Plan Overview

A Data Management Plan created using DMPTool-Stage

Title: SCC-CIVIC-PG Track B Journal of Civic Technology

Creator: Lane Rasberry - **ORCID:** 0000-0002-9485-6146

Affiliation: University of Virginia (UVA) (virginia.edu)

Principal Investigator: Jonathan Kropko

Data Manager: Lane Rasberry, Dave Ghamandi

Project Administrator: Dave Ghamandi

Template: NSF-CISE: Computer and Information Science and Engineering

Project abstract:

There exist multiple networks of civic technology (civic tech) community organizations in the United States which organize local people to identify barriers to accessing government or public resources and to increase accessibility with technical solutions. As an example, consider Code for America, which is a national organization supporting 80 regional chapters or "brigades" each convening frequent meetings for technically skilled volunteers, nonprofit or government representatives, and community stakeholders to discuss problems and design solutions to them. Strengths of groups like these are that they have community participation; they are able to reach community consensus on setting priorities; they propagate ethics of diversity, equity, and inclusion throughout the network; and their methods are proven to create usable technology. Shortcomings of civic tech communities are that they lack access to conventional resources and find it difficult to establish institutional partnerships. In comparison, universities also seek to serve communities with technology, and they are comparatively well resourced and have eager students, faculty, and even funders who would contribute to such projects. Shortcomings of universities are that they struggle to establish and maintain meaningful and independent community partnerships. If it were possible to match the stakeholder participation of civic tech

communities with the development capacity of universities, then these community and academic sectors would more effectively achieve their objectives, the technological solutions would be used immediately by stakeholders who requested them, and the research documentation practices of universities would make these solutions reusable for other communities which wished to adopt them.

Start date: 09-30-2022

End date: 03-30-2023

Last modified: 08-07-2023

Copyright information:

SCC-CIVIC-PG Track B Journal of Civic Technology

This project produces an interdisciplinary, multi-media rich, community-inclusive academic journal. Consequently, this project will accept as many popular filetypes as possible. This project has expectations of processing 4 types of media: text, datasets, code, and "files", which include videos, images, slides, or even undefined filetypes. This project does not include non-digital media.

Data produced by this project is anticipated to be "small", meaning that the team anticipates being able to archive all data in the home university's institutional repository, that it can be mirrored in Zenodo, and that reusers who are familiar with media copying should have little difficulty re-posting media from this project into other repositories as desired.

More details are as follows:

Text documents will be available as plain text (.txt), portable text (.PDF), and LaTex (.tex). Preferred dataset format is .csv. Slides will be OpenDocument Presentation (.ppt) and as .pdf. Code may be in various programming languages, shared in github and exportable as text. Video will be as .mp4 and .ogv.

Datasets presented by this project are estimated to be less than 10MB each. Text presented by this project is estimated to be less than 10MB. Images will be as needed for presentation, and estimated to be under 100MB. Video is estimated to be 2-4 hours of content and 1TB.

Digital photographs of public will be retained as JPEG (.jpg) if originally created in that format. Otherwise, files will be converted and retained in TIFF version 6 uncompressed (.tif) formats. Geospatial metadata, if applicable, will be retained with geo-referenced TIFF (.tif) files.

During the data collection process, some media including texts, datasets, code, and messages will be disseminated to collaborators in order to facilitate open dialogue about methodologies and tentative conclusions. By the end of the project the products will be archived in appropriate ways with open and FAIR reuse by others as a goal.

The institutional repository is a research university which makes recommendations for metadata. These standards are published at https://data.library.virginia.edu/ . Additionally, outcome metadata will be registered in Wikidata, which publishes metadata standards including field-specific recommendations dependent on subject matter. Although not all-inclusive of all project activities, an example outline of Wikidata metadata standards is hosted in WikiProject Informatics at https://www.wikidata.org/wiki/Wikidata:WikiProject_Informatics including guides for software, hardware, algorithms, and protocols.

