Binghamton University
The Open Repository @ Binghamton (The ORB)

Library Scholarship | University Libraries

2-2021

Digital scholarship services on a budget: using open source methods for growth of research sharing

Amy Gay

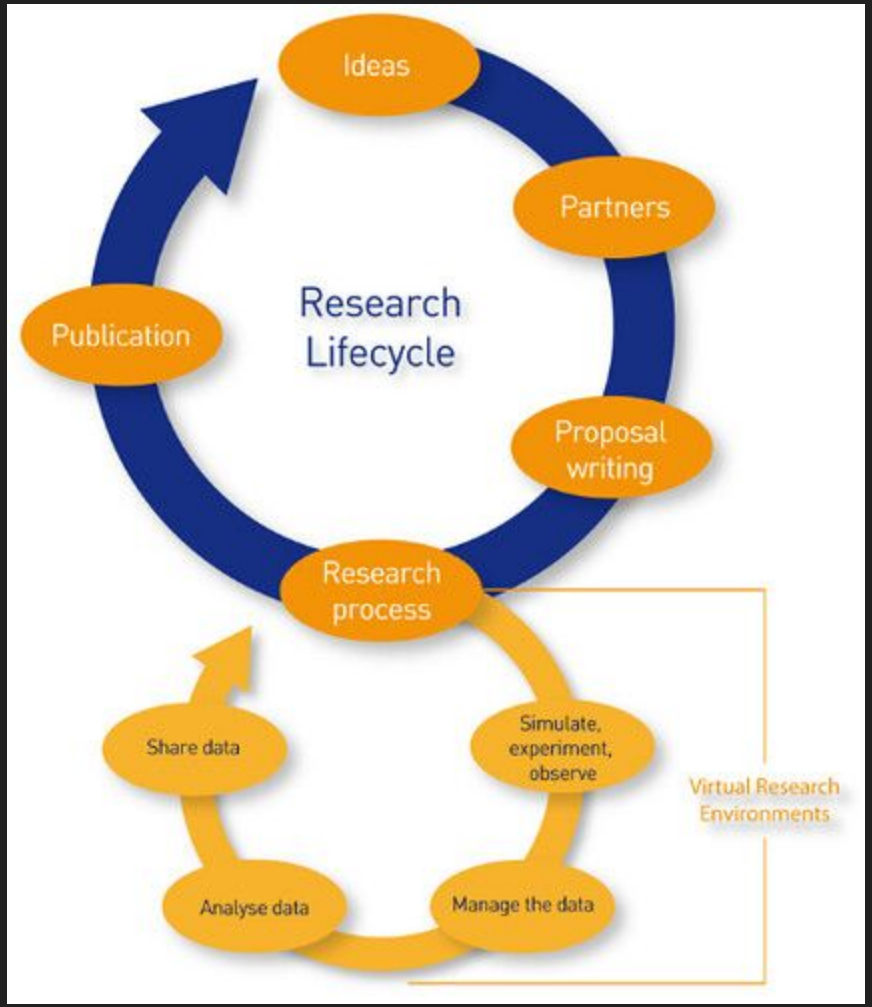
Follow this and additional works at: https://orb.binghamton.edu/librarian_fac

Recommended Citation
Gay, Amy, "Digital scholarship services on a budget: using open source methods for growth of research sharing" (2021). *Library Scholarship*. 60.
https://orb.binghamton.edu/librarian_fac/60

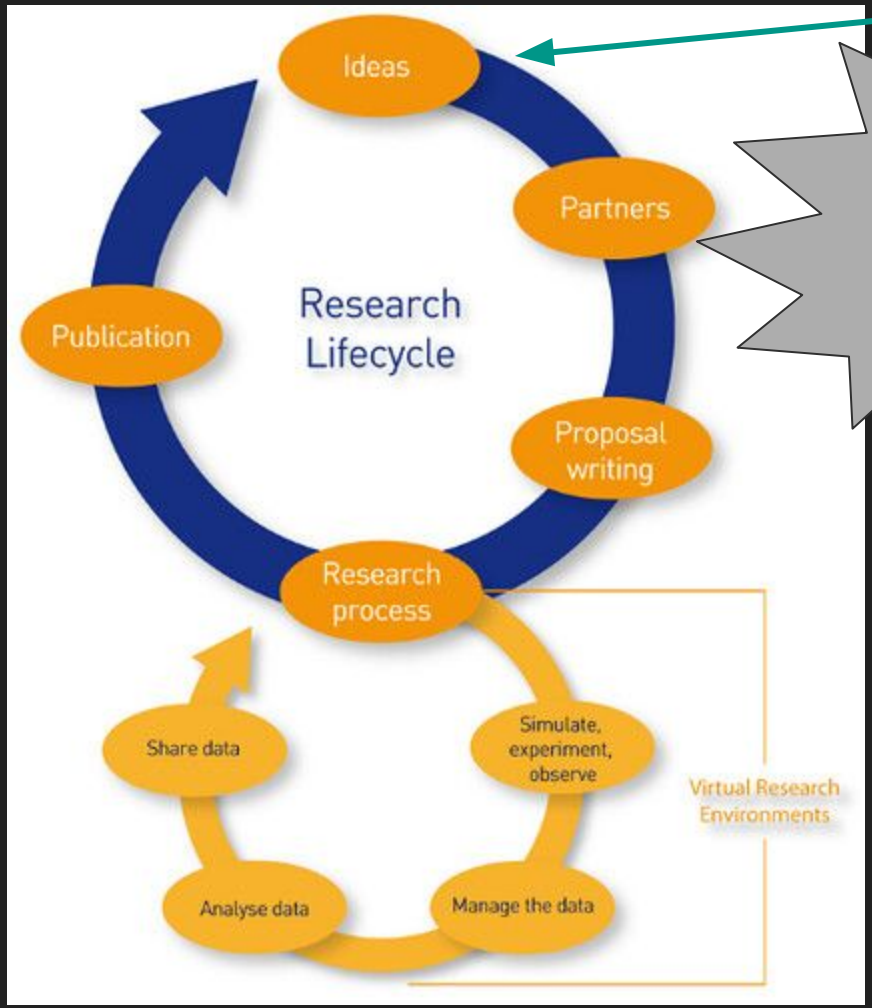
This Presentation is brought to you for free and open access by the University Libraries at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in Library Scholarship by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.

