



# Working with Foundational Open Science Skills (FOSS) for Dissertation Data and Reproducibility

*How to plan a dissertation using data management  
and FAIR principles to store reproducible data*



Laura W. Dozal, iSchool, PhD Candidate

The hope is to guide interested researchers through examples on how I learned and implemented Open Science into my dissertation research

## Roots for Resilience (R4R)

### Semester with R4R

*The Roots for Resilience Program provides training and support to select graduate students on open, reproducible science and computational infrastructure to enhance research focused on resiliency in the environment.*

<https://datascience.arizona.edu/r4r>

“The Roots for Resilience Program provides training and support to select graduate students on open, reproducible science and computational infrastructure to enhance research focused on resiliency in the environment and society.” I had the opportunity to be a fellow and work alongside other graduate students learning and implementing open science frameworks into their own research projects.

We met twice a week to 1) learn about open, reproducible research practices and 2) discuss our own research and got a better understanding of how to implement FOSS into our work.

A how to get involved...

# What's in Store for Today!

- Foundational Open Science Skills
- Applications for Dissertation

Presentation Agenda This presentation will be in two parts

- First will go over what I learned during FOSS
- Second will go over how I applied it and how you might use it in your own work

# Foundational Open Science Skills (FOSS)

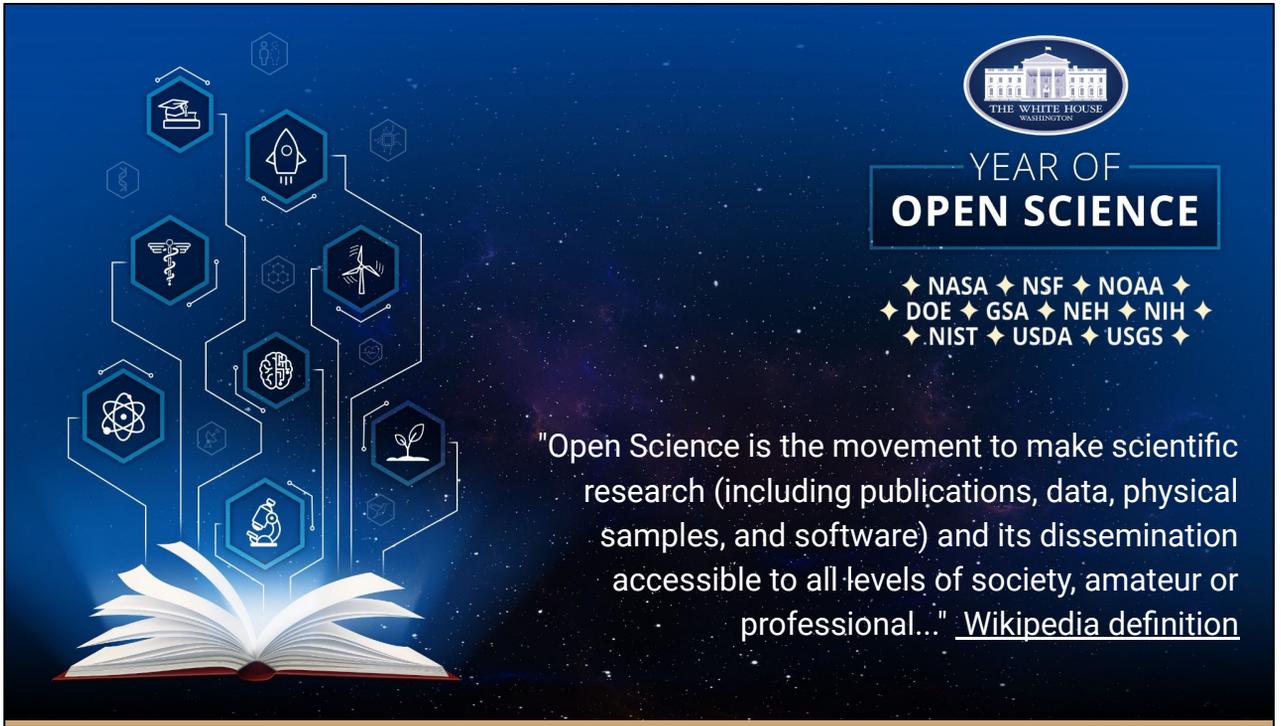
What is open science

FAIR & CARE principles

Project Management, Version  
Control and reproducibility

Cyverse and docker/containers

---



There are many considerations of what open science is, 2023 has even been called the year of open science and so everyone jumped on attempting to be the expert. Here is a broad definition of open science and how I intend to implement it in my dissertation work.

# Six Pillars of Open Science

Open Access Publications	Open Data	Open Educational Resources	Open Methodology	Open Peer Review	Open Source Software
"Open access is a publishing model for scholarly communication that makes research information available to readers at no cost, as opposed to the traditional subscription model in which readers have access to scholarly information by paying a subscription (usually via libraries)." -- <a href="http://OpenAccess.nl">OpenAccess.nl</a>	"Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike." - <a href="#">Open Data Handbook</a>	"Open Educational Resources (OER) are learning, teaching and research materials in any format and medium that reside in the public domain or are under copyright that have been released under an open license, that permit no-cost access, re-use, re-purpose, adaptation and redistribution by others." - UNESCO	"An open methodology is simply one which has been described in sufficient detail to allow other researchers to repeat the work and apply it elsewhere." - <a href="#">Watson (2015)</a>	A manuscript review process that includes some combination of Open Identities, Open Reports, Open Participation, and even Open Interaction	"Open source software is code that is designed to be publicly accessible – anyone can see, modify, and distribute the code as they see fit. Open source software is developed in a decentralized and collaborative way, relying on peer review and community production." - RedHat

## Open Access to Publications

- Make research publications free and accessible

## Open data

- Make data accessible and reusable

## Open Educational Resources

- Make learning, teaching and research materials accessible

## Open methodology

- Make your methods well documented and accessible for other to use and perfect

## Open Peer Review

- Make the article publication review process open

## Open Source Software

- Make software source code accessible and modifiable

# Ethics of Open Science



If you hope to make your data and processes actually usable, it's good to have some guidelines in place. The FAIR principles can be those guidelines, but it is good to note that they can be changed and enhanced depending on the project and your data. When collecting, saving, and managing your data it's important to follow some type of ethics, especially if you're working with sensitive information. Other principles like CARE and even some design justice and data feminism principles are important to look into.

