD

Director, Spatial Information
Systems Laboratory, SDSC
UC San Diego
zaslavsk@sdsc.edu



Partnering with libraries: The Research Informationist

Davidoff F, Florance V. The informationist: a new health profession? *Ann Intern Med*. 2000 Jun 20;132(12):996-8. doi: 10.7326/0003-4819-132-12-200006200-00012. PMID: 10858185.

Rankin, J. A., Grefsheim, S. F., & Canto, C. C. (2008). The emerging informationist specialty: a systematic review of the literature. *Journal of the Medical Library Association : JMLA*, 96(3), 194–206.
<https://doi.org/10.3163/1536-5050.96.3.005>

Grants and Funding: Extramural Programs (EP)

[Home](#) > [EP Home](#)

Awards for NLM Administrative Supplements for Informationist Services in NIH-funded Research Projects (Admin Supp)

(See [//www.nlm.nih.gov/ep/AdminSupp.html](http://www.nlm.nih.gov/ep/AdminSupp.html))

Informationists are information specialists, usually health sciences librarians, who have graduate training and practical experience that provides them with disciplinary background in biomedical, behavioral or biological sciences and in library and information sciences/informatics. Their cross training provides informationists with a unique perspective on the acquisition, synthesis, management and use of information in research. Informationists work as team members with research scientists and health professionals, and are sometimes called in-context information specialists for that reason. The awards bring informationists into research settings and measure the value of their contributions to the research. The Supplement provides funding for up to 2 years to an active NIH-funded researcher, in order to bring needed information expertise into the research team.

Fiscal Years

- [2017](#)
- [2015](#)
- [2014](#)
- [2012](#)

F.A.I.R. Data-Related Informationist Collaborations

Data Documentation

- Metadata standards/Metadata application profiles
- Controlled vocabularies
- Ontologies
- Persistent identifiers (DOIs/ORCiDs/RRIDs)
- Authority control
- Provenance, Versioning, Rights & Licenses

Research Data Management (RDM)

- Data formats/documentation/organization
- Data dissemination/attribution
- Data archiving

Reproducibility/Methods replication/Data Reuse/Data Repurposing

- Protocol registry/README
- Preservation

“Rich metadata” (contextual information) that includes identifiers

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” (Machine-readable)

```
{ "@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 & CO, 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/MMLJOM", "identifier": "https://doi.org/10.7910/DVN/QHKPAI/MMLJOM", "contentUrl": "https://dataverse.harvard.edu/api/access/datafile/4002447"}]}
```

Foundational RDM Practices Covered in DMPs

- Document your methods and use a metadata standard, where possible
- Save copy of files in open data formats
- Name files and folders using a naming convention and use it consistently
- Version files to keep track of latest and past versions (ISO8601 YYYYMMDD)
- Backup files: 3-2-1 rule (3 copies = 2 separate local - 1 remote/cloud)
- Secure participants' data (data classifications/encryption/limit access/PWs)
- Assign a license allowing for data reuse (Creative Commons)
- Deposit data and code in long-term repositories for public access (Re3data)
- Cite the location of data and code in publications for others to locate
- Plan for any files that need to be retained and preserved after a project ends

“Collections as Data”

“Collections as data development aims to encourage computational use of digitized and born digital collections”

“Shared documentation helps others find a path to doing the work”

“Collections as data development values interoperability”

“Data as well as the data that describe those data are considered in scope.”

*Always Already Computational: Collections as Data project team (2017; v2).
The Santa Barbara Statement on Collections as Data.
<https://collectionsasdata.github.io/statement/>*

Collections-based data: Visiting the archives

- Historical maps
- Image/text-based data
- Diaries
- Inventories
- Herbaria
- Industrial directories

COLLECTIONS EXPLORE CREATE SHARE EMBED PRINT HELP

MEDIA INFORMATION X > Part of New England. Related T

COLLECTION NAME:
JCB Map Collection

Record ▾

Accession Number:
12416

File Name:
12416-2

Call number:
*Z E585 1706 / 3-SIZE

Map title:
Part of New England.

Place of Publication:
London

Publisher:
Sold by John Thornton at ye Platt in the Minories. And by Will: Fisher at ye Posturn gate on Tower hill

Publication date:
[1706]

Map size height:
41.9 cm.

Map size width:
46.5 cm. [both pages]

Item description:
engraved map; following p. 18

Geographical description:
Chart of New England showing part of present-day Rhode Island, Connecticut, Long Island, Massachusetts, New Hampshire, and Maine. Cartographic elements include soundings, sea banks or shoals, degrees of latitude, rhumb lines, scale, and compass rose.

Original in the John Carter Brown Library at Brown University

Abies balsamea (L.) Mill.

Overview

[Full Metadata](#) ↗

Catalog No.

PBRU00052545

Family

Pinaceae

Collector

H. P. Sartwell

Date Collected

State/Province

New York

Country

United States of America

Locality

Files

[DWC](#)



[DIGITAL-NEGATIVE](#)



[JP2](#)



Content

[Views](#) ▾

[Files](#) ▾

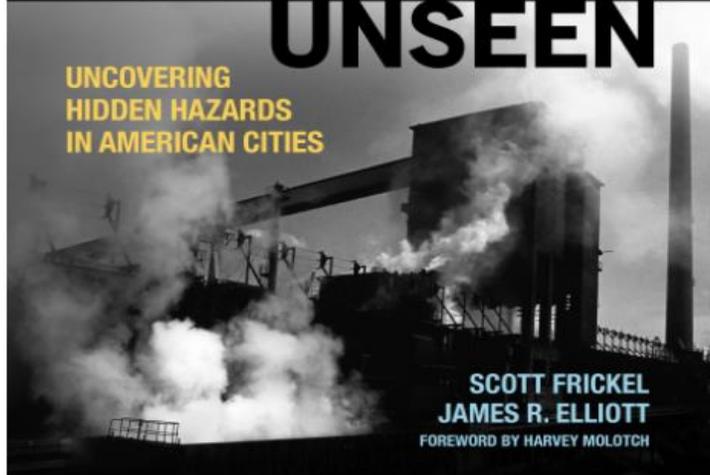


```
▼<sdr:SimpleDarwinRecordSet xmlns:dc="http://purl.org/dc/terms/" xmlns:dwc="http://rs.tdwg.org/dwc/terms/" xmlns:sdr="http://rs.tdwg.org/dwc/xsd/simpledarwincore/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://rs.tdwg.org/dwc/xsd/simpledarwincore/ http://rs.tdwg.org/dwc/xsd/tdwg_dwc_simple.xsd">
▼<sdr:SimpleDarwinRecord>
  <dwc:basisOfRecord>PreservedSpecimen</dwc:basisOfRecord>
  <dwc:catalogNumber>PBRU00052545</dwc:catalogNumber>
  <dwc:scientificName>Abies balsamea</dwc:scientificName>
  <dwc:scientificNameAuthorship>(L.) Mill.</dwc:scientificNameAuthorship>
  <dwc:kingdom>Plantae</dwc:kingdom>
  <dwc:phylum>Charophyta</dwc:phylum>
  <dwc:class>Equisetopsida</dwc:class>
  <dwc:order>Pinales</dwc:order>
  <dwc:family>Pinaceae</dwc:family>
  <dwc:genus>Abies</dwc:genus>
  <dwc:specificEpithet>balsamea</dwc:specificEpithet>
  <dwc:taxonID>815565</dwc:taxonID>
  <dwc:municipality>Penn Yan</dwc:municipality>
  <dwc:stateProvince>New York</dwc:stateProvince>
  <dwc:country>United States of America</dwc:country>
  <dwc:recordedBy>H. P. Sartwell</dwc:recordedBy>
  <dc:license>http://creativecommons.org/licenses/by/4.0/</dc:license>
  <dc:rightsHolder>Brown University</dc:rightsHolder>
  <dc:modified>2019-08-09 15:54:23</dc:modified>
  <dwc:collectionID>65888b1e-0a64-4307-95a2-7bfbdb917590</dwc:collectionID>
  <dwc:collectionCode>BRU</dwc:collectionCode>
  <dwc:occurrenceID>c0c8d7e2-d026-470a-8178-f2716af9df4b</dwc:occurrenceID>
  <dc:references>http://portal.neherbaria.org/portal/collections/individual/index.php?occid=1555378</dc:references>
  <dwc:institutionCode>Brown</dwc:institutionCode>
  <dwc:acceptedNameUsage>Abies balsamea (L.) Mill.</dwc:acceptedNameUsage>
</sdr:SimpleDarwinRecord>
</sdr:SimpleDarwinRecordSet>
```



SITES UNSEEN

UNCOVERING HIDDEN HAZARDS IN AMERICAN CITIES



SCOTT FRICKEL
JAMES R. ELLIOTT
FOREWORD BY HARVEY MOLOTCH

Frickel, S., Elliott, J., & Molotch, H. (2018). *Sites Unseen: Uncovering Hidden Hazards in American Cities*. New York: Russell Sage Foundation. doi:10.7758/9781610448734

INTERNET ARCHIVE

Providence Public Library Directories

Directories from the Providence Public Library.

