Aid Analysis - Chinese and US aid to Sub-Saharan Africa

A Data Management Plan created using DMPTool

Creator: Ryan Kenny

Affiliation: University of Texas at Austin (UT)

Template: NA

Project abstract:

An Investigation into the Drivers of Chinese and American Aid to Sub-Saharan Africa Research Questions: (1) What country characteristics (government type, freedom, wealth/poverty) drive US and Chinese Aid? (2) How does Chinese aid to Africa differ from United States Aid by sector? (3) Which countries recieve the most/least aid from each?

Last modified: 05-09-2019

Aid Analysis - Chinese and US aid to Sub-Saharan Africa

Data Collection

What data will you collect or create?

Name	Description	Source	Year
Polity_IV	Autocratic/Democratic Nature	Polity Project	2000-2017
Human Freedom Scores	Index of Freedoms by Country	Cato	2008-2016
AidData - Global Chinese Finance	Chinese ODA and OOF by Country	AidData	2000-2014
OECD- US Aid	US ODA by Country	OECD	2005-2017
World Bank	GDP per Capita by Country	World Bank	ALL

How will the data be collected or created?

Structure of folders and files: Datasets will be located in files by version: Raw/Clean/Analysis/Final. The datasets will be named with the following convention: Original for the raw data as it is saved in a RAW DATA folder, not to be manipulated. Clean data used in the project will have the following naming convention: name_RKEDIT_FINALPROJECT. Other versions of data cleaning are kept on desktop folder: CLEAN DATA and have the naming convention: name_RKEDIT_v1, v2 and so on.

Versioning: Cleaning of data is done through excel. As such, instructions for replication are kept in a worddoc and each version will take on a new naming convention as seen above. Word document will be kept in GitHub with Clean Data and updated as needed

Quality Assurance: quality assurance is achieved by ensuring completeness and accuracy by comparing each data version to tools located on the source website (ie. comparing total 2014 chinese ODA to Zambia on AidData's visualization tool to each version of the dataset).

Documentation and Metadata

What documentation and metadata will accompany the data?

Metadata and data dictionaries are created via Markup in the README file in my Github repository, excel spreadsheets with specific variable dictionaries and original documentation of variable definition from each unique source.

Ethics and Legal Compliance

How will you manage any ethical issues?

These datasets are all open source. The nation-state is the unit of analysis so there is no need for anonymization. All data will be fully cited by source, links to data and any extra citation required by the owner organization.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

All copyright and IP/IRR guidance will be followed based on the specific requirements of the data owner.

Storage and Backup

How will the data be stored and backed up during the research?

Data will be stored on GitHub. Backups are kept on a personal external hard drive and updated at the end of every week. There is a copy of each document on my personal desktop. If there is an incident, data will be recovered via external backup.

How will you manage access and security?

My laptop and external hard drive are password protected. GitHub Repository and Python Collab are also password protected. Google Drive is password protected with 2FA

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

As this project uses only open source data, it will only be retained for one year after completion in my GitHub Repository. The only reason for this period to be extended is if I deem it necessary to continue or update research.

What is the long-term preservation plan for the dataset?

There is no long-term preservation plan for this dataset as it will not be kept for more than one year. In the event that it is kept longer for additional reserach, a long-term preservation plan will be generated at time of determination.

Data Sharing

How will you share the data?

Data is shared in an open-access GitHub Repository. All datasets, including all versions and analysis are stored in this repository for open-access.

Are any restrictions on data sharing required?

No.

Responsibilities and Resources

Who will be responsible for data management?

I am solely responsible for the data management of the listed data sets and their usage for this project. This is a solo project.

What resources will you require to deliver your plan?

I will be doing my analyses through Microsoft Excel, plot.ly and Python (Collab or Anaconda). Other resources include workflow software from Google and this dmptool. If additional software is required, I will update as necessary.