- <https://mitpressonpubpub.mitpress.mit.edu/data-feminism>
- <https://designjustice.mitpress.mit.edu/>

<https://www.go-fair.org/fair-principles/>

CARE Principles ([Global indigenous data alliance](#))

Collective Benefit

- C1. For inclusive development and innovation
- C2. For improved governance and citizen engagement
- C3. For equitable outcomes

Authority to Control

- A1. Recognizing rights and interests
- A2. Data for governance
- A3. Governance of data

Responsibility

- R1. For positive relationships

- R2. For expanding capability and capacity
- R3. For Indigenous languages and worldviews

#### Ethics

- E1. For minimizing harm and maximizing benefit
- E2. For justice
- E3. For future use

# Data Management

Types of Data

Collection

Documentation

Storage

Analysis

Preservation

Ethics

## Data Management Plan [↗](#)

This document takes into consideration data collected and made available for researchers and practitioners.

### 1. Data Description [↗](#)

The data is collected by web Application Program Interface (API) scrapers that target social media images of the anti-gender-violence movement in Mexico using relevant keywords and time frames. A customized content-scraping script was developed using Python libraries and the InstaDownloader package which enables image and metadata collection of the Instagram social media platform. The scraper opens the Instagram application under the project's account @LWDDorientation and searches for a particular hashtag and Instagram account. The hashtags and accounts include #mexicofeminicida, #mexicofeminicida, #mujeramenos, #vivansosqueros, #noestamosotodas. The accounts include @noestamosotodas, @silvapazjusticia, @womansonfire, @borujamixta.

The current data consists of international posts mainly from Mexico, but also includes other countries in North and South America like Chile, Argentina, Brazil and the United States. The languages are Spanish, English and Portuguese. Although the topic particularly targets Mexico, other countries were found in the top search results.

### 2. Data Collection [↗](#)

Metadata collected from the hashtags and accounts include comments, likes, hashtags, URLs, account information, and automated descriptions from Instagram's API. So far, about 2,000 images have been collected within a year's timeframe between November 25, 2021, and November 25, 2023. The date November 25th was chosen because it is the international anti-violence against women day, Día Internacional contra la violencia hacia las mujeres. A second round of data collection will be done to include posts surrounding the topic on international women's day. The goal is to download over 5,000 images and their metadata.

### 3. Data Storage and Backup [↗](#)

The research findings and public facing data storage facilities will be de-identified for the privacy and safety of the participants. De-identification includes the removal of names and demographic information. Only organization names or description of organizations will be used in public facing research findings.

### 4. Data Organization and Documentation [↗](#)

In Github and ReData, the data will be de-identified and have an accompanying metadata file detailing data structure, variable definitions, and data processing steps.

### 5. Data Preservation and Long-Term Access [↗](#)

This collected data will be saved in a private repository as well as in University of Arizona's Research Data Repository, ReData. As this data will be considered my testing data in the machine learning models similar social movement data found on social media will be used as the training data. I hope to keep this data in an open repository to provide more diverse collection of social movement data for other research to learn from.

### 6. Data Sharing and Access [↗](#)

The participants will be provided with links and access to the data to be used in their own research. The focus group data will not be kept for future research. Data collected beforehand and after the focus group will be saved. This collected data will be saved in a private repository as well as in University of Arizona's Research Data Repository, ReData.

### 7. Data Ethics and Legal Compliance [↗](#)

Ethics approval is in process with the University of Arizona IRB board. All data collection follows ethical guidelines based on FAIR and CARE principles.

Considerations for managing your data incorporate many aspects, even some that might not even be mentioned here. A general understanding of what you need to manage your data and remember what it is in the future can be to know the: type of data you have - can be numeric, categorical, ordinal, or even multimodal. How you collect your data (document the steps). Metadata (the details), where and why you store your data where you do, any analysis or summaries of the data, preservation or maintenance of keeping that data. And any ethical implications that should be considered.

# Version Control and Reproducibility

Version control refers to keeping track of the version of a file, set of files, or a whole project.

Reproducibility comes from our obligations and desires to work ethically, honestly, and with confidence that the data and knowledge we produce is done has integrity.

- Repeatable, shareable, understandable



For me, the real value of this program forced me to get into the habit of saving everything!

Don't let data loss happen to you!

Git and Github are the most popular ways to work in version control and reproducibility.

Others are also:

- Microsoft Office's [Track Changes](#) functionality
- Apple's [Time Machine](#)
- Google Docs' [Version History](#)
- [Git](#)

Reproducibility is also a "spectrum of practices", not a single step.

- Docker containers
- Also the practice of making your content memorable i.e. will you remember what you were doing if you take a long break? Will people be able to understand what you were doing if they come to it with fresh eyes?

# Planning for the Dissertation

Creating an Open Science  
Research Space

---

Currently in Data Collection Mode and IRB  
Application



Artist: Agnes Toth; <https://mymodernmet.com/agnes-toth-fine-art-paintings/>

**Narratives Help: We receive small bits of information that represent a story from one or various perspectives**

### **Social Media Platform Fragmentation**

Some research explores the idea that the mediatization of social media has increased fragmentation of information among social media sites because of the affordances provided by these platforms (Sadler, 2021).

### **Bias and Interpretation**

Some say fragmentation leads to a biased perspective of a real-time event that could be impactful to a community (Chen et al, 2020) and in turn enable a fragmented interpretation of what is happening (Sanchez-Moreno et al, 2020; Mosseri, 2021).



## Mixed methods approach with Combined data

**Data Collection** - Instagram, Government, Non-Governmental Institutions, Citizen Data, and Datasets for Computational Models

## Methods

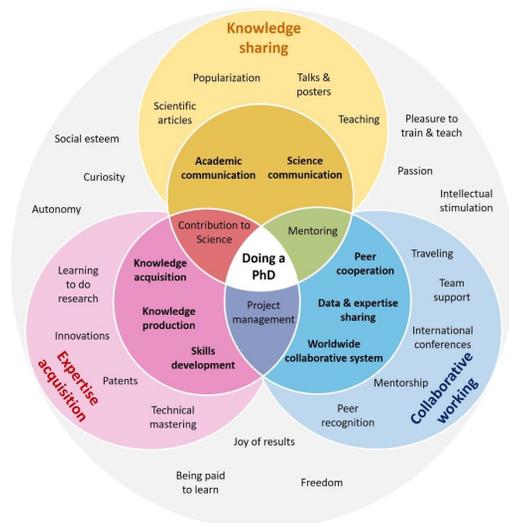
1. Network Narrative
2. Visual Interpretation and Focus

1. Group
2. Knowledge Graph Building

# Open Science Implications

## This project's need for:

- Data Ethics while sharing/maintaining data
- Interpretability of analysis for users
- Repeatability for Researchers and Practitioners
  - E.g. Well Commented Code
  - Detailed methodology and analysis
- Links to resources used (data, tutorials, literature)



Bernery et al. (2022) Figure 2: The positive aspects of doing a PhD.