Share
Favorite
RSS

ABOUT
COLLECTION

24 RESULTS

Sort by: **VIEWS** · TITLE · DATE PUBLISHED · CREATOR

Search this Collection

Metadata
 Text contents

Part Of
The Providence Public Library
American Libraries

Media Type

texts 24

Year

1990 1
 1985 1
 1980 1
 1978 1
 1976 1
 1972 1
[More](#)

Collection

American Libraries 24
 Providence Public Library Directories 24
 The Providence Public Library 24
 z4mom Favorites 9
 abarton08 Favorites 1
 annietaiter Favorites 1
[More](#)

Language

English 24

Polk's Providence (Providence County, R.I.)

1,315 views

Providence (Providence County, R.I.) City Directory

2,048 views

Providence Rhode Island City Directory

2,649 views

Providence (Providence County, R.I.) City Directory

1,829 views

Providence (Providence County, R.I.) City Directory

1,787 views

Polk's Providence (Providence County, R.I.)

1,132 views

Polk's Providence (Providence County, R.I.)

1,143 views

Polk's Providence (Providence County, R.I.)

1,387 views

Polk Providence Rhode Island City Directory

110 views

Polk's Providence (Providence County, R.I.)

1,418 views

Polk's Providence (Providence County, R.I.)

1,132 views

Polk's Providence (Providence County, R.I.)

1,143 views

Polk's Providence (Providence County, R.I.)

1,387 views

Polk Providence Rhode Island City Directory

110 views

Polk's Providence (Providence County, R.I.)

1,418 views

Mining Spatio-temporal Data on Industrialization from Historical Registries

David Berenbaum, Dwyer Deighan, Thomas Marlow, Ashley Lee, Scott Frickel, Mark Howison

Despite the growing availability of big data in many fields, historical data on socioenvironmental phenomena are often not available due to a lack of automated and scalable approaches for collecting, digitizing, and assembling them. We have developed a data-mining method for extracting tabulated, geocoded data from printed directories. While scanning and optical character recognition (OCR) can digitize printed text, these methods alone do not capture the structure of the underlying data. Our pipeline integrates both page layout analysis and OCR to extract tabular, geocoded data from structured text. We demonstrate the utility of this method by applying it to scanned manufacturing registries from Rhode Island that record 41 years of industrial land use. The resulting spatio-temporal data can be used for socioenvironmental analyses of industrialization at a resolution that was not previously possible. In particular, we find strong evidence for the dispersion of manufacturing from the urban core of Providence, the state's capital, along the Interstate 95 corridor to the north and south.

Subjects: **Computer Vision and Pattern Recognition (cs.CV)**; Information Retrieval (cs.IR)

Journal reference: Journal of Environmental Informatics 34(1): 28-34 (2019)

DOI: [10.3808/jei.201700381](https://doi.org/10.3808/jei.201700381)

Cite as: [arXiv:1612.00992](https://arxiv.org/abs/1612.00992) [cs.CV]

(or [arXiv:1612.00992v1](https://arxiv.org/abs/1612.00992v1) [cs.CV] for this version)

Submission history

From: Mark Howison [\[view email\]](#)

[v1] Sat, 3 Dec 2016 17:54:03 UTC (1,962 KB)

 [brown-ccv / directoreadr](#)

forked from [samwbell/directoreadr](#)

 [Code](#)

 [Pull requests](#)

 [Actions](#)

 [Projects](#)

 [Wiki](#)

 [Security](#)

 [Insights](#)

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

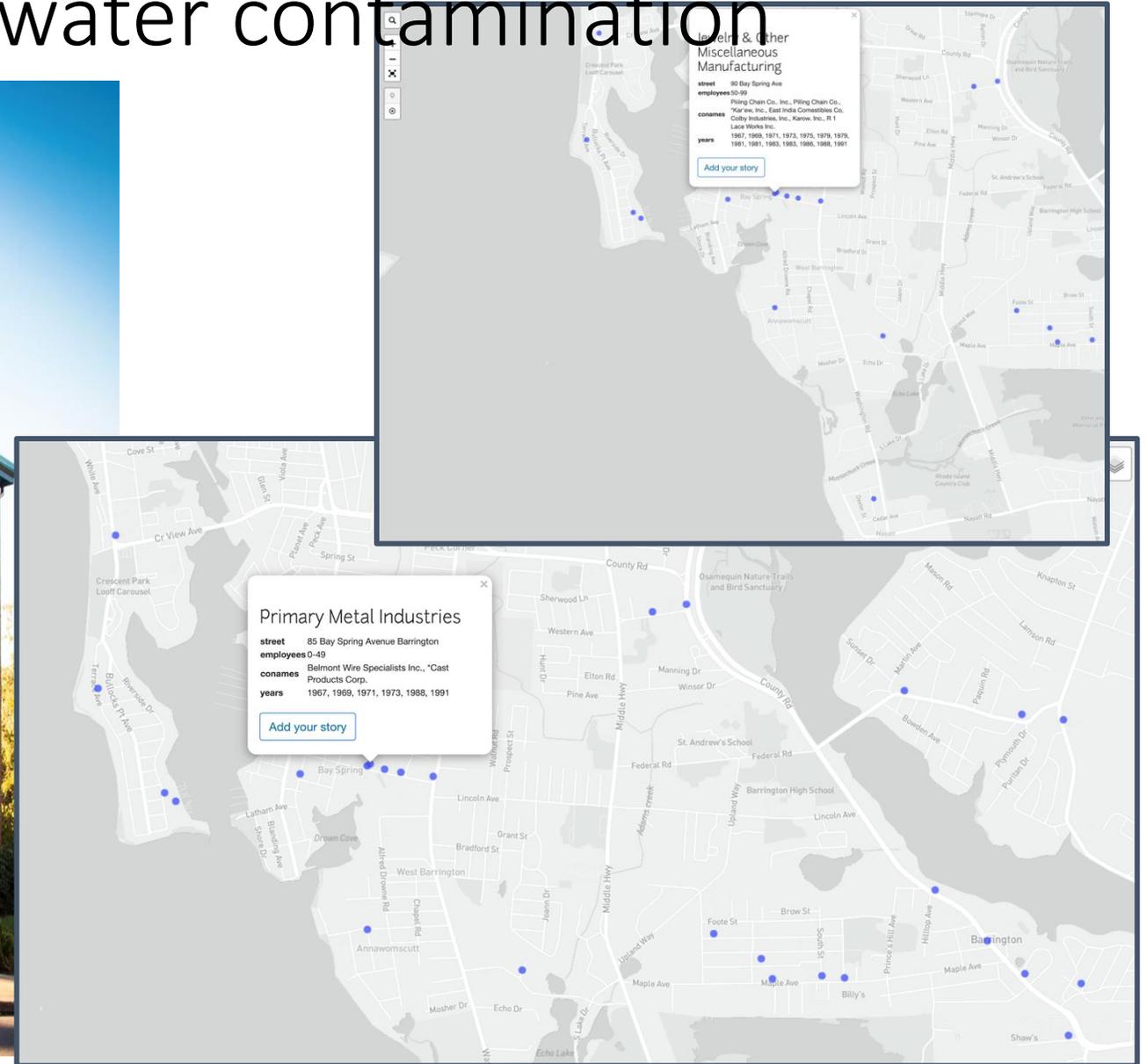
[Databases Primer](#)

Creator: Xuying Xin
Mentor: Dave Fearon

[Geodatabase Primer](#)

Creators: Andrew Battista, Tom Brittnacher, Zenobie Garrett, Jennifer Moore and Carrie Pirmann
Mentor: Mara Blake

Maps of Historical Industrial Activity and Brownfields Potential sites of soil and water contamination





Unearthing Providence



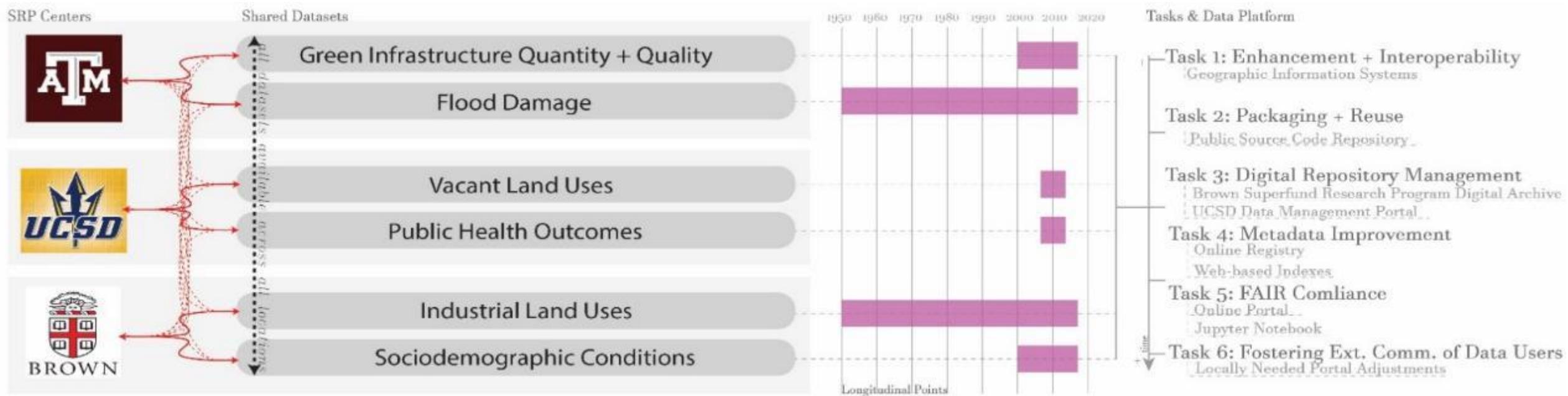
 Add a location

Learn about location [privacy](#)

A platform for historical data-driven analyses of industrial land use, story-mapping, and community organizing.