FAIR data and open content principles are the default philosophies for this project's access and sharing. This project will not handle sensitive data or media, so privacy is less of a concern, but the project team includes a diversity coordinator and the team will routinely check its practices to address any privacy or diversity issues which arise. All though products of this project will be FAIR and open, this project may use external tools, datasets, or resources which have usage restrictions and which we cannot copy or redistribute. When this is necessary, this project will catalog the metadata of the resource so that the research process is reproducible and others can attempt to access the resource in the same way that this project did.

The intent of this project is to comply with high standards of FAIR and open practices common in activist communities advocating for the Open Movement. In general, this means CC0 (Creative Commons Zero) data licenses, CC-By 4.0 (Creative Commons attribution) media licenses, and software licenses comparable to the GNU

General Public License. Products from this project will be open and ideally FAIR, particularly metadata.

This project includes community partnerships and has a goal of including underrepresented demographics in this field, including women, African diaspora, and Indigenous peoples. When this project makes records related to demographics it will do so in consultation with a diversity coordinator. Although created in advocacy for Indigenous peoples, the 2021 The CARE Principles for Indigenous Data Governance are an inspiration for general community data management practices and the project team will refer contributors to this publication.

FAIR data and open content principles are the default philosophies for this project's access and sharing. Most non-data project outcomes are available for re-use, re-distribution, and remixing with attribution according to the Creative Commons Attribution 4.0 International license. (CC-By 4.0 International). Data project outcomes will be available with a Creative Commons Zero (CC0) license, so not even attribution is required, but we will provide attribution instructions for those who wish to provide attribution for reasons including pointing to our validated original dataset. Tools such as code will be available in GitHub but also archived in our institutional repository, and with a license comparable in free software philosophy to the GNU General Public License.

This project has three routine archiving practices in place: store outcomes in our university's institutional repository; also put them into Zenodo as a public archive; and finally put them into a consumer-facing location where the public is likely to find them. The university repository should be as stable as the university library itself; Zenodo seems as stable as CERN; and the consumer-facing location is contemporary most accessible and findable place, but typically controlled by a corporation or nonprofit organization and web repositories tend to be ephemeral.

Tying all these together, resource metadata will point to the various locations where anyone can find project outcomes. Also the metadata will assist anyone with citing the resource if they need to document its provenance. Metadata will primarily refer to the university institutional repository, as this is the source most under the control of the project team for the longest time.

While the university institutional repository and Zenodo can accept almost any filetype, consumer-facing repositories have more variance. GitHub is one such repository and familiar for storing code. In contrast, as this project is multidisciplinary and will collect datasets from fields which we cannot currently anticipate, we may deposit some datasets in specialized repositories which we find through community or subject matter expert referral, and which we would only need if and when community contributors shared such content. In the best case, we proactively share data in our standard archives and the fitting data repository; in the worst case, we only share data in the standard archives, so it is still accessible even if it is less findable through currently popular search aggregation services.

Much of the project metadata will be mirrored in Wikidata, which is a popular information referral service mirrored by commercial big tech companies including Facebook, Apple, Amazon, and Google. Those companies' use of the data is unpredictable, but historically, they copy and redistribute whatever they find in Wikidata.

This project has a home institutional repository in the University of Virginia Library. The repository name is Libra, and it contains subdivisions such as Libra Data. The library archiving practice can be simplified into two parts:

depositing of the files, and creation of file metadata. Both files and metadata stay in the library indefinitely. Because of licensing, anyone may copy and redistribute project files or metdata. Metadata, being FAIR and easier to mirror, will circulate to other search aggregators much more than the files themselves. Metadata includes persistent identifiers and an exchange of equivalent identifiers in other services, like for example, in the way that Wikidata or services such as SciCrunch assign their own identifiers to resources while also copying the identifiers of other services.