So what does this have to do with open science?

- **Data Ethics while sharing/maintaining data:** My research constitutes using public data and collecting data from public social media accounts.
- **Interpretability for Focus Group Participants:** I will also work with members of the social movement and hope to have my findings visualized and easily understandable for participants to use in discussion.
- **Repeatability for Researchers and Practitioners:** I hope my research can be an example for future practitioners and expand on the topic or methods used.
  - E.g. Well Commented Code
  - Detailed methodology and analysis
- **Links to resources used (data, software/packages, literature):** Might as well share what I've found and not let my work be done in vain

# Applying Open Science to (My) Dissertation

- Data Analysis
    - Downloadable document(s)
    - Interactive graphs and visualizations
    - Open code
  
  - Data Management
    - Governance
    - Data Plan
    - Reproducibility
-

# Downloadable Documents



The screenshot shows a website with a dark blue header. On the left, there is a navigation menu with the following items: 'Social Movement Narrative' (with a book icon), 'Home', and 'Code'. The main header contains the text 'Social Movement Narrative' and a search bar with a magnifying glass icon and the word 'Search'. On the right side of the header, there is a logo consisting of a diamond shape with a book icon inside, followed by the text 'SM\_Narrative'.

The main content area has a white background. On the left side, there is a vertical navigation menu with the following items: 'Social Movement Narrative', 'Home', and 'Code'. The main content area contains the following text:

## Social Movement Narrative: Identifying frames in online social movement images.

Space to build the informational website for the research project: Narrative Frames in Topical Images: Mixed Methods Approach to Understanding Social Movements

[Download the dissertation prospectus](#)

*This is part of the FOSS 2023 FOSS Capstone that will be used for dissertation research!*

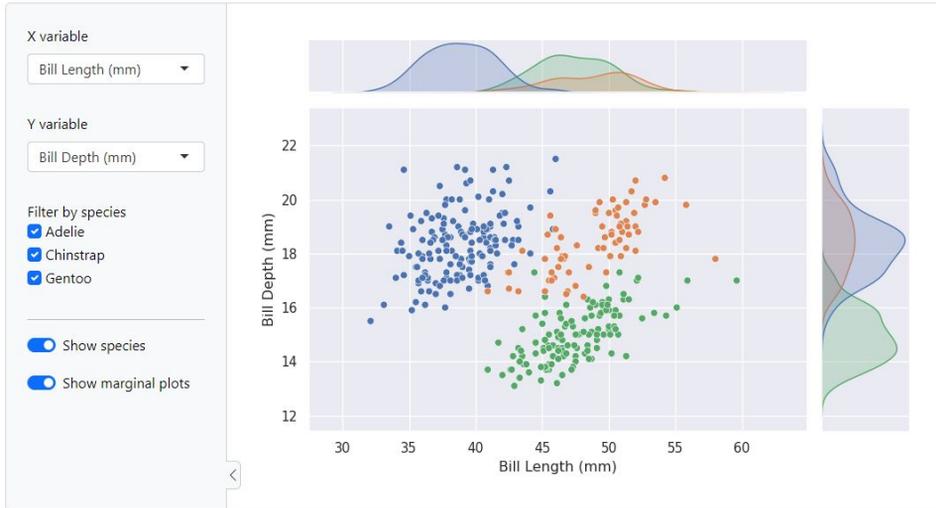
---

Last update: 2023-11-07

My contribution will be via a github website (open source) that works off its own repository/version control but enables us to create a website that will (hopefully) be user friendly.

Meaning, downloadable documents and code, and interactive plots!

# Interactive Graphs and Visualizations



Shiny (Open source)

The screenshot displays the Instaloader application interface. On the left, there is a sidebar with navigation options: Overview, Repositories, Projects, Packages, and People. The main area shows a list of datasets with columns for Dataset, Project, Topic, Date Donated, Description, Characteristics, Subject Area, # Instances, Associated Task, Attributes Type, and # Views. The datasets listed include Corel Image Features, CMU Face Images, MNIST Database of Handwritten Digits, Facebook Comment Volume, Gesture Phase Segmentation, Pseudo-Periodic Synthetic, Optical Recognition of Handwritten Digits, Internet Advertisements, Smartphone Dataset for Android, ImageNet, CIFAR-10, Pedestrian in Traffic Dataset, News Popularity in Multiple Languages, YouTube Spam Collection, Swarm Behaviour, and Skin Segmentation.

Below the dataset list, the logo for the National Institute of Statistics and Geography (INEGI) is displayed, along with the text "Government agency".

Types of data being collected:

Instagram Data

Government, Non-Governmental Institutions, Citizen Data

Datasets for Training Computational Models

Although Open Science is free, space is not.. A sample dataset will be available after processing and cleaning.

The data process will also be de-identified and documented for research reproducibility

Where you found/will find some open data for this project

Open packages: instaloader; shiny

The screenshot shows a GitHub repository page for a document titled "dplPy-for-FOSS Governance and Operations Document". The page includes a navigation bar with options like "Preview", "Code", and "Blame", along with file statistics (103 lines, 64 loc, 8.97 KB) and a GitHub Copilot badge. The main content area features a "Table of Contents" with sections: 1. Introduction (sub-sections: Social Movement Narrative Dissertation, FOSS Reference Hub, Dissertation Committee), 2. Procedures, 3. Community Practices (sub-sections: Open Science Commitment, Land Acknowledgement, Diversity Statement, Code of Conduct), and 4. Attribution, Authorship, and Ownership. Below the table of contents is an "Introduction" section with a paragraph defining the project's purpose. At the bottom, there is a section titled "Social Movement Narrative Dissertation" with a partial paragraph starting "Abstract Social movements have been using social media to share their goals, values, and grievances with the world for the past decade. This".

This document shows the guidelines for how I will handle the data:

Project Governance is the set of rules, procedures and policies that determine how projects are managed and overseen.