Recommendations for Integrated SRP Data: Outcomes of the SRP External Use Cases

- Title: **Developing a Spatial Approach for Toxic Transferal from Industrial and Vacant Land Uses to Green Infrastructure**
- Research Question:
 - *How do vacant and industrial land uses impact green infrastructure conditions often presumed to enhance community resilience and public health?*



Inputs and Actions

- Locations

- Harris County, TX
- San Diego County, CA
- State of Rhode Island

Dataset	Year	Source	Reference	Scale
Green Infrastructure	2016	https://data.tnris.org/	USGS	US Census tract
Vacant Addresses	2016	https://www.huduser.gov/portal/usps/index.html	HUD	US Census tract
Public Health (14 factors)	2016	https://www.cdc.gov/500cities/index.htm	CDC	US Census tract
Social Vulnerability	2016	https://svi.cdc.gov/data-and-tools-download.html	CDC	US Census tract
Industrial Land Uses	2016	Multiple (created from land use data)	Local	US Census tract
Flood Plain	2016	https://www.fema.gov/faq-details/GIS-Data	FEMA	US Census tract

San Diego business location data, 1958, from Polk's San Diego City Directory: <https://doi.org/10.7910/DVN/QHKPAI>
Vacant land time series (1986 – 2018), from San Diego Association of Governments, based on parcel-level data

[Search](#)

Information Search

Map Intersects Within

find a place

Esri, HERE, Garmin, NGA...

[Search Using MapExtent](#)

Bounding Box

Publication Date

Location Keyword

Original Keyword

AI Keywords

Cited Authors

Cited Organizations

Temporal Extent (text)

Temporal Extent (slider)

Metadata Properties

Metadata Type

Metadata Collection

http://datadiscoverystudio.org/

Filters

Results [By Relevance](#) 1,669,856 items [«](#) [<](#) [1](#) [of >10k Pages](#) [>](#)

Yasur Volcano Crater DEM, Republic of Vanuatu, Oct 2016
 Publication: 2019-07-22 Source: OpenTopography (U) Last Modified: 2020-06-18
 DEM of the crater of Yasur Volcano, Vanuatu
[Studio](#) [Get Data](#) [Item Details](#) [Add to Collection](#)

Palmerston North, Manawatu-Whanganui, New Zealand 2018
 Publication: 2020-02-27 Source: OpenTopography (U) Last Modified: 2020-06-18
 Lidar was captured for Palmerston North City Council by AAM New Zealand between August and September 2018. The dataset was generated by AAM New Zealand and their subcontractors. The survey area includes city of Palmerston North, Ashhurst, Longburn and the surrounding area. Data management and distribution is by Land Information New Zealand. Prepared DEM and DSM files are available through the LINZ Data Service: Palmerston North, Manawatu-Whanganui, New Zealand 2018 Digital Elevation Model Palmerston North, Manawatu-Whanganui, New Zealand 2018 Raster Tile Index
[Studio](#) [Get Data](#) [Item Details](#) [Add to Collection](#)

Christchurch and Ashley River, Canterbury, New Zealand 2018
 Publication: 2020-02-26 Source: OpenTopography (U) Last Modified: 2020-06-18
 Lidar was captured for Environment Canterbury Regional Council by Aerial Surveys between July 2018 to March 2019. The dataset was generated by Aerial Surveys and their subcontractors. The survey area includes Christchurch City and the Ashley River, as well as the surrounding area. Data management and distribution is by Land Information New Zealand. Prepared DEM and DSM files are available through the LINZ Data Service: Christchurch and Ashley River, Canterbury, New Zealand 2018 Digital Elevation Model Christchurch and Ashley River, Canterbury, New Zealand 2018 Digital Surface Model Christchurch and Ashley River, Canterbury, New Zealand 2018 Raster Tile Index
[Studio](#) [Get Data](#) [Item Details](#) [Add to Collection](#)

Huntly, Waikato, New Zealand 2015-2019
 Publication: 2020-02-10 Source: OpenTopography (U) Last Modified: 2020-06-18
 Lidar was captured for Waikato Regional Council by Aerial Surveys between February 2015 to January 2019. The dataset was generated by Aerial Surveys and their subcontractors. The survey areas include Huntly, Port Waikato, and the Hauraki Plains. Data management and distribution is by Land Information New Zealand.
[Studio](#) [Get Data](#) [Item Details](#) [Add to Collection](#)

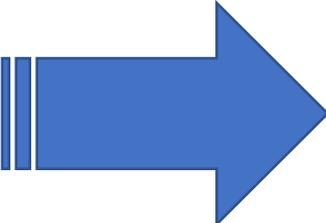
2019 Ridgecrest, CA Post-Earthquake Lidar Collection
 Publication: 2020-01-28 Source: OpenTopography (U) Last Modified: 2020-06-18
 Airborne lidar survey of surface ruptures and ground failure features associated with the 4 and 5 July 2019 Ridgecrest Earthquake sequence. For additional information about this dataset see: Hudnut, K.W., B. Brooks, K. Scharer, J.L. Hernandez, T.E. Dawson, M.E. Oskin, R. Arrowsmith, C.A. Goulet, K. Blake, M.L. Boggs, S. Bork, C.L. Glennie, J.C. Fernandez-Diaz, A. Singhania, D. Hauser, S. Sorhus (2020), Airborne Lidar and Electro-Optical Imagery Along Surface Ruptures of the 2019 Ridgecrest Earthquake Sequence, Southern California, in press, Seismological Research Letters (preprint available upon request)
[Studio](#) [Get Data](#) [Item Details](#) [Add to Collection](#)

Auckland South, New Zealand 2016
 Publication: 2020-01-13 Source: OpenTopography (U) Last Modified: 2020-06-18
 Lidar was captured for Auckland Council by AAM New Zealand between September 2016 through to June 2017. The original dataset was generated by AAM New Zealand and their subcontractors. The survey area covers the southern Auckland suburbs and regions. Data management and distribution is by Land Information New Zealand.
[Studio](#) [Get Data](#) [Item Details](#) [Add to Collection](#)

Lake Isabella Lidar Collection, CA 2015
 Publication: 2019-12-23 Source: OpenTopography (U) Last Modified: 2020-06-18
 Airborne LiDAR data was acquired by CRREL on 21 August 2015 over Lake Isabella and the Lake Isabella Dam, California as part of a coincident regional study. These data were collected using an airborne laser scanning (ALS) system comprised of a Riegl Q680i full-waveform LIDAR sensor, an Applanix POS AV INS system, and custom designed hardware and aircraft integration components. The system was installed in a Partenavia P.68, with an average collection AGL of 2,600' and airspeed of 90 knots. A total of 19-flight-lines were collected during a single flight, with 50% overlap of laser swath coverage, given a 60 deg across track field-of-view.
[Studio](#) [Get Data](#) [Item Details](#) [Add to Collection](#)

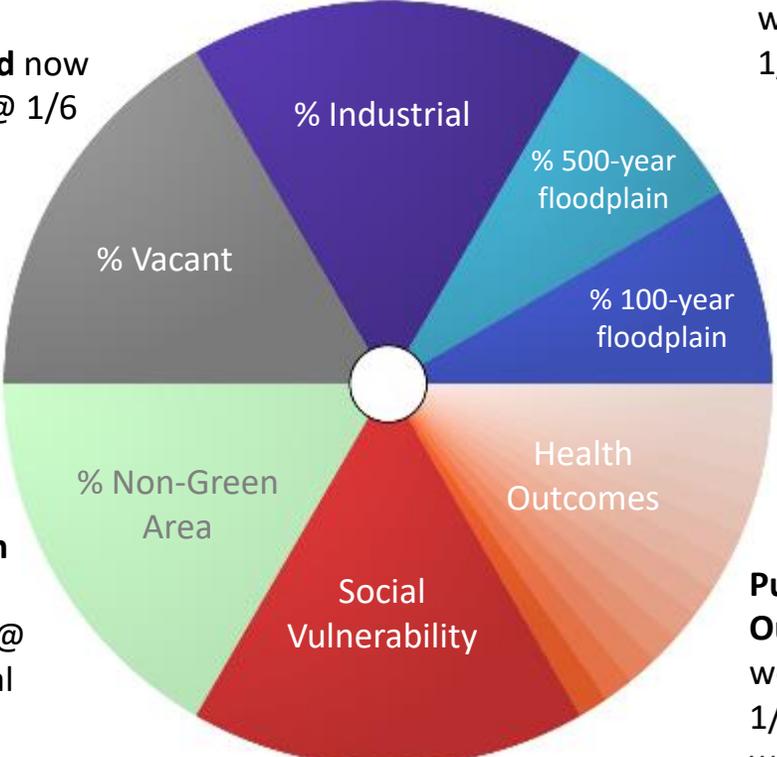
Kern Canyon Lidar Collection, CA 2008
 Publication: 2019-12-06 Source: OpenTopography (U) Last Modified: 2020-06-18
 Between September 2008 and September 2009, data were collected along the Kern Canyon fault starting south-east of Bakersfield and extending north to the southern end of Kings Canyon National Park. This

