Data Archiving for Dissemination within a Gulf Nation

Brian Mandikiana of SESRI

In collaboration with
Lois Timms-Ferrara and Marc Maynard
of Data Independence LLC

IASSIST & CARTO May 2018

Overview

Social and Economic Survey Research Institute (SESRI)

- Established in 2008 at Qatar University

- Has collected over 60 nationally representative household surveys

SESRI Data Archive

- Established in 2017

- Building from the ground up

By Ezz505 (Template: Communtion Section-ER) [CC BY-SA 3.0 (https://creativecommons.org /licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], via Wikimedia Commons

Mission Statement

The SESRI Data Archive Mission is...

To advance globally accepted standards of economic and social survey research data for storage, management, preservation and dissemination.

Source: Courtesy University of Texas Libraries and the US Central Intelligence Agency

Data Sharing Culture in the Gulf

- Availability of microdata from countries in the Gulf region has been limited

- Not only in terms of comparative analysis across countries, but also within individual countries

- Government departments account for the majority of data collected

- Data is mostly used for administrative purposes

- Data sharing an uncommon feature in the region

Challenges: Organizational

- Undefined mission for data archive
- A lack of archiving policies and procedures
- SESRI staff focused mostly on analysis and writing research papers and policy briefs
- Attention to preservation and archiving work not integrated into overall SESRI processes
- Maintaining a type of persistent, reusable version of the resulting data files

Challenges: Legal and Contractual

- Cultural and social sensitivity (but censorship is less than commonly believed)

- Similar to all survey projects, protecting respondent confidentiality and privacy (for instance, releasing geography at an appropriate level of aggregation)

- Contractual arrangements with funders can impose restrictions on preservation, accessibility and re-use

- Developing appropriate review and approval processes

Challenges: Recruitment and Staff

- Hiring to fulfill the archive's human resource needs for both on going work and backlog of past datasets

- The need to hire a lead archivist, two data processors, one developer, and multiple students

- Research data archiving being uncommon in the region, appropriately skilled workers are rare Need for multilingual employees (Arabic and English)

- National focus on hiring local candidates before considering international recruitment

HELP WANTED

Challenges: Technical

- No accumulated metadata

- Survey data and documentation not prepared with broader use in mind

- No standardization in naming conventions and identifiers

- No version control system

- No existing data dissemination channels

An OAIS model**?!** ...I NEED AN *OASIS!*

©Brett Wilson Artistry, 2018.

Approach and Architecture

- Development of the SESRI Archive proceeded on two tracks: general policy development and practical data processing workflow implementation.

- Managing such a process requires substantial planning

- General Policies
  - Developed and discussed a Data Management Policy with SESRI staff.
  - Selection of studies to implement draft policies and strategic planning
  - Data Preservation Policy

Data Processing Workflow

- Data Archive carefully considered and designed a system for storing, and processing new data files.

- Thinking through required steps, identifying dependencies, and documenting their effectiveness, pays dividends in the long-term as the data collection and the Data Archive grows.

Data Processing Workflow

- Data archive advocates for data processing to be conducted using software packages that allows for storing code

- Adding notes directly in the code to document logic behind certain decisions

- Align Workflow with processing procedures

- These steps are essential in promoting data consistency and quality

Data Availability
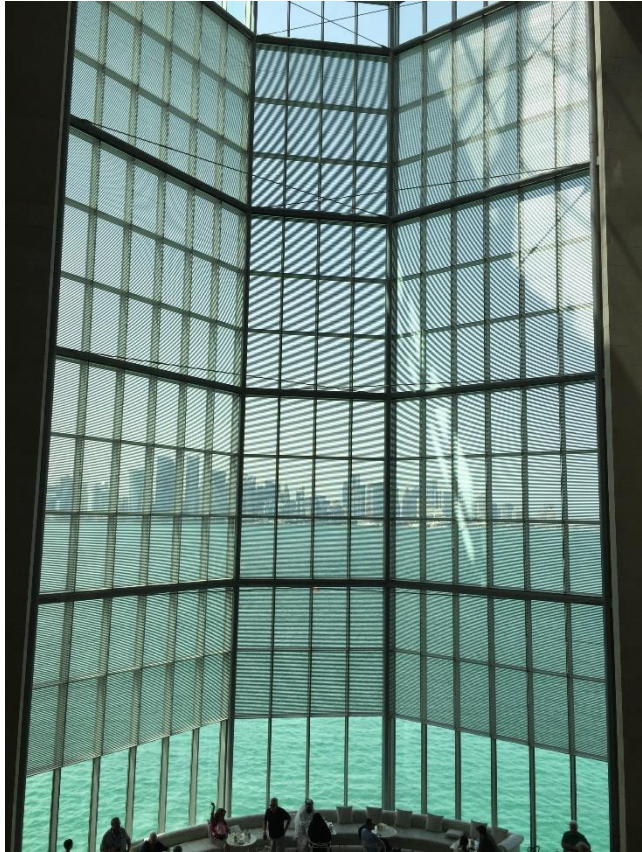
Data Availability

- Data Center Portal for Dissemination

- User and Access Support

- Promotion - marketing the archive, conferences, social media

- Measuring Success – indicators inform action

In Conclusion

- Data archiving supports quantitative research which is essential to better policy making in the Gulf

- Data archiving is beneficial to growth of quantitative research in the Gulf region

- Clear communications with stakeholders are critical

- International relationships are helpful

- Encouraging re-use of survey data would improve data quality, facilitate research, and increase availability of metadata and documentation that supports analysis

Thank you!

Islamic Cultural Center, Doha, Qatar 2017.

SESRI Data Archive:
http://www.qu.edu.qa/research/sesri/

Authors

Brian Mandikiana:
bmandikiana@qu.edu.qa

Lois Timms-Ferrara:
Lois@DataIndependenceNow.com

Marc Maynard:
Marc@DataIndependenceNow.com