"The set of policies, regulations, functions, processes, procedures and responsibilities that define the, management and control of projects, programmes or portfolios." - [APM \(2012\), open.edu](#)

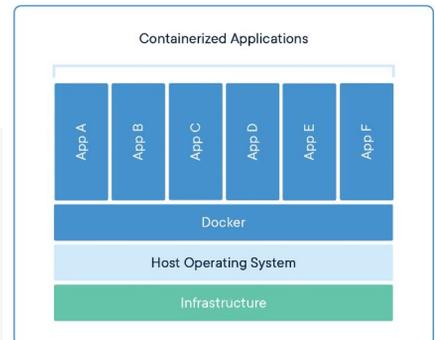
[Example gov. Documents \(Featuring Dr. Heidorn\)](#)

# Version Control and Containers for Reproducible Science

## Repository Organization

```
.
├── AUTHORS.md
├── LICENSE
├── README.md
├── mkdocs.yml
├── requirements.txt
├── code
│   ├── src
│   │   ├── *.py
│   │   └── execution_sample.ipynb
│   ├── Dockerfile
│   └── tests_data
│       ├── csv
│       └── rwl
├── docs
│   ├── assets
│   ├── stylesheets
│   ├── index.md
│   ├── installation.md
│   ├── manual.md
│   ├── Data_Management_Plan.md
│   └── Governance_Operations.md
└──
```

<- Governing file for website building  
<- Requirements file for pip installation (required by website)  
<- Python files part of dp1Py  
<- runnable example (executable Jupyter notebook)  
<- Docker script in charge of container creation  
<- Data from third party sources used for testing (in rwl and csv formats).  
<- Folder for images and additional graphic assets  
<- Folder containing style-related code for the website  
<- Main website home page  
<- Installation steps for dp1Py  
<- Manual for dp1Py  
<- Data Management Plan (example) applicable for this repository  
<- Governance & Operations (example) file applicable for this repository



## CONTAINERS

Containers are an abstraction at the app layer that packages code and dependencies together. Multiple containers can run on the same machine and share the OS kernel with other containers, each running as isolated processes in user space. Containers take up less space than VMs (container images are typically tens of MBs in size), can handle more applications and require fewer VMs and Operating systems.

This is the set-up of my repository which enables easy version control of the dissertation website, as well as a safe space for the code to be saved for future use.

Containers are helpful here because they essentially package your code (library and package installations and all) and send it to you to run and use on your own machine via docker. Once you have docker installed, your code is in one format that will work on all operating systems the same way.

Software containers, such as those managed by Docker(Cyverse) or Singularity (HPC), are incredibly useful for reproducible science for several reasons:

**Environment Consistency:**

Containers encapsulate the software environment, ensuring that the same versions of software, libraries, and dependencies are used every time, reducing the "it works on my machine" problem.

**Ease of Sharing:**

Containers can be easily shared with other researchers, allowing them to replicate the exact software environment used in a study.

**Platform Independence:**

Containers can run on different operating systems and cloud platforms, allowing for consistency across different hardware and infrastructure.

**Version Control:**

Containers can be versioned, making it easy to keep track of changes in the software environment over time.

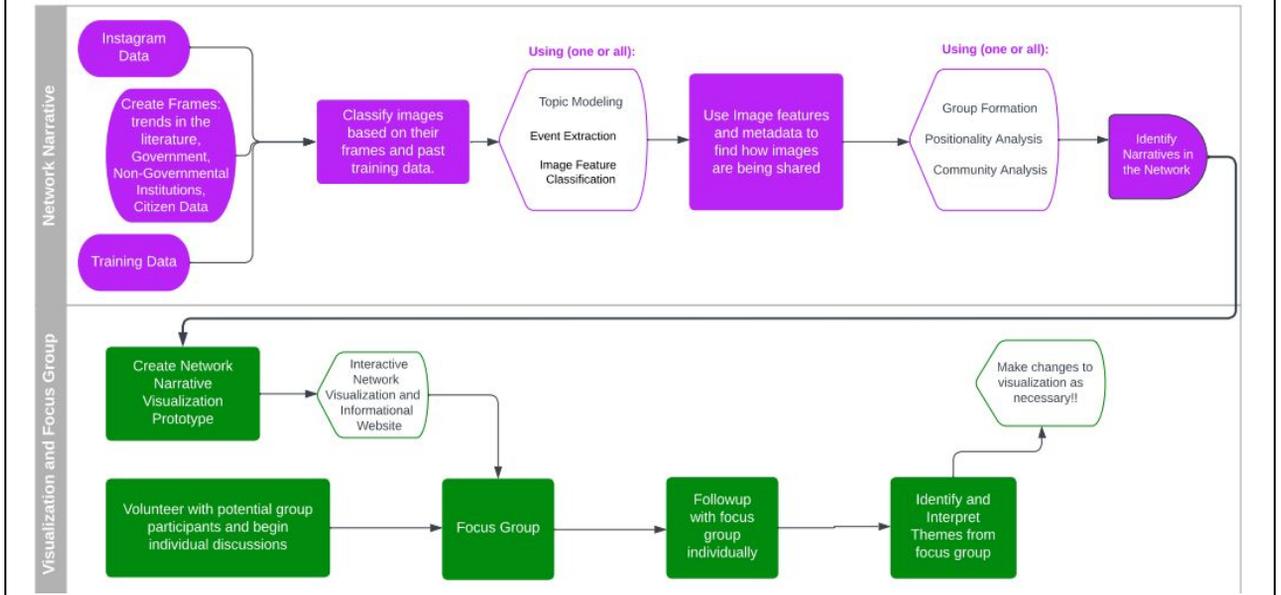
**Scalability:**

Containers can be easily scaled and deployed on cloud infrastructure, allowing for reproducible science at scale.