ToxPi Setup



Vacant land now weighted @ 1/6 of total

Non-green area now weighted @ 1/6 of total



Industrial land now weighted @ 1/6 of total

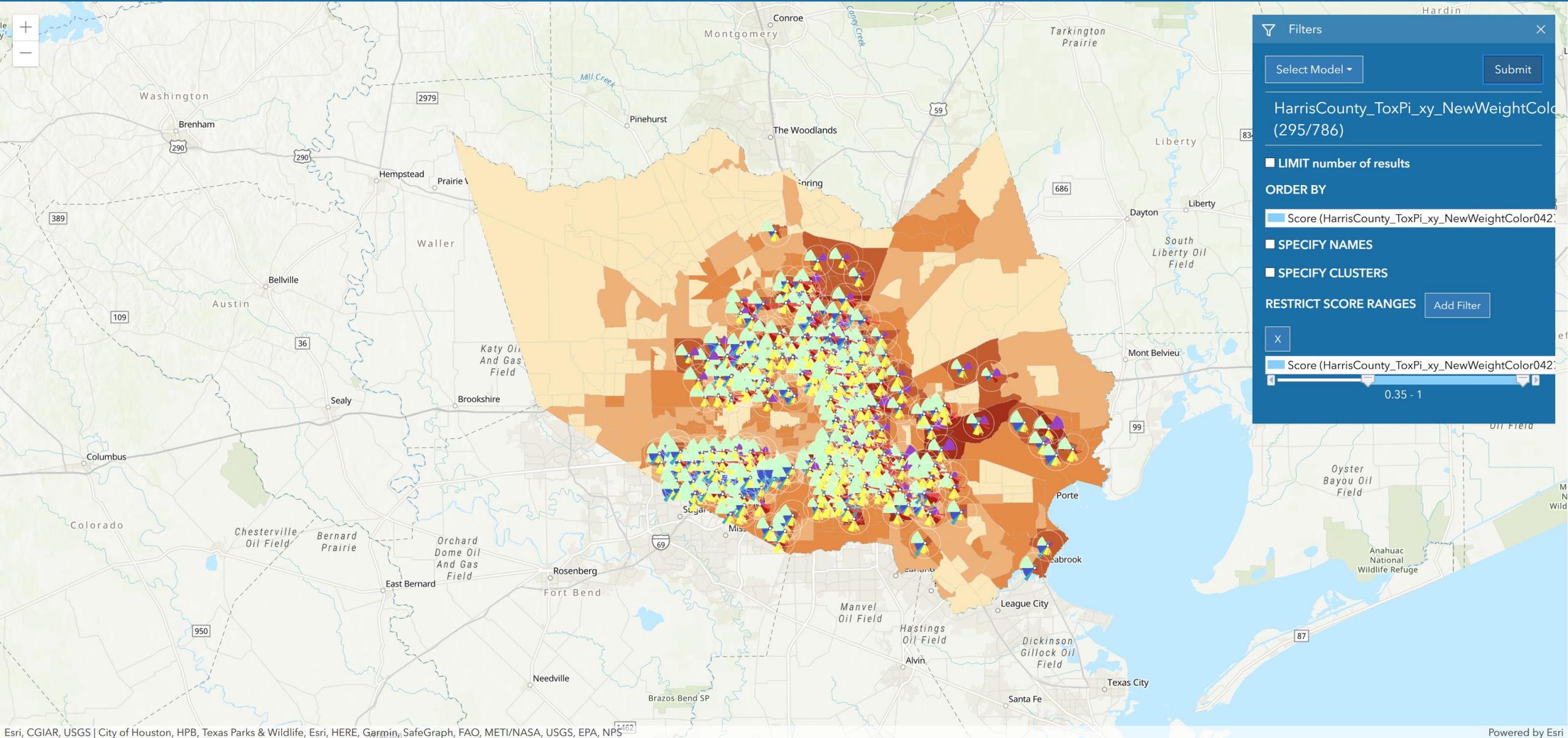
Floodplains weighted @ 1/6 of total

Social Vulnerability now weighted @ 1/6 of total or 4 categories (each @ 1/24)

Public Health Outcomes weighted @ 1/6 of total with 14 factors (each @ 1/78)