Digital Scholarship Services on a Budget: Using Open Source Methods for Growth of Research Sharing

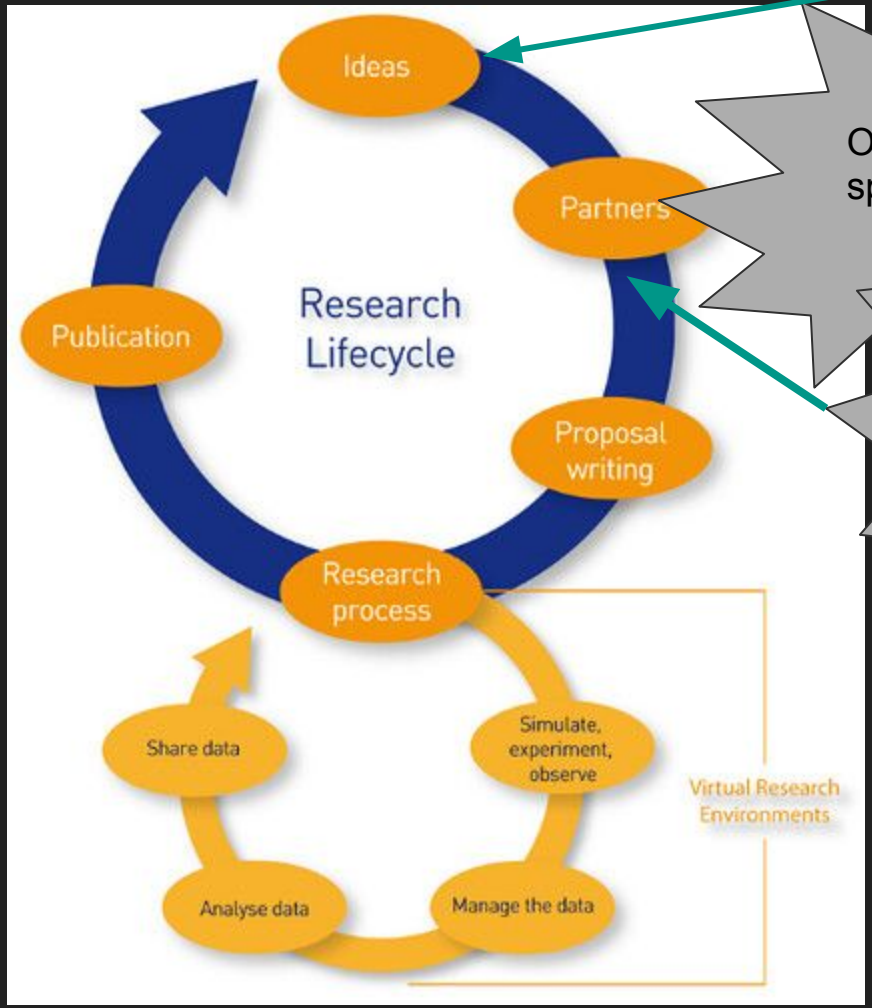
Amy Gay
Digital Scholarship Librarian
Binghamton University