Isolation:[¶](#)

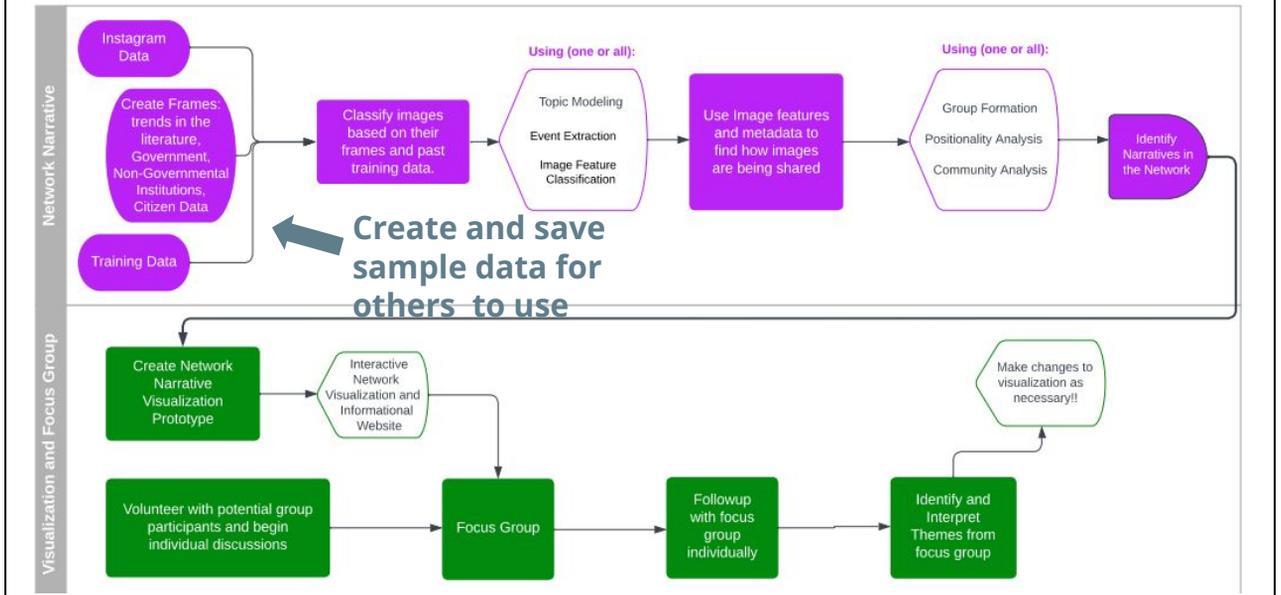
Containers isolate the software environment from the host system, reducing the risk of conflicts with other software and ensuring a clean and controlled environment.

# Network Narrative Data Workflow



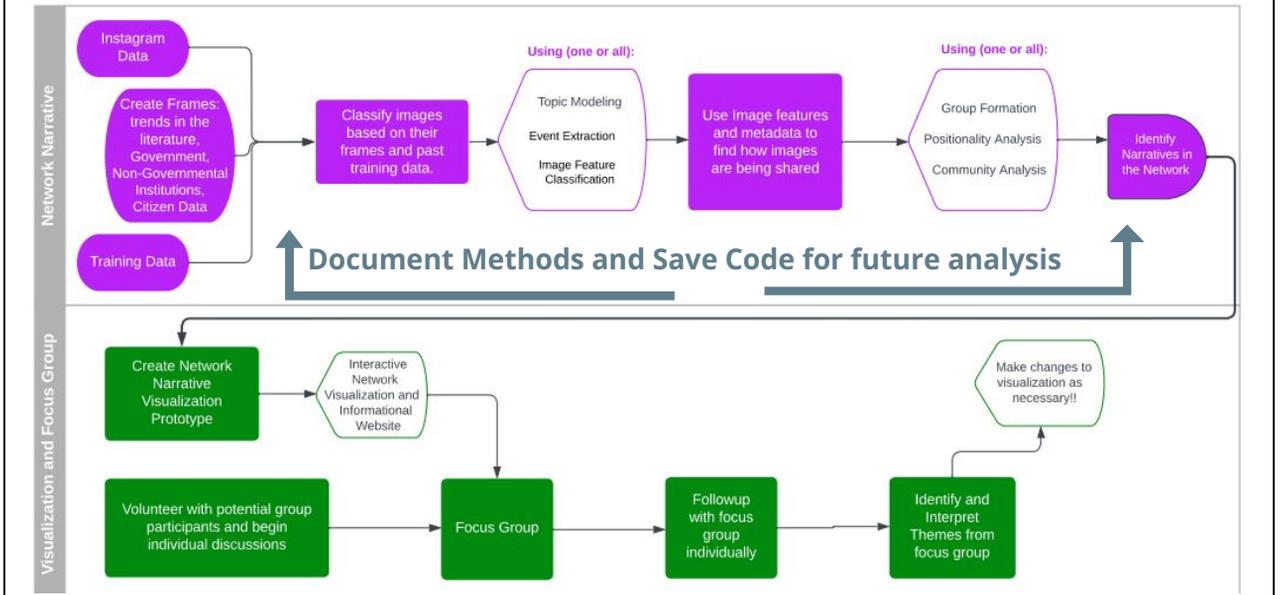
The data pipeline shows how the first parts of the data project will be managed and what aspects of the workflow fit with Open Science

# Network Narrative Data Workflow



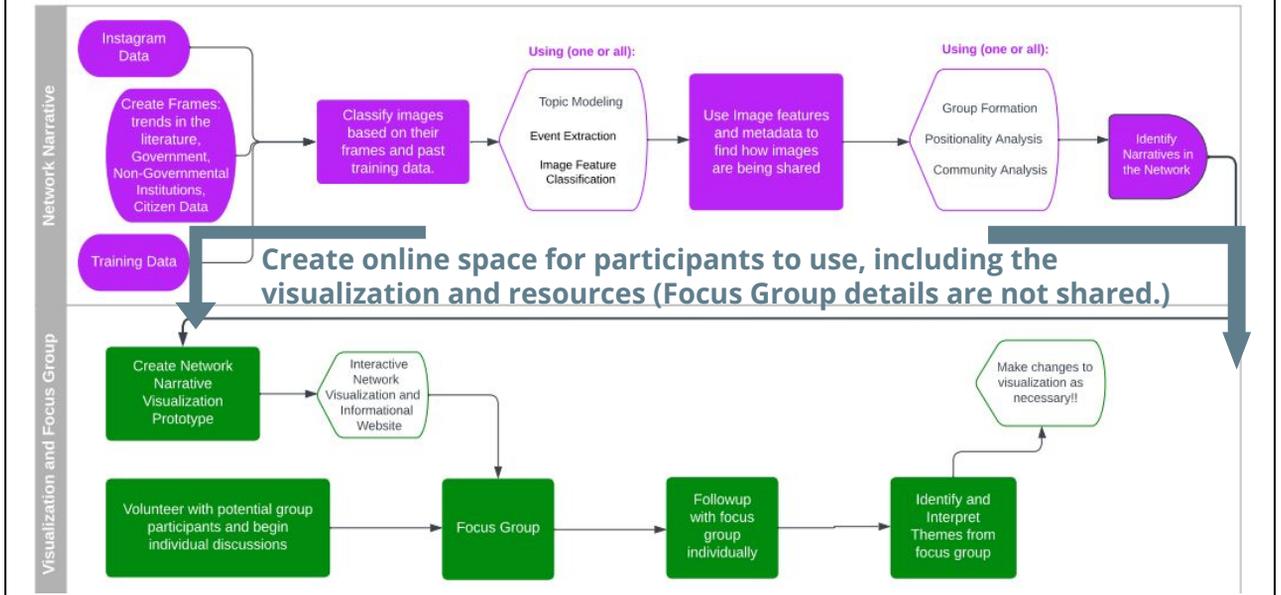
The data pipeline shows how the first parts of the data project will be managed and what aspects of the workflow fit with Open Science