ToxPi Interface

☰ ToxPi GIS



ToxPi Interface

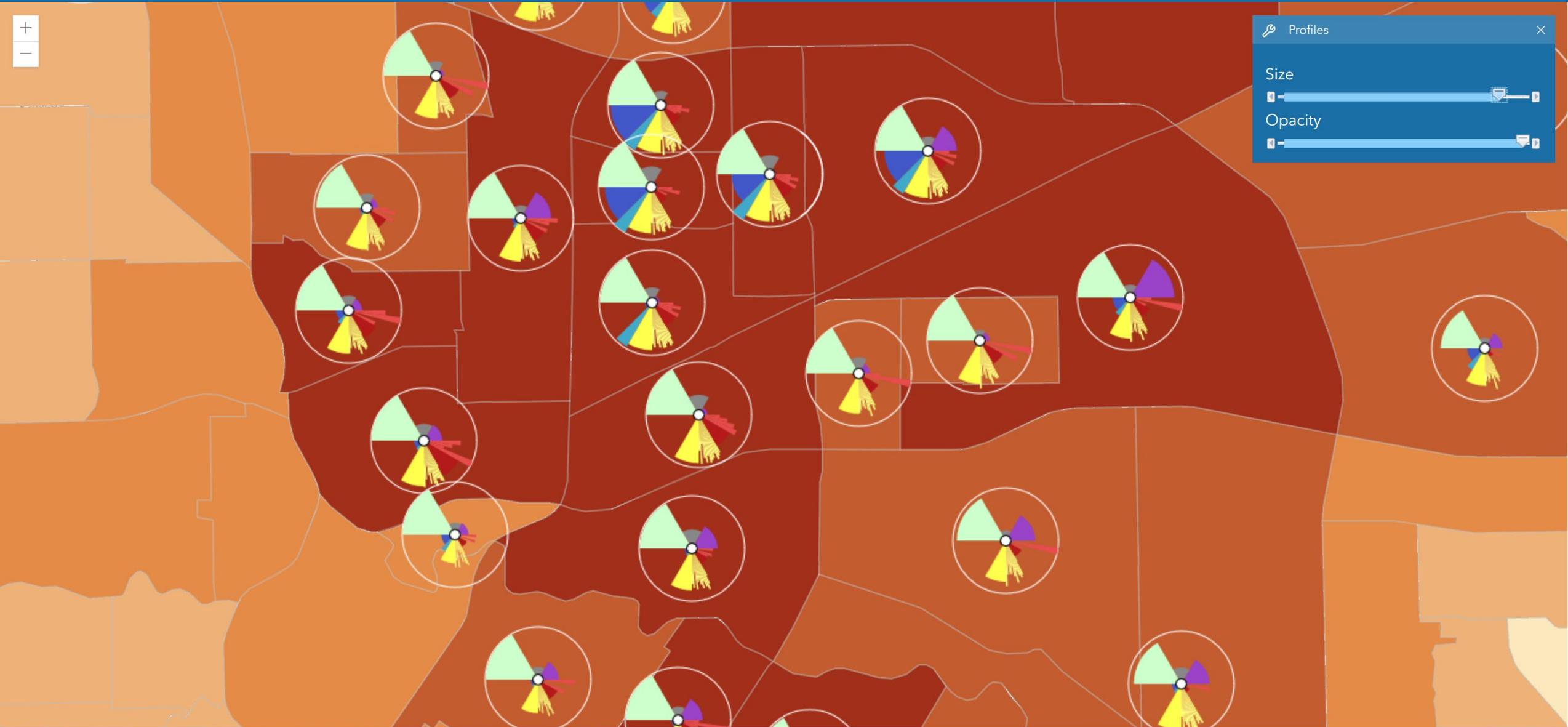
☰ ToxPi GIS



Profiles ✕

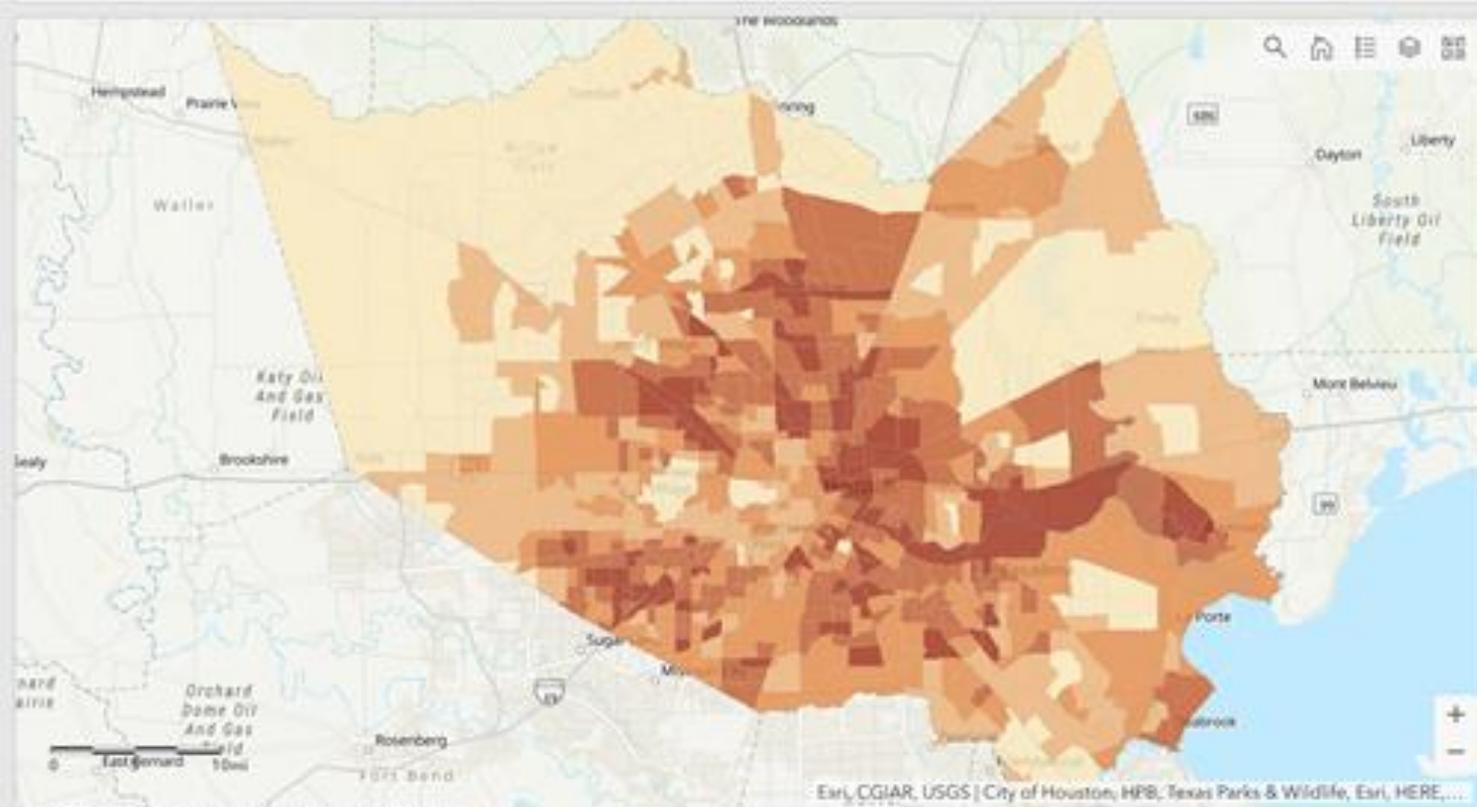
Size

Opacity

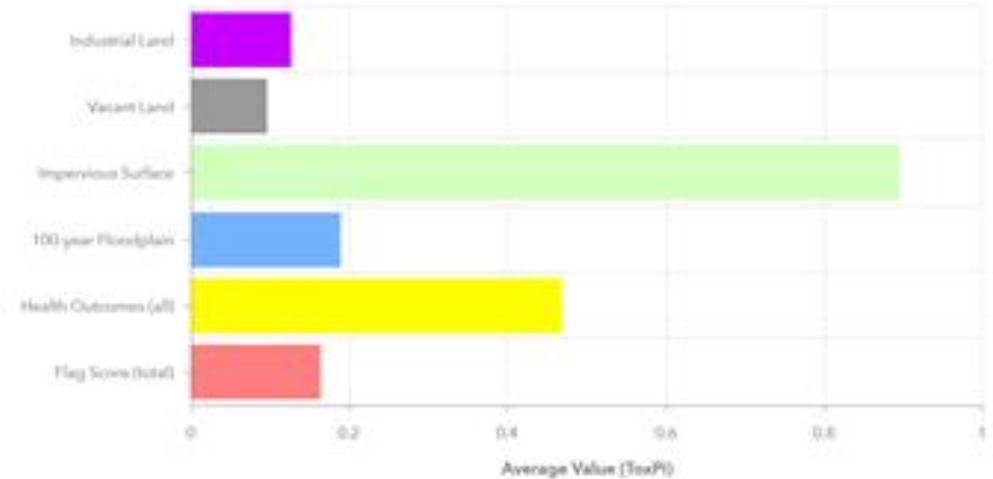


Interactive Online Dashboard

Vulnerability in Harris County (ToxPi*GIS data)

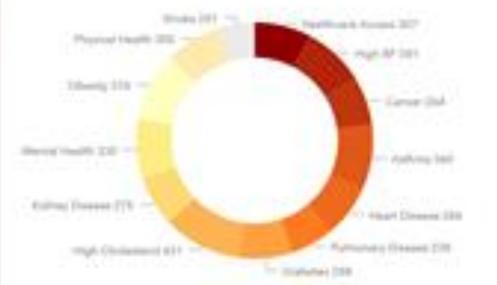


Vulnerability to Toxic Transferal, Harris County

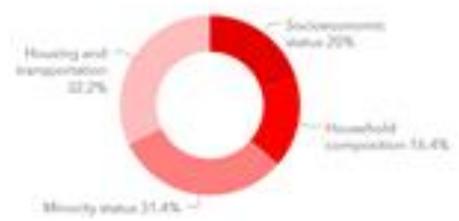


Last updated: a few seconds ago

Health Outcomes, Harris County



Social Vulnerability ('flag score' categories), Harris County



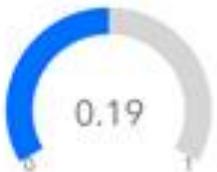
Last updated: a few seconds ago

Average Vulnerability Score

0.32

Last updated: a few seconds ago

Mean % Floodplain



Last updated: a few seconds ago

Top 10 Census Tracts by Total Vulnerability (ToxPi Score), Harris County



Last updated: a few seconds ago

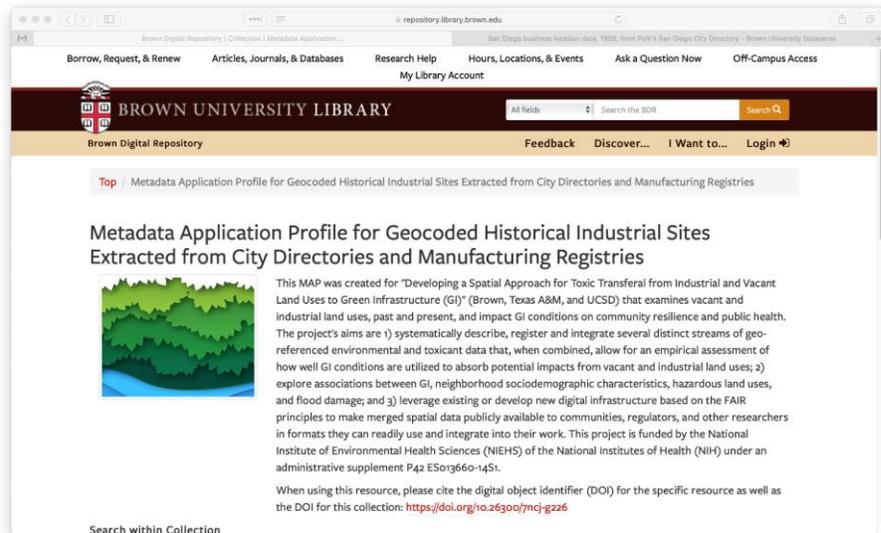
Last updated: a few seconds ago

San Diego County | **Harris County** | Rhode Island

Esri, CGIAR, USGS | City of Houston, HPB, Texas Parks & Wildlife, Esri, HERE,...

Making these datasets F.A.I.R.