"research lifecycle" by dff.jisc is licensed under CC BY-SA 2.0

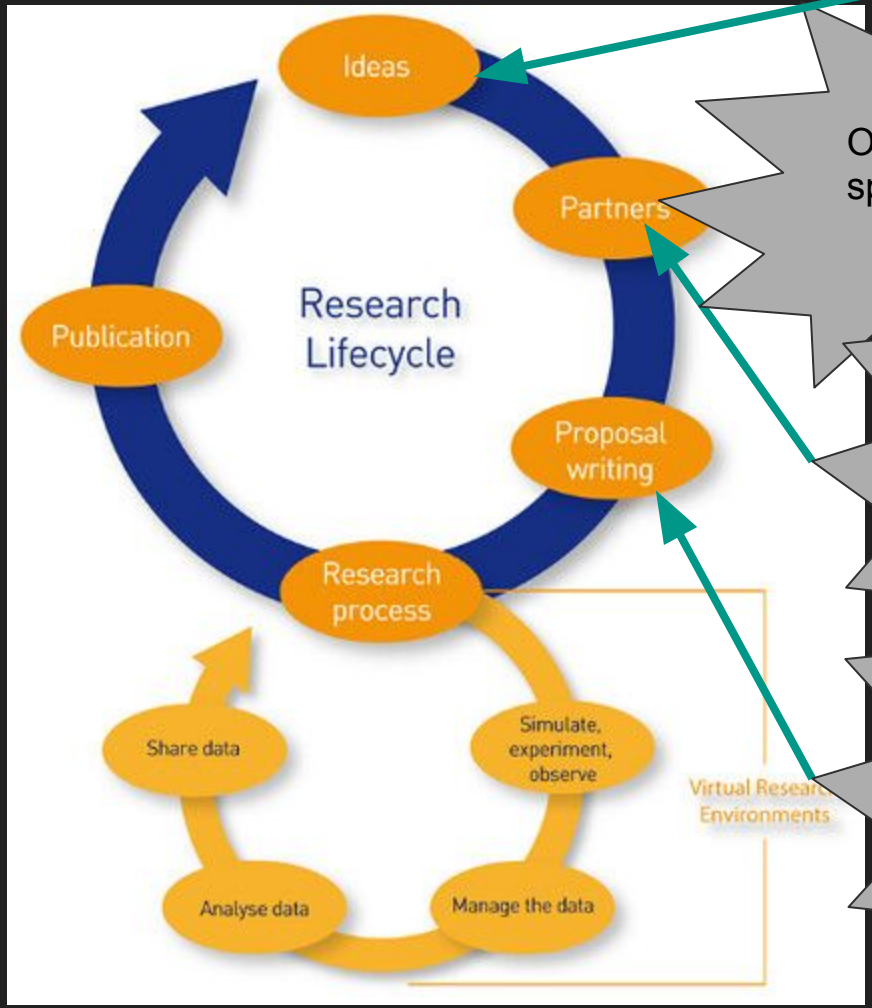


Open data to spark ideas



Open data to spark ideas

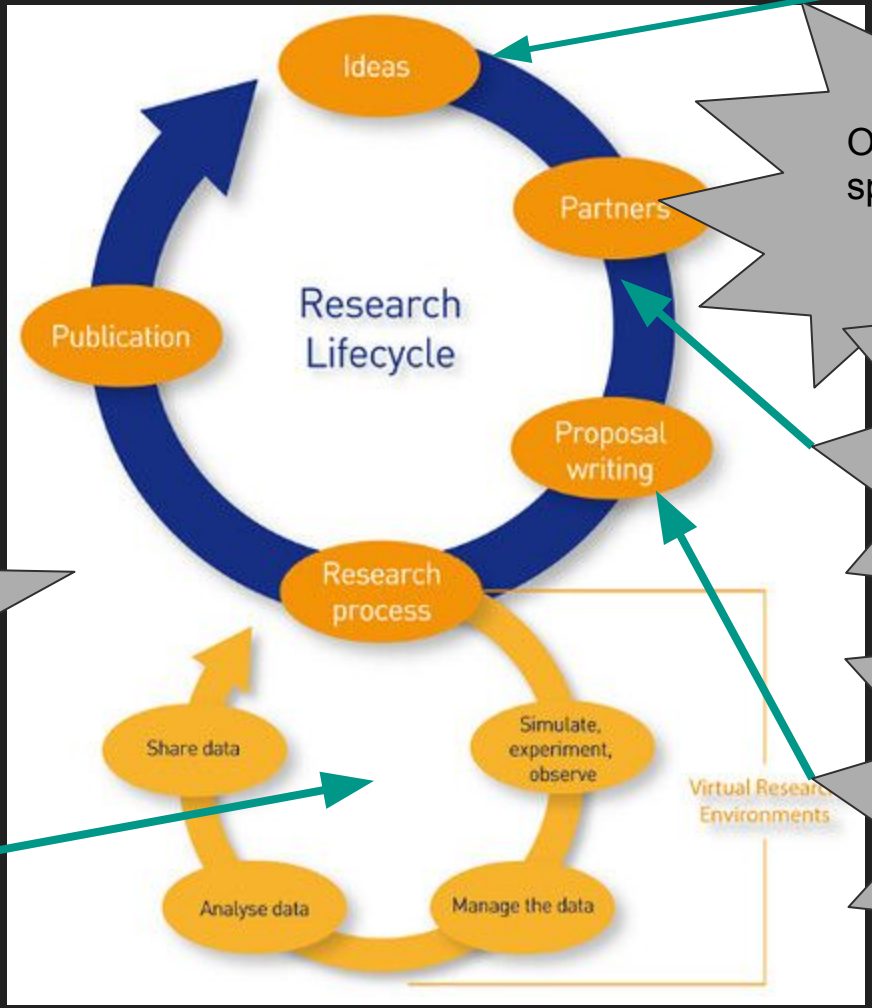
Liaison to find research partners (or be the partner)



Open data to spark ideas

Liaison to find research partners (or be the part r)