# Network Narrative Data Workflow



The data pipeline shows how the first parts of the data project will be managed and what aspects of the workflow fit with Open Science

# Network Narrative Data Workflow

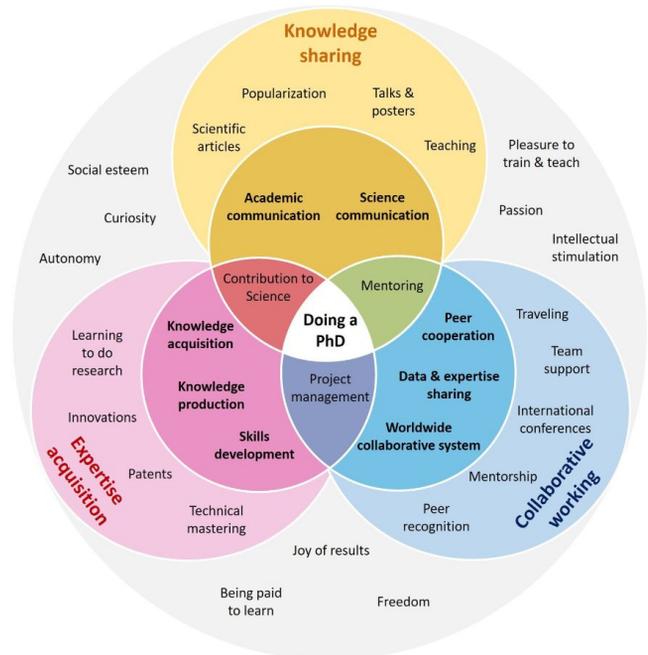


The data pipeline shows how the first parts of the data project will be managed and what aspects of the workflow fit with Open Science

# Open Science For Research

## Some of many open science uses:

- Data Ethics while sharing/maintaining data
- Interpretability of analysis for users
- Repeatability for Researchers and Practitioners
  - E.g. Well Commented Code
  - Detailed methodology and analysis
- Links to resources used (data, tutorials, literature)



Bernery et al. (2022) Figure 2: The positive aspects of doing a PhD.

So what does this have to do with open science?

- **Data Ethics while sharing/maintaining data:** My research constitutes using public data and collecting data from public social media accounts.
- **Usability for Focus Group Participants:** I will also work with members of the social movement and hope to have my findings visualized and easy for participants to understand and use for better discussion.
- **Repeatability for Researchers and Practitioners:** I hope my research can be an example for future research and expand on the topic or methods used.
  - E.g. Well Commented Code
  - Detailed methodology and analysis
- **Links to resources used (data, software/packages, literature):** Might as well share what I've found and not let my work be done in vain

# Thank You to R4R and iColloquium



Use the QR Code to get the slides!

# Extra Slides



# Ethics of Open Science

<b>To be findable:</b>	F1. (meta)data are assigned a globally unique and eternally persistent identifier. F2. data are described with rich metadata. F3. (meta)data are registered or indexed in a searchable resource. F4. metadata specify the data identifier.
<b>To be accessible</b>	A1. (meta)data are retrievable by their identifier using a standardized communications protocol. A1.1 the protocol is open, free, and universally implementable. A1.2 the protocol allows for an authentication and authorization procedure, where necessary. A2 metadata are accessible, even when the data are no longer available.
<b>To be interoperable</b>	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.
<b>To be re-usable:</b>	R1. meta(data) have 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 their provenance. R1.3. (meta)data meet domain-relevant community standards.

If you hope to make your data and processes actually usable, it's good to have some guidelines in place. The FAIR principles can be those guidelines, but it is good to note that they can be changed and enhanced depending on the project and your data. When collecting, saving, and managing your data it's important to follow some type of ethics, especially if you're working with sensitive information. Other principles like CARE and even some design justice and data feminism principles are important to look into.

- <https://mitpressonpubpub.mitpress.mit.edu/data-feminism>
- <https://designjustice.mitpress.mit.edu/>

<https://www.go-fair.org/fair-principles/>

## CARE Principles

### Collective Benefit

- C1. For inclusive development and innovation
- C2. For improved governance and citizen engagement
- C3. For equitable outcomes

### Authority to Control

- A1. Recognizing rights and interests
- A2. Data for governance
- A3. Governance of data

### Responsibility

- R1. For positive relationships

- R2. For expanding capability and capacity
- R3. For Indigenous languages and worldviews

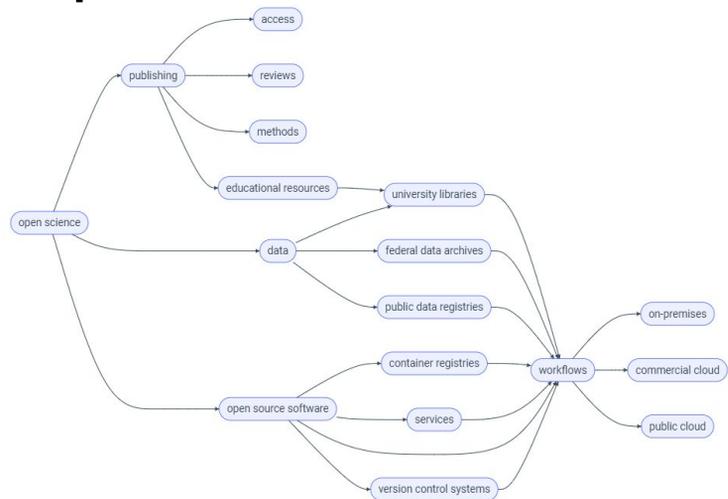
#### Ethics

- E1. For minimizing harm and maximizing benefit
- E2. For justice
- E3. For future use