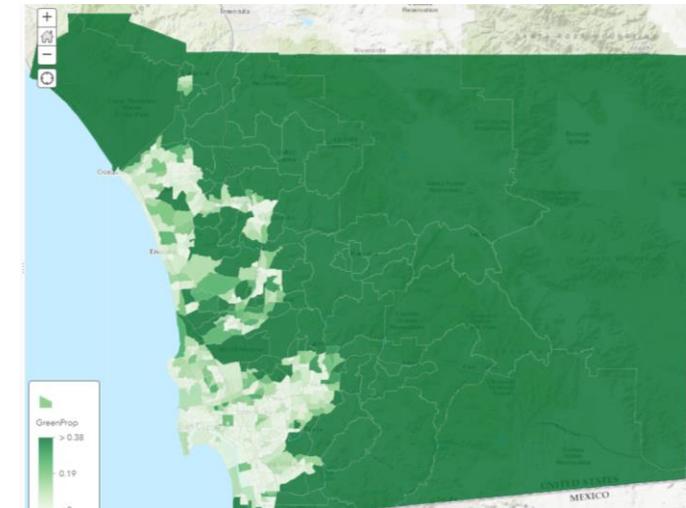
- **Registries and repositories in which the data are shared already:** Dataverse, ArcGIS Online, Brown Digital Repository, Data Discovery Studio
- **Metadata standards used:** ISO 19115/19139, schema.org
- **Data formats and services:** shapefiles, feature services, CSV
- **Vocabularies:** CINERGI Ontology (automated metadata enhancement), domain vocabularies (NAICS)
- **Identifiers:** DOIs
- **Demonstrated access, interoperability and reuse:** Jupyter notebooks, dashboards



The screenshot shows a web browser window displaying a metadata application profile. The header includes the Brown University Library logo and navigation links. The main content area is titled "Metadata Application Profile for Geocoded Historical Industrial Sites Extracted from City Directories and Manufacturing Registries". Below the title is a small map showing a green landscape with a blue river. The text describes the project's goals: to develop a spatial approach for toxic transfer from industrial and vacant land uses to green infrastructure (GI) in Brown, Texas A&M, and UCSD. It details the project's aims, including describing and registering streams, assessing environmental conditions, and exploring associations between GI and sociodemographic characteristics. The profile also includes a search bar and a "Search within Collection" link at the bottom.



The screenshot shows a Harvard Dataverse entry for "San Diego business location data, 1958, from Polk's San Diego City Directory". The page features the Harvard Dataverse logo and navigation options. The main title is "San Diego business location data, 1958, from Polk's San Diego City Directory". Below the title is a "Description" section with a document icon, providing details about the dataset's origin and content. The "Subject" and "Keyword" sections are also visible. A "Related Publication" section lists a relevant article. The page includes an "Access Dataset" button and a "3 Downloads" indicator. The bottom of the page has tabs for "Files", "Metadata", "Terms", and "Versions".



FAIR Checklist + Jupyter Notebooks

- All registered datasets can be found through ArcGIS Online based on their metadata, via online user interface or via Python code in a Jupyter Notebook
- The datasets can be loaded in notebooks for additional analysis

FAIR Checklist

Dataset	Source	Registered in	Data accessible via	Interoperable	Reusable
Green Infrastructure	TNRIS	AGOL, DDS; with respective GUIDs	Download, web services; ISO-19115 metadata accessible	Shapefiles, feature services, CSV; CINDERGI ontology, schema.org/Json-LD generated	Provenance; domain standards compliant, rich metadata
Vacant Addresses	USPS	Same as above	Same as above	Same as above	Same as above
Public Health (14 factors)	CDC	Same as above	Same as above	Same as above	Same as above
Social Vulnerability	CDC	Same as above	Same as above	Same as above	Same as above
Industrial Land Uses	local	Same as above	Same as above	Same as above	Same as above
Floodplain	FEMA	Same as above	Same as above	Same as above	Same as above
Vacant Lands	SANDAG	Same as above	Same as above	Same as above	Same as above
San Diego historical business locations	City of San Diego	AGOL, DDS, Brown, Dataverse; DOIs, GUIDs	Download; ISO 19115 metadata	Shapefiles, CSV; NAICS codes	Same as above

Jupyter Notebook

The screenshot shows a Jupyter Notebook titled "Accessing EUC Data - Demo" with a last checkpoint of 04/16/2020. The notebook interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations and execution. The code cell contains the following Python code:

```
In [57]: from IPython.display import display
for item in bioreg_result:
    display(item)
```

The output of the code cell displays four interactive map widgets. Each widget shows a map of a specific region with a legend and descriptive text. The maps are:

- Harris County - All Data (w/ Additional Flood Plain & Public Health Data)**: Shapefile containing relevant data from different factors (% vacant, % industrial, % green, SVI, flood plain, Public Health) collected to support green infrastructure analysis for SRC EUC-17. Also includes GEOID (2016 US Census tract identifier). Last Modified: March 04, 2020. 0 comments, 16 views.
- Rhode Island - All Data (w/ Additional Flood Plain & Public Health Data)**: Shapefile containing relevant data from different factors (% vacant, % industrial, % green, SVI, flood plain, Public Health) collected to support green infrastructure analysis for SRC EUC-17. Also includes GEOID (2016 US Census tract identifier). Last Modified: March 04, 2020. 0 comments, 12 views.
- San Diego - All Data (w/ Additional Flood Plain & Public Health Data)**: Shapefile containing relevant data from different factors (% vacant, % industrial, % green, SVI, flood plain, Public Health) collected to support green infrastructure analysis for SRC EUC-17. Also includes GEOID (2016 US Census tract identifier). Last Modified: March 04, 2020. 0 comments, 8 views.
- Rhode Island - National Flood Hazard Layer, Digital Flood Insurance Rate Map Data**: Digital Flood Insurance Rate Map (DFIRM) data for Rhode Island is collected here to support green infrastructure analysis in SRC EUC-17. Published with support from NIEHS SRP program. Last Modified: February 21, 2020. 0 comments, 18 views.

Vocabularies:

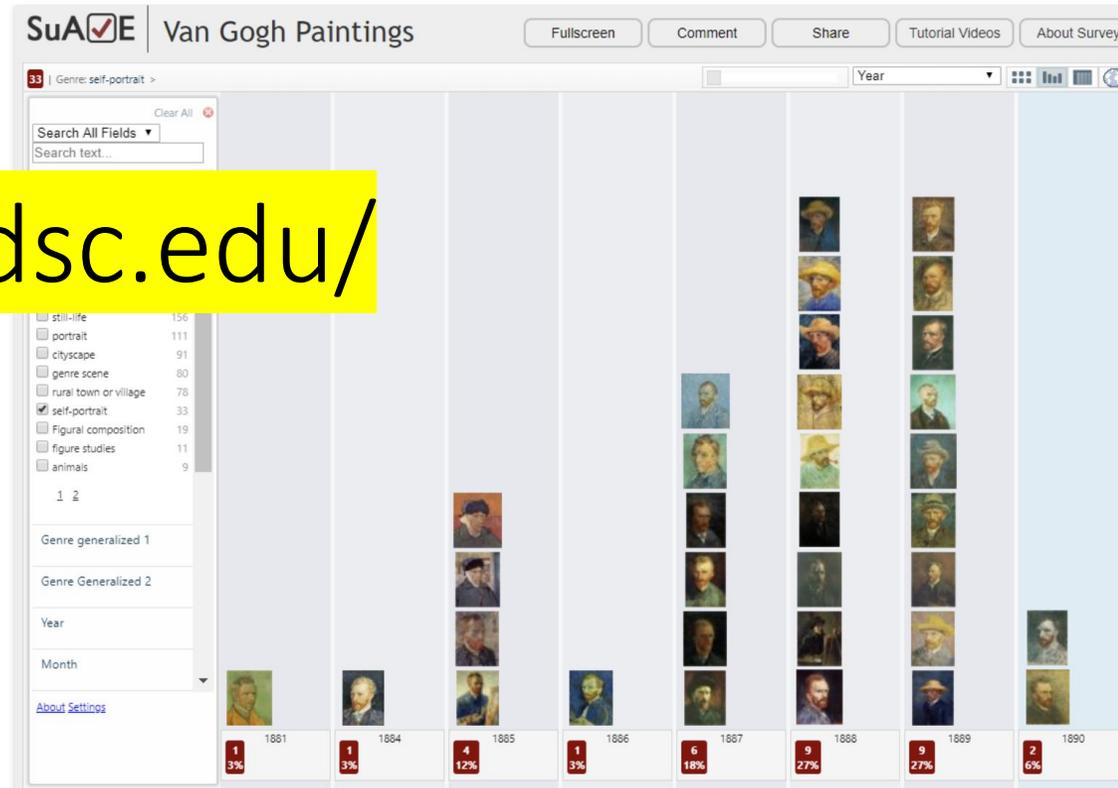
- CINERGI Ontology
 - Developed in OWL using Protégé
 - Accessible via website:
http://hydro10.sdsc.edu/cinergi_ontology/
 - Archived in Github:
<https://github.com/CINERGI/ontologies>
 - Semantic services via Scigraph:
<http://ec-scigraph.sdsc.edu:9000/scigraph/docs/>
- NAICS (North American Industry Classification System) codes

Public Repositories:

- Harvard/Brown Dataverse
 - Library-hosted repositories
- ArcGIS Online (for spatial data)
- Using data from sources hosted in
 - Socrata (CDC)
 - ArcGIS Enterprise server
 - Local Government web sites

Software/GitHub:

- Managing and sharing code for extracting data from business directories
- Sharing ontology
- Archive in Zenodo/Integrate with Binder
- ReproZip (NYU)



SuAVE: Survey Analysis via Visual Exploration.