Data management plans and grant language



Open data to spark ideas

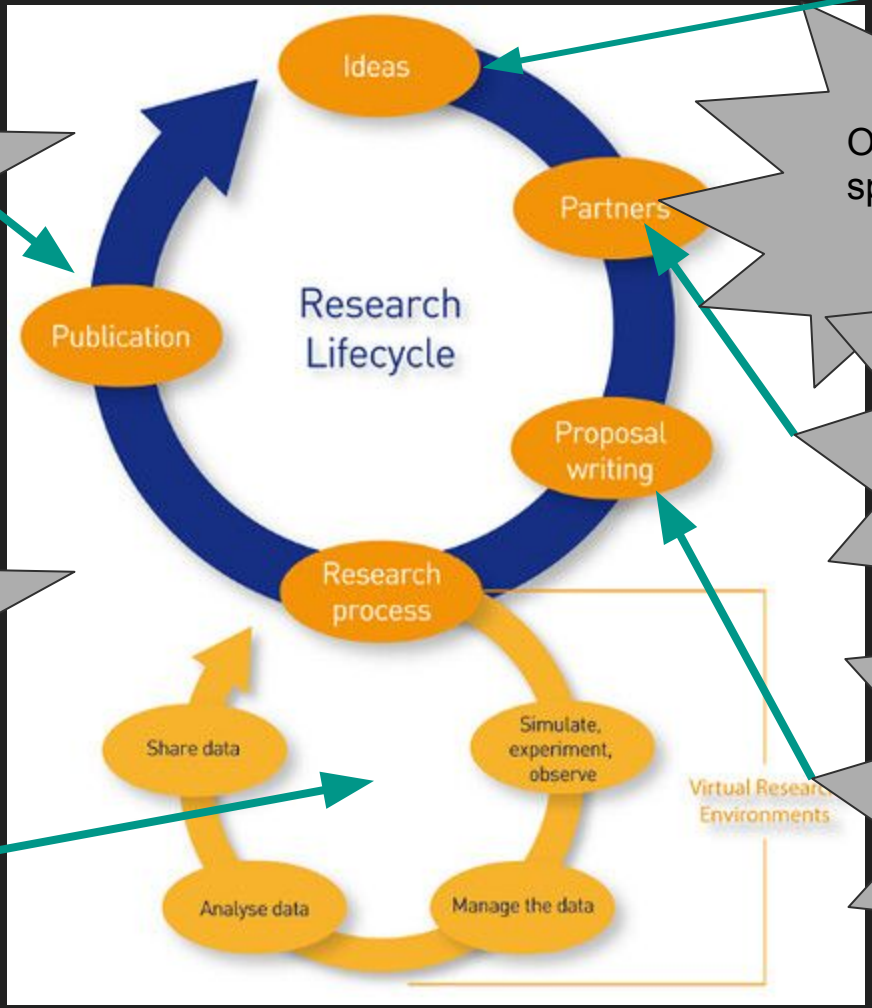
Liaison to find research partners (or be the partner)

Data management plans and grant language

Offer spaces, trainings, and access to materials and platforms

Repositories; open sharing; digital media

Offer spaces, trainings, and access to materials and platforms



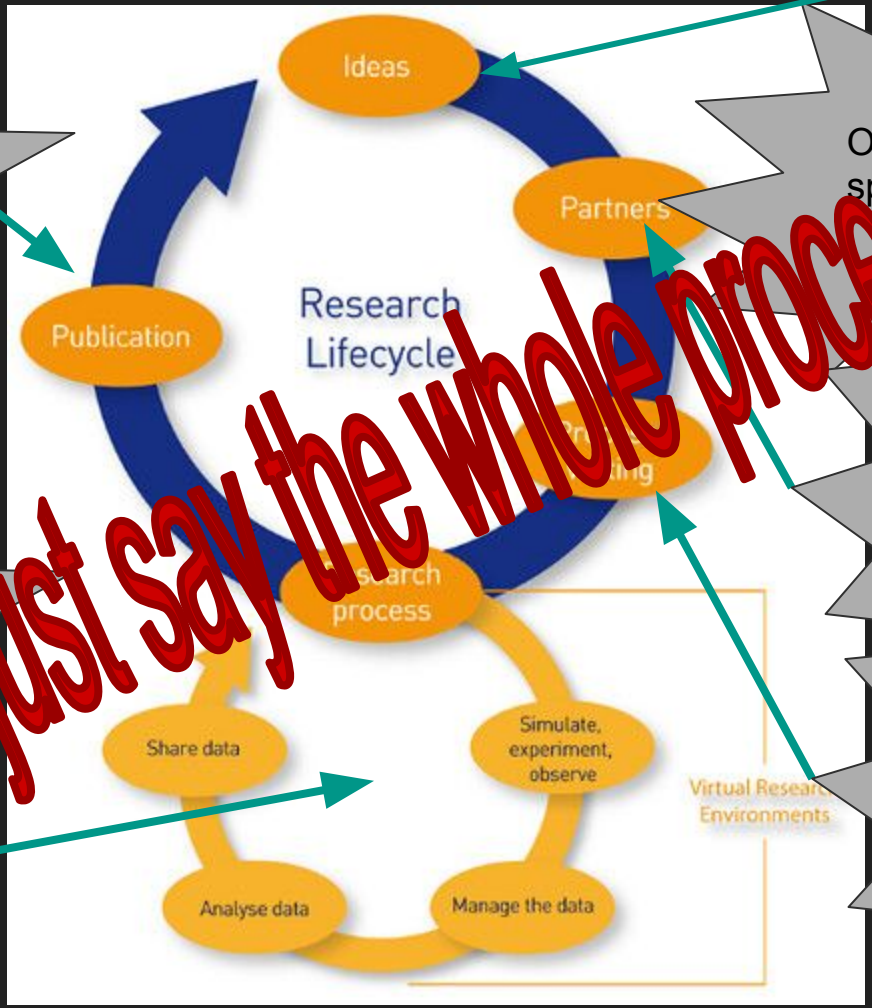
Open data to spark ideas

Liaison to find research partners (or be the partner)