# Open Science Implications

## The need for:

- Data Ethics while sharing/maintaining data
- Usability for Focus Group Participants
- Repeatability for Researchers and Practitioners
  - E.g. Well Commented Code
  - Detailed methodology and analysis
- Links to resources used (data, tutorials, literature)



Mermaid Diagram: Conceptual relationships of Open Science and cyberinfrastructure

- **Data Ethics while sharing/maintaining data:** My research constitutes using public data and collecting data from public social media accounts.
- **Usability for Focus Group Participants:** I will also work with members of the social movement and hope to have my findings visualized and easy for participants to understand and use for better discussion.
- **Repeatability for Researchers and Practitioners:** I hope my research can be an example for future research and expand on the topic or methods used.
  - E.g. Well Commented Code
  - Detailed methodology and analysis
- **Links to resources used (data, software/packages, literature):** Might as well share what I've found and not let my work be done in vain

# Include a Personal Website!

lvdozal.github.io

**Laura W. Dozal**



PHD Candidate at the University of Arizona's School of Information (School). Methods Expert in Network Analysis, Machine Learning, and Data Mining

**Data Science: Projects and Publications**

**Dissertation Research: Feature Extraction for Network Visualization, Focus Groups with Human Computation, and Knowledge Graph Creation**

- Createad a customized image and metadata collector script using the Installoader package built in Python. Collected 2,000+ images and metadata from Instagram.
- Narrative Frame Network Visualization (*forthcoming*)
  - Image feature extraction and information diffusion analysis
- Visual Interpretation and focus group (*forthcoming*)
- Knowledge Graph Implementation (*forthcoming*)
  - Knowledge base and scene graph building

[Python](#) [Instagram](#) [Installoader](#) [Scrapy](#) [Spyder](#) [Anaconda](#)

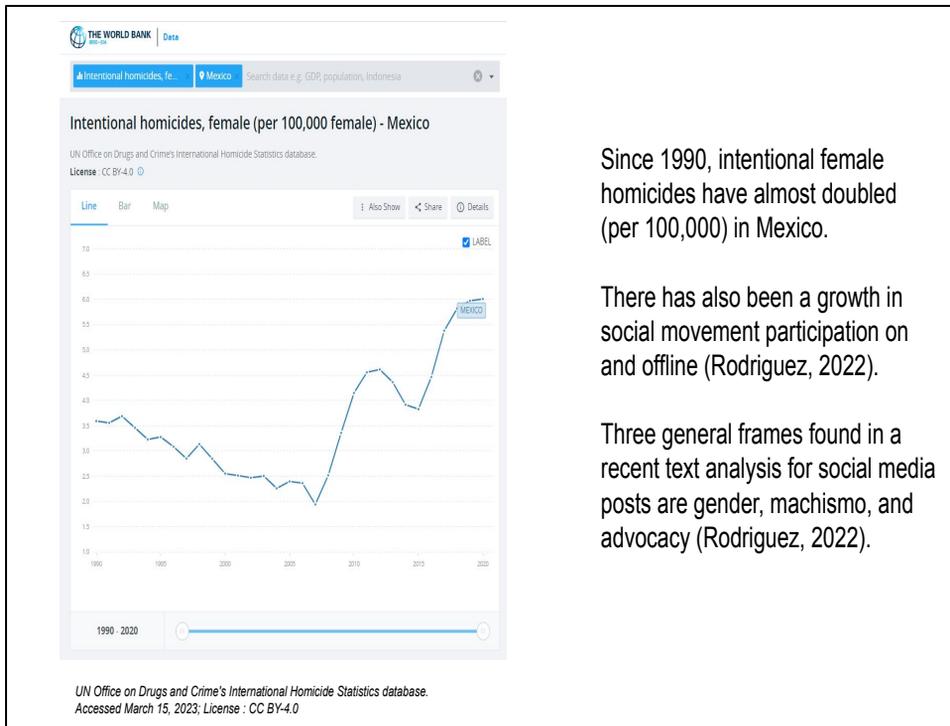
**Street Art Network Analysis: Applications of Bi-Partite and Bi-Dynamic Line Graphs**

In this analysis, Street Art images are considered as a type of visual information that can represent a specific perception of a community as a member of a community space. Dynamic Bi-partite network analysis was used to understand how different neighborhoods are connected through artist attributes and how they might differ. The results show that specific neighborhood traits, urban, population, culture contribute to stronger ties within the Street Art community network. Street Art as Visual Information: Mixed Methods Approach to Analyzing Community Spaces

Download My CV  
[View My LinkedIn Profile](#)  
[View My GitHub Profile](#)

Hosted on GitHub Pages — Theme by orderedlist

This will help support you credibility, and it's good for resume building. You can link you dissertation to this website.



Since 1990, intentional female homicides have almost doubled (per 100,000) in Mexico.

There has also been a growth in social movement participation on and offline (Rodriguez, 2022).

Three general frames found in a recent text analysis for social media posts are gender, machismo, and advocacy (Rodriguez, 2022).

Why does the full narrative matter?

Since 1990, intentional female homicides have almost doubled (per 100,000) in Mexico.

With these numbers there have also been a growth in social movement participation on and offline (Rodriguez, 2022).

In 2020, over 20 thousand people marched in CDMX for anti-gender violence during March 8th international women's day ([El Universal, 2021](#)). Similarly, chants and other images of protest and advocacy have increased online (Edmé, 2020).

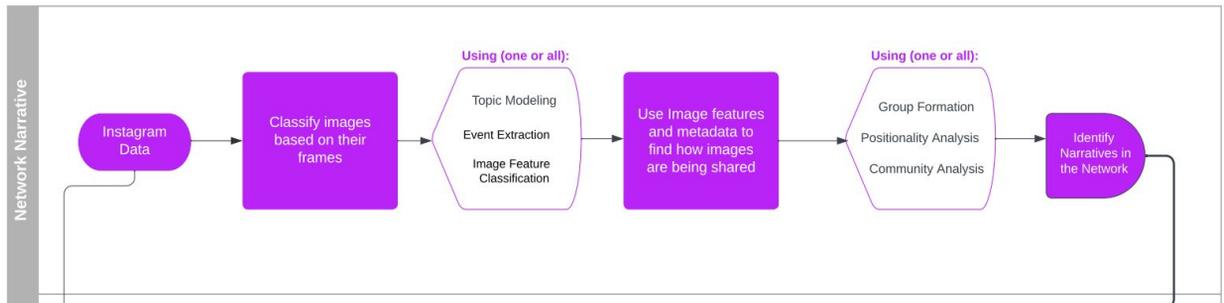
Essentially, Much of Mexico has used international women's day to protest and spread awareness for violence against women efforts

There have been three overarching frames found online through a recent text analysis by Rodriguez, 2022. These are gender, machismo, and advocacy.