SuAVE is a new online platform for visual exploratory analysis of surveys and image collections. It integrates visual, statistical and cartographic analyses and lets users annotate and share images and distribution patterns. It also provides a gateway into advanced data science and machine learning tools by integrating with R and Jupyter notebooks.

<http://suave.sdsc.edu/>

From FAIR Representation to Comparative Analysis

Green Land Cover, %%

Home Gallery Map Scene Notebook Groups Content Organization

Rhode Island - All Data (w/ Additional Flood Plain & Public Health) Overview Data Visualization



Shapefile containing relevant data from different factors (% vacant, % industrial, % green, SVI, flood plain, Public Health) collected to support green infrastructure analysis for SRC EUC-17. Also includes GEOID (2016 US Census tract identifier).

Open in Map Viewer
Open in Scene Viewer
Open in ArcGIS Desktop
Share
Metadata

Feature Layer by izaslavsky_ucsd
Created: Mar 4, 2020 Updated: Apr 13, 2020 View Count: 67

Home Gallery Map Scene Notebook Groups Content Organization

Harris County - EUC All Data (w/ Additional Flood Plain & Public H Overview Data Visualization



Shapefile containing relevant data from different factors (% vacant, % industrial, % green, SVI, flood plain, Public Health) collected to support green infrastructure analysis for SRC EUC-17. Also includes GEOID (2016 US Census tract identifier).

Open in Map Viewer
Open in Scene Viewer
Open in ArcGIS Desktop
Share
Metadata

Feature Layer by izaslavsky_ucsd
Created: Feb 28, 2020 Updated: Sep 21, 2020 View Count: 214

Home Gallery Map Scene Notebook Groups Content Organization

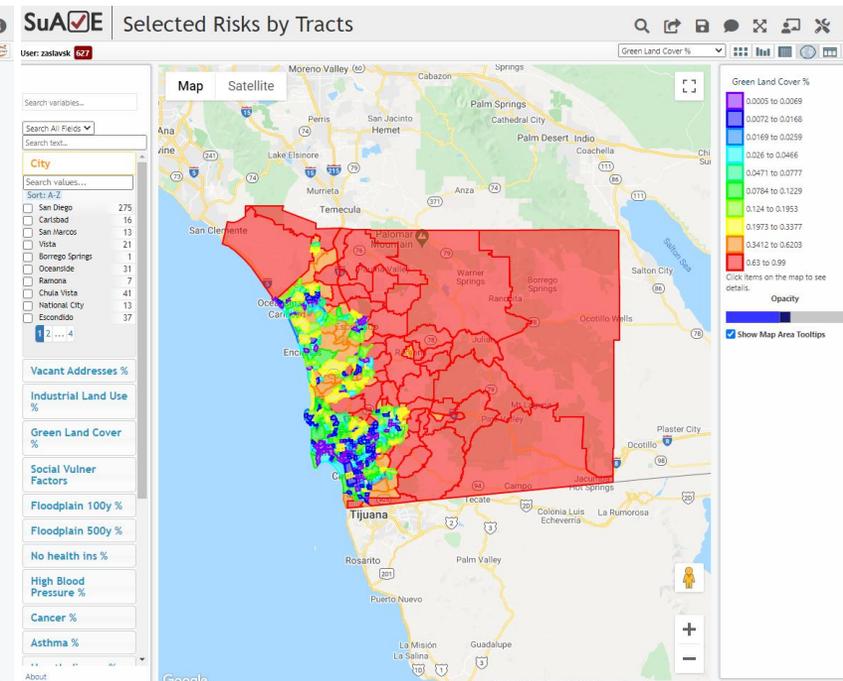
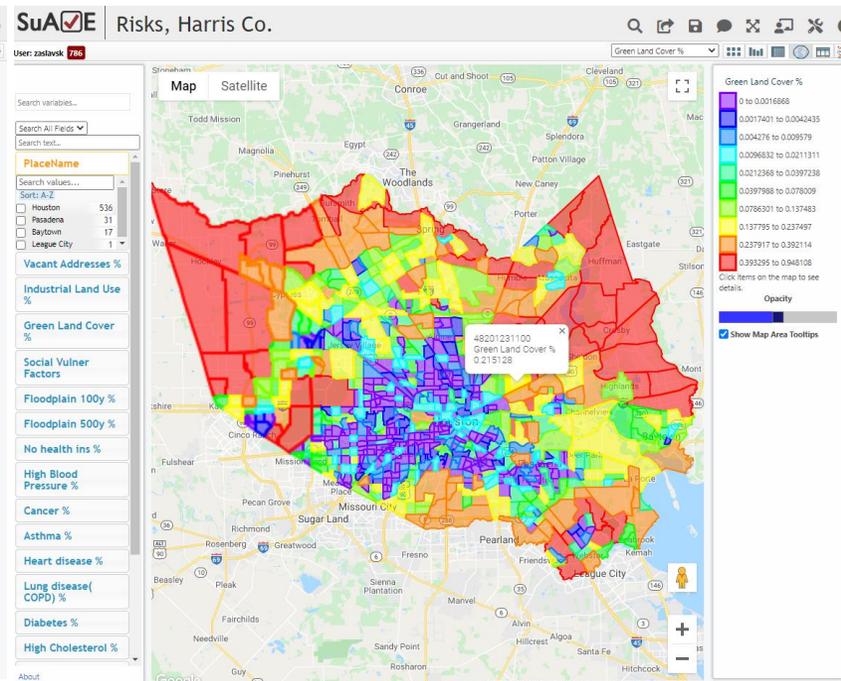
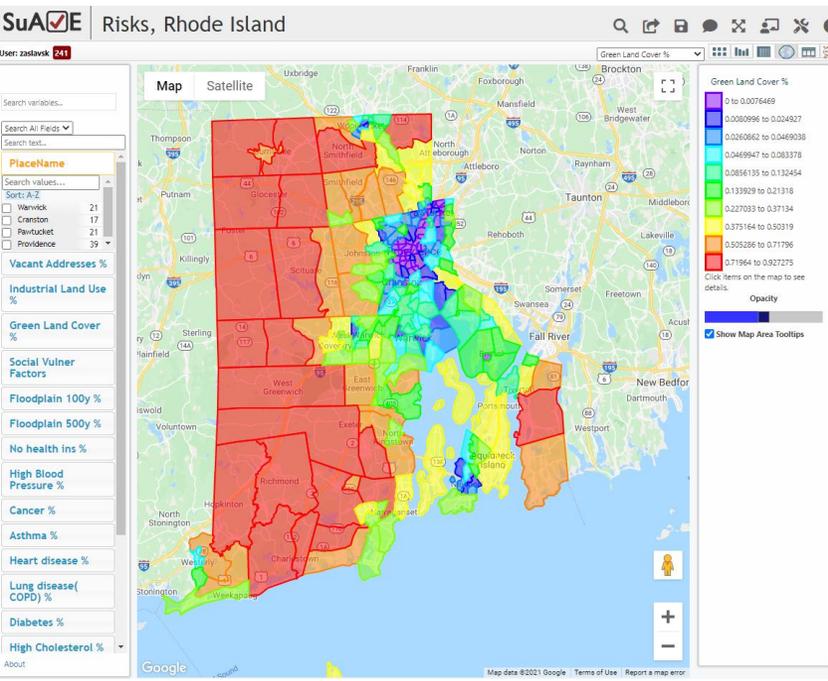
San Diego - All Data (w/ Additional Flood Plain & Public Health Da Overview Data Visualization



Shapefile containing relevant data from different factors (% vacant, % industrial, % green, SVI, flood plain, Public Health) collected to support green infrastructure analysis for SRC EUC-17. Also includes GEOID (2016 US Census tract identifier).