Data management plans and grant language

Data management plans and grant language

Offer spaces, trainings, and access to materials and platforms



Open data to spark ideas

Liaison to find research partners (or be the partn

Data management plans and grant language

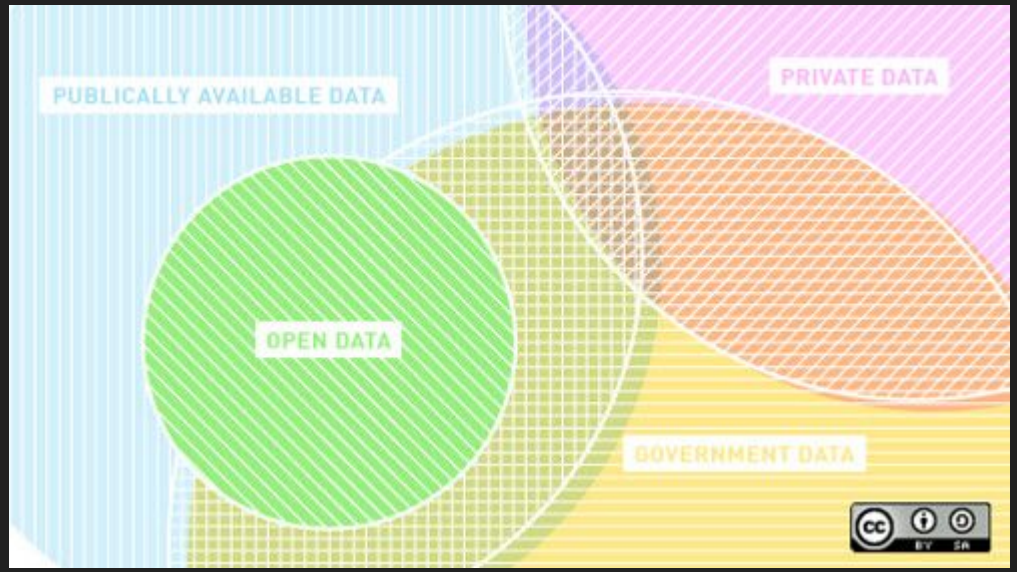
Let's just say the whole process

I. Open Source for Research Purposes

Ideas - Open Datasets

Open Datasets are data made publicly available and free to use without paywall or password restrictions

- This includes use, re-use, and re-distribution
- Can advance research in that field at a faster pace
- More efficient than constantly re-inventing the wheel



"What “open data” means – and what it doesn’t" by opensourceway is licensed under CC BY-SA 2.0

Where to find Open Datasets

GitHub Awesome Public Datasets (contains a section for Agriculture)

Kaggle Open Datasets (I ran a search for agriculture and 238 results came up)

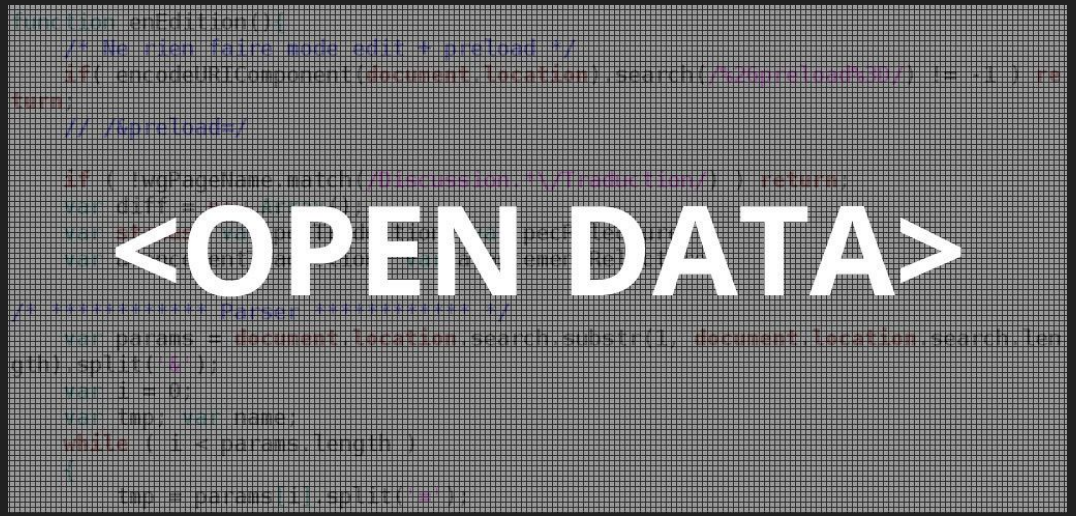
World Bank Open Data

Data.world (1,693 datasets tagged as agriculture currently)

Google Dataset Search

Unicef Data

World Health Organization Data Collection



"Open data" by thedescrier is licensed under CC BY 2.0

Other places to search for open datasets

Directory of Open Access Repositories

Registry of Open Data on AWS

U.S. Government Open Data

U.K. Open Data

NYC Open Data

Buzzfeed News Data

Reddit Datasets

UCI Machine Learning Repository

…. And many other repositories I am missing

Partners - Searching Profiles

ORCiD offers a way for researchers to create a unique identifier attached to their name

- Free to create a profile (although organizations can pay to be members)
- One-stop-shop for all research, grants, peer reviews, projects, and creative works
- Add C/V information
- Add research interests
- If people cite them using a particular format, their ORCiD ID will still be a way to connect their work back to them

Partners - Other places to search profiles

Google Scholar

Academia.edu

ResearchGate

Institutional repositories tend to have profiles of researchers attached - here is ours as an example: Expert Gallery

Social media platforms, like LinkedIn, Twitter, Facebook

Partners - Librarians as research partners

It can and does happen!

