DDI-Lifecycle and Colectica at the UCLA Social Science Data Archive

Jeremy Iverson (Colectica)
Elizabeth Stephenson (UCLA SSDA)
Overview

- UCLA Social Science Data Archive Overview
- Search for a Repository Solution
- Implementing DDI-L with Colectica and other systems
- Future Directions
Mission

To provide a foundation for social science research with faculty support throughout an entire research project involving original data collection or the reuse of publicly available studies.
Serve all UCLA quantitative researchers

- Provide reference, cataloging/metadata
- Long term archiving
- Support in data rescue
- Data Management
- Security
UCLA Social Science Data Archive

- One archivist
- One programmer analyst
- Student interns
- Manage survey/quantitative data stored on media from punch cards to cloud
Existing system was becoming difficult to maintain

- Custom-built system, but limited IT resources
- Limited search capability

Search for a solution started 3 years ago
Requirements

- Easier to find for data
- Easier for faculty to document data
- More visibility for data (web-based)
- Support a disaster plan
- Document the entire data lifecycle
- Operate according to TRAC processes and policies

- Off the shelf
  - leverage existing tools
  - work with partners
Options Considered

- Fedora
- Islandora
- Dryad
- California Digital Library – Merrit
- DuraSpace
- Hydra(ngea)
- Nesstar
Issues

- Would have needed a lot of IT resources
- Deployment can be difficult
- Cost was way over what we could afford
- Technology really didn’t fit with what we wanted to do
Solution

- Collaboration: Best practices in curation
- Multi-institutional
- Shared resources
- LOCKSS: Supports disaster plan
- Shared catalog tool

- Collaboration: Archive and user
- Multi-institutional
- Open source
- SafeArchive tool
- Online data access and analysis tool

- Collaboration: Survey researchers
- Visibility for UCLA scholars
- Licensed software
- DDI life-cycle data management tools
- Data and metadata documentation tool
One-size fits all solutions don’t; interoperability does.
Colectica Pluses

- “Colectica offered us a nearly out of the box solution and met our criteria for curation.”

- “For small, domain specific archives Colectica provides a much easier and more technically manageable option than Fedora. “

- “In Colectica designer, a researcher or project manager can pretty easily learn to use the tool during the questionnaire design phase.”
Faculty want to work with colleagues at other institutions

Unless the remote colleagues also have Colectica, they could not use Designer to collaborate

“In the future I would see a use for a web-based tool”
Colectica Components

- Colectica Designer
- Colectica Repository
Workflow for existing datasets

- Import SPSS format files into Colectica Designer
- Add additional details
  - codes
  - question text
  - other curation information.
- Create a file of the DDI metadata and a codebook.
- Load into a Dataverse as part of our participation in DataPASS.
- Link the studies to our Repository so that those searching the repository will be able to access the data.
Operational Schematic without Collaboration Tools
Los Angeles County Social Survey, 1998

Abstract

The Los Angeles County Social Survey (LACSS) continues the Los Angeles Metropolitan Area Studies (LAMAS) and the Southern California Social Surveys (SCSS). A set of core questions focuses on attitudes and views of living in Los Angeles, neighborhoods, public services, and political views. The sample of respondents is a random digit dial sample of all telephone households in Los Angeles County and contains over samples of African American, and Asian American households. The 1998 survey N = 694; White= 43%; African American =11%; Latino =34%; Asian American =6%. Main topics in 1998 focus on ethnic relations. Respondents were randomly selected to answer a series of questions from one of three ballots concerning immigration and affirmative action; social distance questions varied depending on the respondent’s ethnicity. Interviewing used a computer-assisted telephone interviewing (CATI) system. The questionnaire was deployed in English or Spanish depending on language spoken by respondent. This study is provided "as is" and questions should be addressed to the Data Archive.

Citation Information

Title
Los Angeles County Social Survey, 1998

Alternate Title
LACSS

Creator
University of California, Los Angeles, Institute for Social Science Research. (UCLA ISR)

Publisher
University of California, Los Angeles, Survey Research Center. (UCLA SRC), University of California, Los Angeles
Colectica workspace

This workspace is used when entering info for DDI modules.
Example codebook from Colectica (PDF)

Abstract

The Higher Education General Information Survey (HEGIS) series was designed to provide comprehensive information on various aspects of postsecondary education in the United States and its territories (American Samoa, Guam, Puerto Rico, the Virgin Islands, and the Marshall Islands) and Department of Defense schools outside the United States. Data are available for both public and private two-year and four-year institutions. The HEGIS Financial Statistics component for 1968-1969 offers financial data for 2,814 institutions of higher education. Variables include current funds revenues by source, current funds expenditures by function, physical plant assets, indebtedness on physical plant, endowment assets, and changes in fund balances.

Purpose

Sample

Study Citation

Alternate Title: HEGIS
Publisher: Inter-university Consortium for Political and Social Research [distributor].
Copyright: doi:10.3886/ICPSR02094.v1
Survey Design

- How many people were living or staying in this house, apartment, or mobile home?
- Were there any additional people staying here April 1, 2010 that you did not list?
- Is this house, apartment, or mobile home?
- What is your telephone number? We may call if we don’t understand an answer.

For each additional person (PersonCounter):
- Conditional: PersonCounter != 1
- How is this person related to Person 1?
- Does Person (PersonCounter) sometimes live or stay somewhere else?
- What is Person (PersonCounter)’s name?
- What is Person (PersonCounter)’s sex?
Future Directions

- Migrate more data into the system
- Explore Colectica for Excel
- Configure the look at feel of Colectica Portal
- Go Live summer 2013
Thank you

Jeremy Iverson
http://www.colectica.com/

Elizabeth Stephenson
http://dataarchives.ss.ucla.edu/