Open in Map Viewer
Open in Scene Viewer
Open in ArcGIS Desktop
Share
Metadata

Feature Layer by izaslavsky_ucsd
Created: Mar 4, 2020 Updated: Apr 13, 2020 View Count: 95



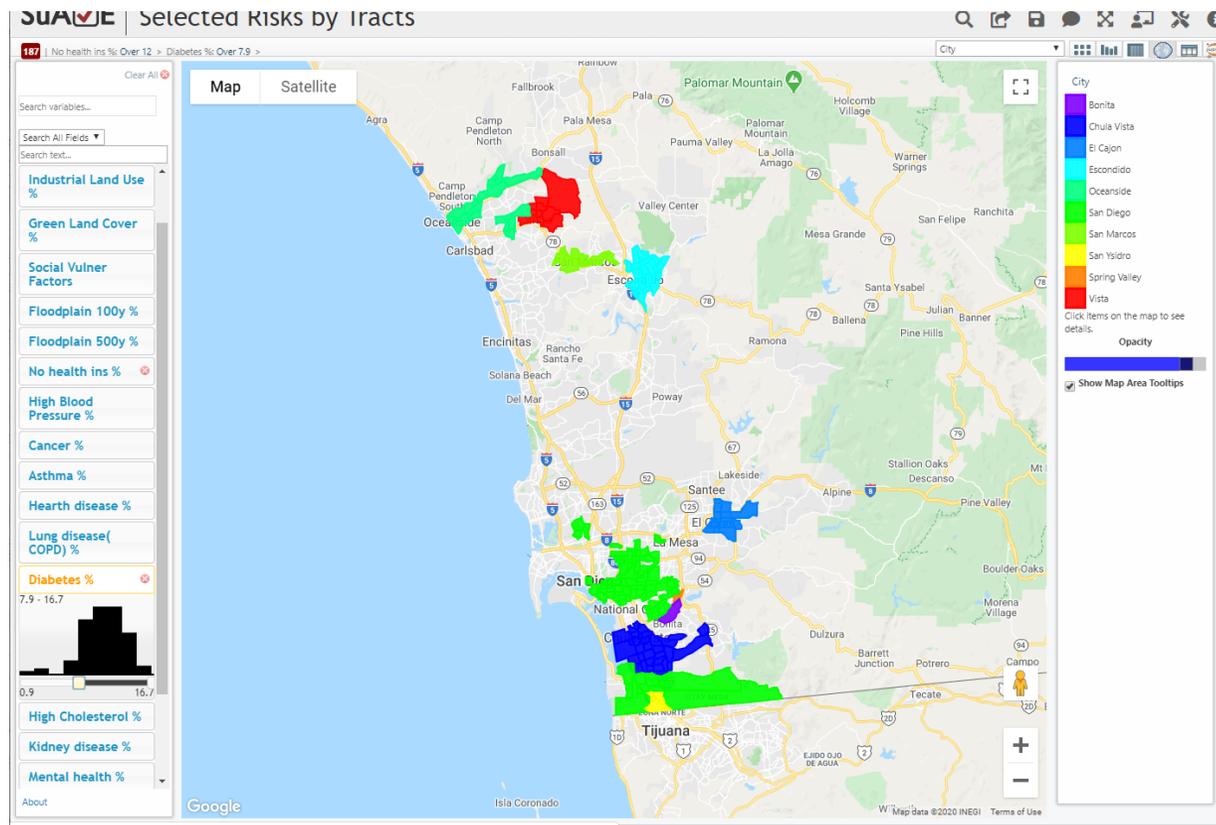
http://suave2.sdsc.edu/main/file=zaslavsk_Risks_Rhode_Island.csv&view=map

http://suave2.sdsc.edu/main/file=zaslavsk_Risks_Harris_Co_.csv&view=map

http://suave2.sdsc.edu/main/file=sdhhsa_Selected_Risks_by_Tracts.csv&view=map

Communication Strategy

- Health outcomes and risk factors by census tracts used by San Diego County HHSA to examine spatial distribution of populations at risk by multiple indicators and develop strategies for COVID-19 pandemic response



Accessible from <http://suave2.sdsc.edu/gallery/sdhhsa>

Home Gallery Map Scene Groups Content Organization

San Diego - All Data (w/ Additional Flood Plain & Public Health Data)

Overview Data Visualization

Shapefile containing relevant data from different factors (% vacant, % industrial, % green, SVI, flood plain, Public Health) collected to support green infrastructure analysis for SRC EUC-17. Also includes GEOID (2016 US Census tract identifier).

Feature Layer by rlagoc_UCSDOnline5

Created: Mar 4, 2020 Updated: Apr 13, 2020 View Count: 34

Description

EUC All Data by Census Tract (w/ Additional Flood Plain & Public Health Data)

Location: San Diego

Description:
Shapefile containing data from the four factors (% vacant, % industrial, % green, & SVI) relevant for green infrastructure analysis for SRC EUC-17, with additional Flood Plain and Public Health data included. Data is organized by US Census tract, and includes GEOID (2016 US Census tract identifier). File combines and simplifies multiple layers, described in detail below:

- Site Boundary** - Simple shapefile delineating the border of the three study areas.
- Census Tracts** - Shapefile delineating 2016 US Census tracts and related data.
- Vacant Addresses** - Percent (proportion) of vacant properties, by 2016 US Census tract (10 categories, brown-to-tan color scheme).
- Industrial Land Use** - Percent (proportion) of industrial land use, by 2016 US Census tract (purple for industrial parcels, lavender for non-industrial parcels).
- Green Infrastructure** - Percent (proportion) of green land use / land cover, by 2016 US Census tract (green for green infrastructure, lavender for non-green infrastructure).
- Social Vulnerability** - Total CDC social vulnerability index (SVI) "Flags" -out of a possible 15- for 2016 US Census tract. Each flag denotes that the census tract was among the most vulnerable in the county for a given social factor (e.g. poverty, lack of vehicle access). See: <https://svi.cdc.gov/>
- Flood Plain** - Shapefiles showing the percent (proportion) of floodplain area in each US Census tract; one for 100-year (1% annual chance) floodplain, one for 500-year (0.2% annual chance) floodplain
- Public Health** - Shapefiles showing the prevalence of a given health outcome in a given US Census tract (See below for detailed breakdown of different outcomes included).

Details

Source: Feature Service
Created from: SanDiegoAllDataInclFloodHealth, Shapefile
Data Last Updated: Mar 4, 2020, 11:32:00 AM
Size: 5 MB
★★★★★

Share

Owner: rlagoc_UCSDOnline5

Tags

Data are described in ArcGIS Online and available through ArcGIS Enterprise Server

Lessons Learned & Recommendations

- It is important to quickly demonstrate benefits of FAIR data sharing, by developing online applications and services that show how the data can be accessed and used together
- For our data, following established data formats and using common repository platforms was the key. Due to diversity of data variables and hit-and-miss automatic semantic tagging, iterative improvement is essential.
- Sharing the data encouraged its use beyond the applications initially intended (e.g., in the context of COVID-19 analysis)
- Data interoperability and reuse issue; informing data management planning
- Following FAIR principles as in <https://www.force11.org/group/fairgroup/fairprinciples> or <https://www.openaire.eu/how-to-make-your-data-fair> is helpful

Outputs and Outcomes: Products

Short-Term:

- Shared, online, searchable data downloads
- Interactive online dashboard
- “Green Infrastructure and Health Equity Improvement Zones” in underserved communities of San Diego

Medium Term:

- *Presentations:*
 - Newman GD, Malecha ML, Karaye IM, Frickel S, Marlowe TW, Zaslavsky I, Pezzoli K. Comparing Regional Drivers of Toxics Transferal Risk: Applying the Toxics Mobility Vulnerability Index in San Diego County, CA; Harris County, TX; and the State of Rhode Island. NIEHS Superfund Research Center Annual Meeting December 2020 – *Winner of Superfund Meeting Poster Award*
- *Publications:*
 - Malecha ML, Kirsch KR, Karaye IM, Horney JA, Newman G. (2020) Advancing the Toxic Mobility Inventory: Development of a Toxics Mobility Vulnerability Index and application to Harris County, TX. *Sustainability: The Journal of Record*. 13(6): 282-291. DOI: 10.1089/sus.2020.0067.
 - Newman, G., & Malecha, M., “Integrating ToxPi Outputs with Arc GIS Dashboards to Identify Neighborhood Threat Levels of Contaminant Transferal During Flood Events.” In press at *Journal of Spatial Sciences*

Long Term:

- *Future Funding:*
 - Future funding for green infrastructure from Biden Administration Options
 - Competitive supplemental funding for future activation of these systems (NSF Smart and Connected Communities)
- *Publications in Preparation:*
 - Comparative Regional Drivers of Toxics Transferal Risk: Applying a Modified/Updated *Toxics Mobility Vulnerability Index* in San Diego County, CA; Harris County, TX; and the State of Rhode Island, projected submission to *Journal of Environmental Informatics*
 - Contaminants Linked to Longitudinal Industrial Land Uses across Multiple High-Risk Flood Zones, projected submission to *Journal of Urban History*

Galen Newman, PhD

PI, Community Engagement Core
Associate Professor, Landscape
Architecture and Urban Planning
Texas A&M University
gnewman@arch.tamu.edu

Jennifer A. Horney

Co-PI, Community Engagement
Core
Professor, Epidemiology
University of Delaware
horney@udel.edu

Scott Frickel, PhD

PI, Community Engagement
Core
Professor, Sociology
Brown University
scott_frickel@brown.edu

Andrew Creamer

Science Data Specialist
Brown University
andrew_creamer@brown.edu

Keith Pezzoli, PhD

PI, Community Engagement
and Research Translation
Professor, Urban Studies and
Planning
UC San Diego
kpezzoli@sdsc.edu

Ilya Zaslavsky, PhD

Director, Spatial Information
Systems Laboratory, SDSC
UC San Diego
zaslavsk@sdsc.edu