What to consider if you are asked to be a research partner or see it as a recommendation:

- What is the “ask”? Is it hardware? Software? Personnel support? Other?
- How much of your time (or how many librarians) are necessary and how much involvement required?
- Can funding be provided to the libraries from the grant (for many grants, this is seen as a positive if funding goes to those who are assisting in some way)

Do not be invisible labor -- know your worth and show your worth -- libraries and the librarians within them should be considered valuable resources

Proposal Writing - data management plans

The DMPTool is a free, open source application that helps researchers who are developing data management plans

- Libraries can set up accounts through link above for their institution or individuals can sign-up
- You can create your own installation at your institution by following instructions on their GitHub page
- Contains templates from common grant funders, walking researchers through required fields
- Contains generic templates for

  Developing DMPs as well

Proposal Writing - initial repositories considered

Does your library offer an institutional repository? If their research fits, this should be the first offer. If not, other options include:

- Open Science Framework (developed by the Center for Open Science)
  - Free, open source platform
  - Can work on a research project from inception to completion
  - Customize what is made publicly available and what is available to specific people
  - Offers collaboration across institutions
- Figshare (files up to 5GB)
- Zenodo (open source development, meant for big data projects)
- Ways to search for repositories (data-specific):
  - Scientific Data Recommended Repositories
  - Registry of Research Data Repositories (re3data)

Research Process - Data Cleaning

OpenRefine is a free, open source tool for data cleaning

- Install a desktop application to use (with browser based interaction during use)
- Able to keep data private until it is cleaned and ready for sharing
- Available in 15+ languages
- Suggests Wikidata as an option for data repository
- More resources:
  - The Programming Historian tutorial: https://programminghistorian.org/en/lessons/cleaning-data-with-openrefine
  - Thomas Padilla's Gettting Started with OpenRefine: https://thomaspadilla.org/dataprep/

Research Process - Visualization and Analysis

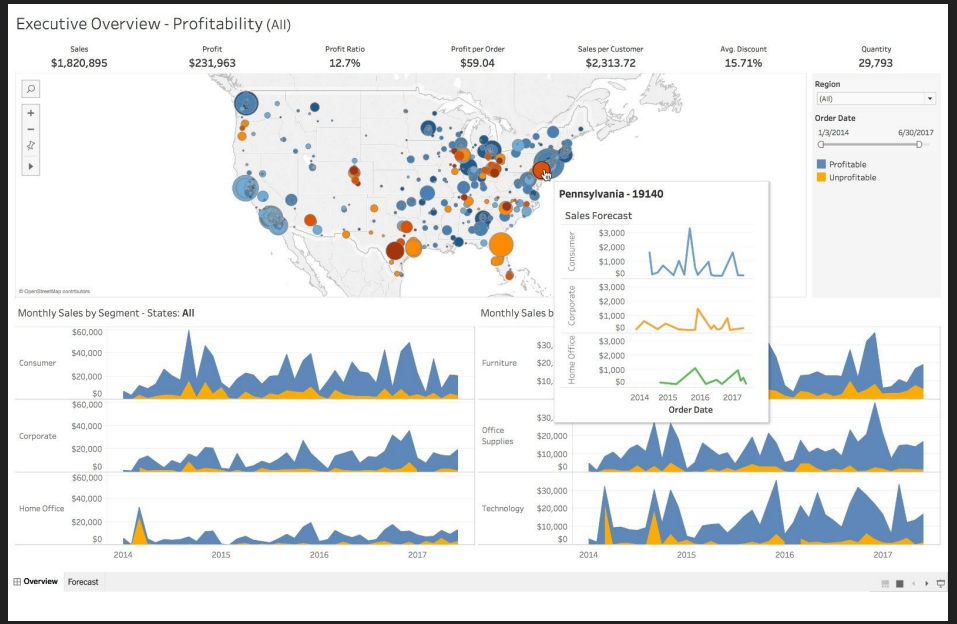
Tableau is software for visualizing and analyzing data

- Tableau Public is free to use, but your visualizations / analyses will be public on their site
- Tableau for Students and Educators