In the unlikely event that the university repository itself fails, and for the purpose of making content more accessible in other repositories with their own promotion practices, files from this project will also be mirrored in Zenodo. This is possible because most project outcomes are small in file size, and of the sort of content that Zenodo can mirror.

The general plan for distribution and dissemination is to register project outputs into syndicated library catalogs so that the metadata of products is open and FAIR and will flow through search aggregators. Metadata will point to archives where the resource is accessible. One of those archives will be the university's own institutional repository; other archives may be used as appropriate for changing circumstances, like for example, a collaborator may also have their own repository where outcomes are also archived.

This project's organizing team solicits crowdsourced creation of case studies in the field of civic technology. Roles in this process are comparable to those of publishing in any conventional academic journal, but for clarity, a description follows: a typical case study includes a report, dataset, code, and illustrative media such as images, slides, or videos. The roles for these activities are diversity coordinator, community contributor, research coordinator, principal investigator, and librarian. The workflow is as follows: The "diversity coordinator" makes public solicitations for project participation and submissions. In response, the "community contributor" submits a case study. This person is a volunteer from outside the project, and they have no program obligation and only join as a service user and supporter. When they make their case study contribution, the "research coordinator" on the project team checks their submission for license compatibility, permission to archive, and compliance with the submission process. If the submission passes, then the "principal investigator" (or delegate) double-checks the submission for suitability. If the submission again passes, then the "librarian" arranges for the submission's publication and archiving. All of these roles are described in the project description, but note that only project staff have responsibilities, while volunteers or community contributors are invited to submit but always checked by staff.

Other research products include the dataset surveying the environment of existing projects in the field of civic technology. These will be collected by the research coordinator, confirmed by the PI, and archived by the librarian.

Supplementary research projects may include slides, notes, blog posts, video recorded presentations, or social media infographics. The project team is in agreement that these are worth sharing, and each team member is responsible for archiving the materials they produce in the university's institutional repository in the folder for this project.

Planned Research Outputs

Dataset - "dataset of civic technology projects"

The dataset of civic technology projects will be a combination of existing datasets describing civic technology projects, our own addition of new such project listings, and the data wrangling we do to make this content uniform. Although we will archive and distribute the dataset through our university's institutional repository and Zenodo, our intended distribution channel is to contribute enriched data back to the original source projects.

Workflow - "civic technology project classification scheme"

There are various published processes for classifying civic technology projects. Using protocols.io's methodology documentation tool, we will publish the classification schemes we applied as models for others to discuss, reuse, or adapt.

Service - "sample database queries"

We will prepare some SPARQL queries for accessing model subsets of data from our database. The queries we present will showcase common use cases for our data, assist new users in using SPARQL query language to access content, and inspire users to adapt the queries for their own purposes.

Text - "Journal of Civic Technology articles"

Project descriptions which people write for review will be published in the Journal for Civic Technology which we establish. We will use the Aperio journal publishing service to perform standard journal archiving into other repositories.

Image - "illustrations"

Illustrations which we create to explain the project, including figures in publications, posters, conference slides, or similar will be deposited into Wikimedia Commons and listed in a project collection in that platform. Although most users will access these images through Wikimedia Commons, they are also indexed and findable through the more FAIR DBpedia.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
dataset of civic technology projects	Dataset	2023-02-28	Open	Zenodo	10 MB	Creative Commons Zero v1.0 Universal	None specified	No	No
civic technology project classification scheme	Workflow	2023-02-28	Open	protocols.io		Creative Commons Attribution 4.0 International	None specified	No	No
sample database queries	Service	2023-02-28	Open	GitHub		Creative Commons Zero v1.0 Universal	None specified	No	No
Journal of Civic Technology articles	Text	2023-02-28	Open	None specified	50 MB	Creative Commons Attribution 4.0 International	None specified	No	No
illustrations	Image	2023-02-28	Open	DBpedia	50 MB	Creative Commons Attribution 4.0 International	None specified	No	No