But these frames do not implement a linear sequence of information messaging on a user's device. This fragmentation can potentially distort user perception, misinform the public, and impact collective action.

[Graph link](#)

# Phase 1: Network Narrative



I know I said knowledge graphs would be used in this step, but KGs can't identify structural aspects of a network, which is what I'd like to do here to understand which frames might be more prevalent or siloed. Here is where I will implement network analysis

Network example: site of production = Instagram; the image itself = the image content; circulation = metadata (and potential network findings); audiencing = users, comments hashtags

Methods used here are ...

RQ1: What aspects of social media enable narrative building from fragmented elements?

- Contextualization

Hypothesis 1 (H1.1): The affordances of Instagram allow for different types of narrative sharing including linear, non-linear, and interactive structures.

Hypothesis 1.2 (H1.2): Fragments of a narrative come about from various perspectives. These perspectives include geographical locations and keyword topics.

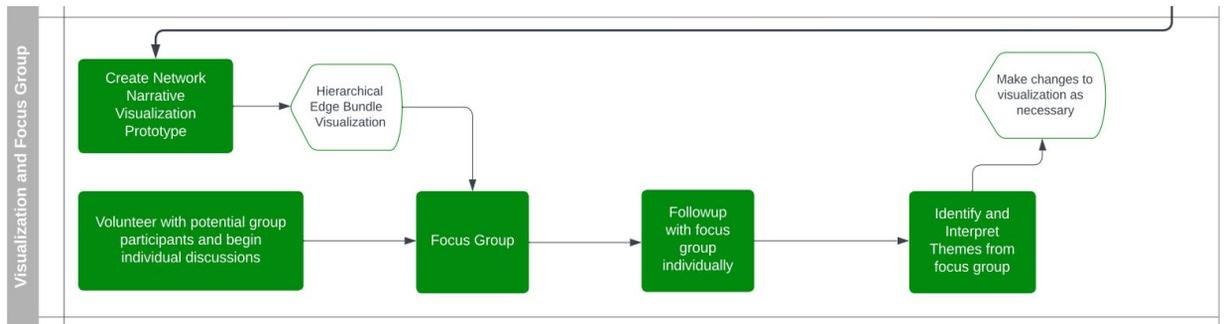
Hypothesis 1.2.1 (H1.2.1): Local and Global narratives are connected but have better contextualization when

focusing on community details.

Based on the data being collected for analysis, there might be aspects of the platform that enable particular types of fragmentation. These include geographical location and user activity.

This hypothesis looks at the date, keywords, and geotagging (when available) to understand how a network outline might be presented. It is argued that location has an impact on how a narrative is structured.

# Phase 2: Visual Interpretation and Focus Group



Plan on using the network analysis as a tool for members of the social movement to use

## **Build Relationships with organizations/potential focus group Participants**

Consider how they might use this tool, and identify themes and topics for analysis in the focus group.

The data collected from .gov, NGO, data citizen scientists will also be compared and support or question the narratives found and prepare for the focus group

## **Focus Group (around 4 or 5 participants)**

### *Implementation 1 (Synchronous)*

Focus Group Review of Narrative Network and properties using Visual Interpretation Modalities (all materials will be sent via email)

Discuss aspects of Discourse analysis, content analysis and other themes brought up by the group (focus groups will be implemented on zoom)

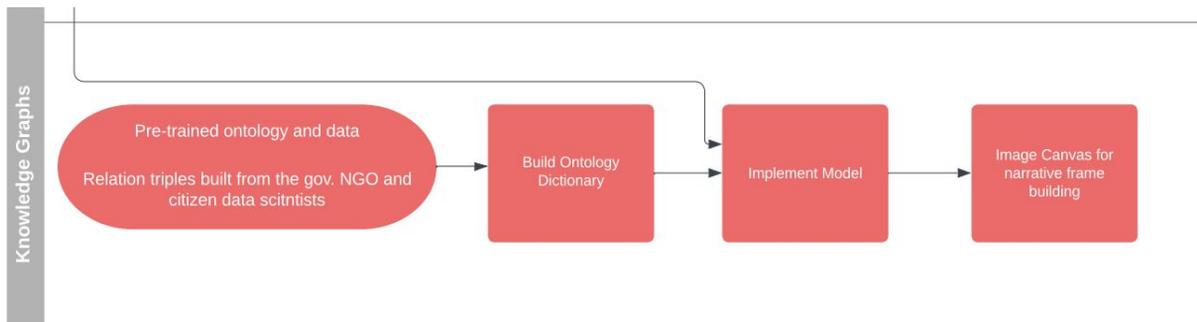
### *Implementation 2 (Asynchronous)*

The SIG-AH and SIG-USE workshop recently implemented an online contribution project where participants access the same content to discuss, interact and ask questions to each other and myself. Participants can also help shape the outcome of how the narratives might work.

## **Non Focus Group**

In the case the focus group doesn't work, I'm hoping to take to University listservs, social media posts, and connections in El Paso, Nogales, cd. Juarez, and Mexico city. Amazon Mechanical Turk would be nice if there is ever extra funding.

# Phase 3: Knowledge Graph Building



The knowledge graph method has academic implications that add to social movement analyses surrounding social media and theoretical concepts of framing and network analysis in the field. Similarly, this phase of the research adds to domain specific research using knowledge graphs and can be a generalizable approach to topical image datasets.

Using protest ontology (Bekker, 2021), narratology (narrative ontology), scene graph ontologies, themes found from the network narrative and focus group, and semantic triple relation building with data collected from gov. NGOs, and citizen data scientist, I will create an ontology dictionary for the KG model.

The model uses layer graphs to describe the ordering of elements in a narrative as well as the spatial relationships of objects in the image, and scene graphs are used to describe the semantic relationship between elements in an image (Fang et al. 2018). The narrative is generated by positioning elements in the images based on the constraints used in the dictionary input (Fang et al. 2018).

Other potential models include Spatio temporal models, (Long short-term memory model networks - LSTMs) that are sequence based for visual storytelling; relation reasoning and temporal graphs; and text-image combo, or other multimodal combos using hierarchical applications

# Tree Ring Data Pipeline