  is free and starts with a trial

  license

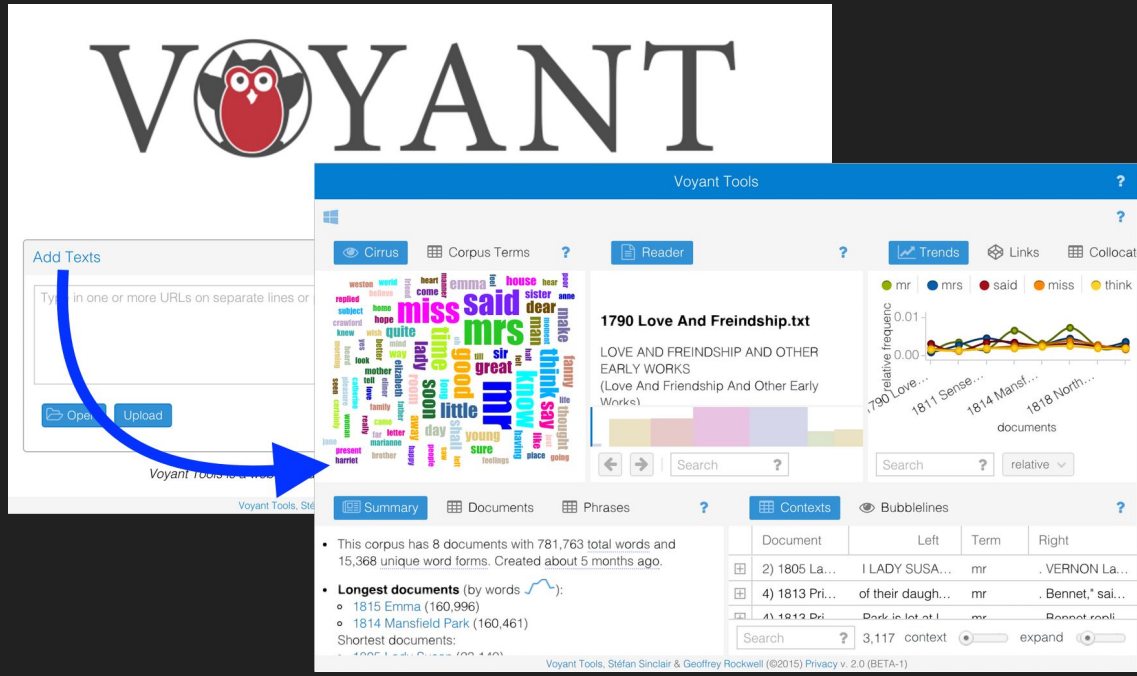
- Tutorial Videos
- TutorialsPoint
- Agricultural example



Research Process

Voyant is a free web-based text analysis tool

- Also works with datasets collected in .csv or .xlsx
- List of visualization and text analysis tools available
- View its Github source code
- Resources:
  - https://voyant-tools.org/docs/#!/guide/tutorial
  - https://www.youtube.com/watch?v=LWyMU0fWX58



Research Process - data cleaning, viz, and analysis

Python Programming

- Free, open source programming language
- Considered most human readable and easiest for first time learners
- Has packages available for data cleaning, visualization, and analysis
- Agriculture projects using Python: https://github.com/topics/agriculture?l=python

R programming

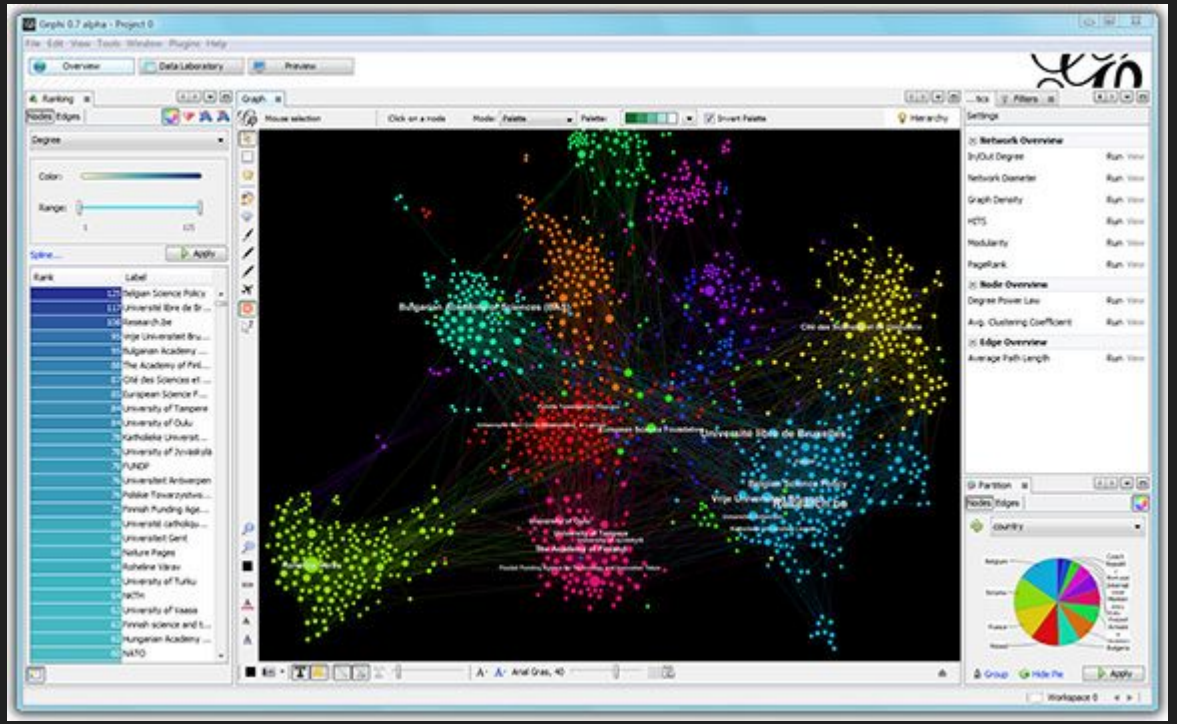
- Free, open source programming language
- Offers packages for data visualization and analysis
- Other common one used for working with data
- Agricultural tutorial / example: https://cran.r-project.org/web/packages/agriTutorial/agriTutorial.pdf

List of resources for both Python and R: https://github.com/pacian/Digital-Humanities-Toolkit#programming-languages-tutorials-programmingscripting-languages-used-to-facilitate-as-text-and-data-analysis-collection-preparation-etc-see-programming-languages-tools--resources

Research Process - network analysis

Gephi is a free, open-source network analysis software that is installed to your desktop

- Used for network analysis, data analysis, biological network analysis, and social network analysis
- Can be used for creating great visuals for poster presentations
- Agriculture research example using Gephi for network visualizations
- Intro to Gephi video tutorial: https://www.youtube.com/watch?v=2FqM4gKeNO4



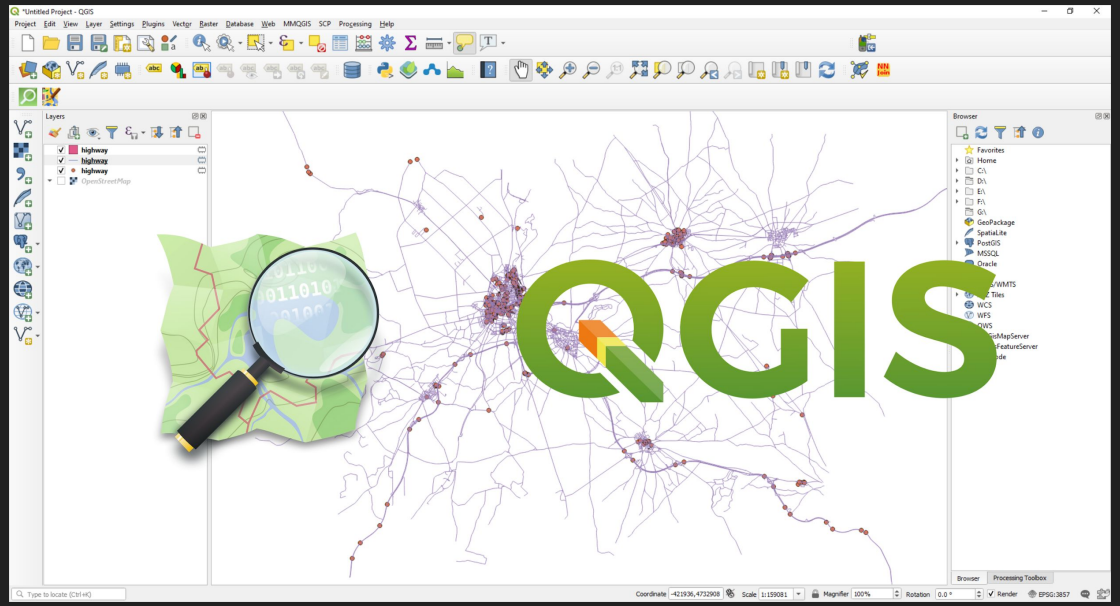
Publication

This can be done through the repositories we explored at the start of the process

- GitHub is another great option and no longer just for code
  - You can create a professional portfolio or profile
  - You can write open books (Tutorial)
  - As you have noticed, many of our resources come from collections here

Publication

QGIS is a free, open-source desktop platform for viewing, editing, and analyzing geospatial data (comparable to the paid Esri tool known as ArcGIS)



- Agriculture examples:
  - Create a Farm Map with Soil and Elevation Data Using QGIS
  - QGIS / Python plugins
  - Tutorial for Precision Agriculture
  - Tutorial for agricultural mapping

Publication

StorymapJS is a free web-based mapping tool to tell stories across pin-pointed locations

- You could develop an interactive map of livestock, plantlife, movement across locations during migration, etc.
- Tutorials:
  - https://vimeo.com/234435187
  - https://lindseywieck.org/fall_2016_sf/storymapjstutorial.html

III. What else can these budgeted services offer?

Bibliometrics

Bibliometrics combines various data and computational methods to analyze research impact

- Offers a way to analyze your own citation network and others citation networks
- These services could be offered in the libraries by combining skills from OpenRefine and Gephi
- Not just services, but workshops!
- Learn more from the tutorials provided by the U.S. National Institute of Health: https://www.nihlibrary.nih.gov/services/bibliometrics/bibliometrics-training-series

Pedagogy

Any of the tools we covered in these processes could be taught to graduate and undergraduates

- Co-teach the method with instructors
- Use for their own theses and dissertations
- Offers students a chance to build their own digital portfolio and research footprint

Train-the-Trainer Workshops

These skills could be offered not just as consultations but also as workshops

- Workshops to learn the skills for their own research
- Teach faculty how to teach these methods to their students
- Develop living toolkits for resource sharing

Wrap-up

- These open source tools are free and can help when on a tight budget and lack of resources
  - They are also suggestions - others may be more suitable
- Challenges can still arise with workload
- You do not have to do everything; learn what is best and what is doable
- If you can partner with other service points on campus, do so
- Take advantage of the free resources to learn and share
- Most important, have fun learning something new and share that knowledge with others

Thank you!

Connect with me on Twitter:
@iamabooknerd

Connect with me on LinkedIn:
https://www.linkedin.com/in/aeg15