



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

Jeremy Iverson (Colectica)
Elizabeth Stephenson (UCLA SSSA)

Overview

2

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

UCLA Social Science Data Archive

3

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

4

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

UCLA Social Science Data Archive

5

- One archivist
- One programmer analyst
- Student interns
- Manage survey/quantitative data stored on media from punch cards to cloud

Search for a Repository Solution

6

- 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

7

- 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

8

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

Issues

9

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

Solution

10



- ⊙ 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

Main Lesson

11

One-size fits all solutions don't;
interoperability does.

Colectica Pluses

12

- “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.”

Downside

13

- 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

14

- Colectica Designer
- Colectica Repository

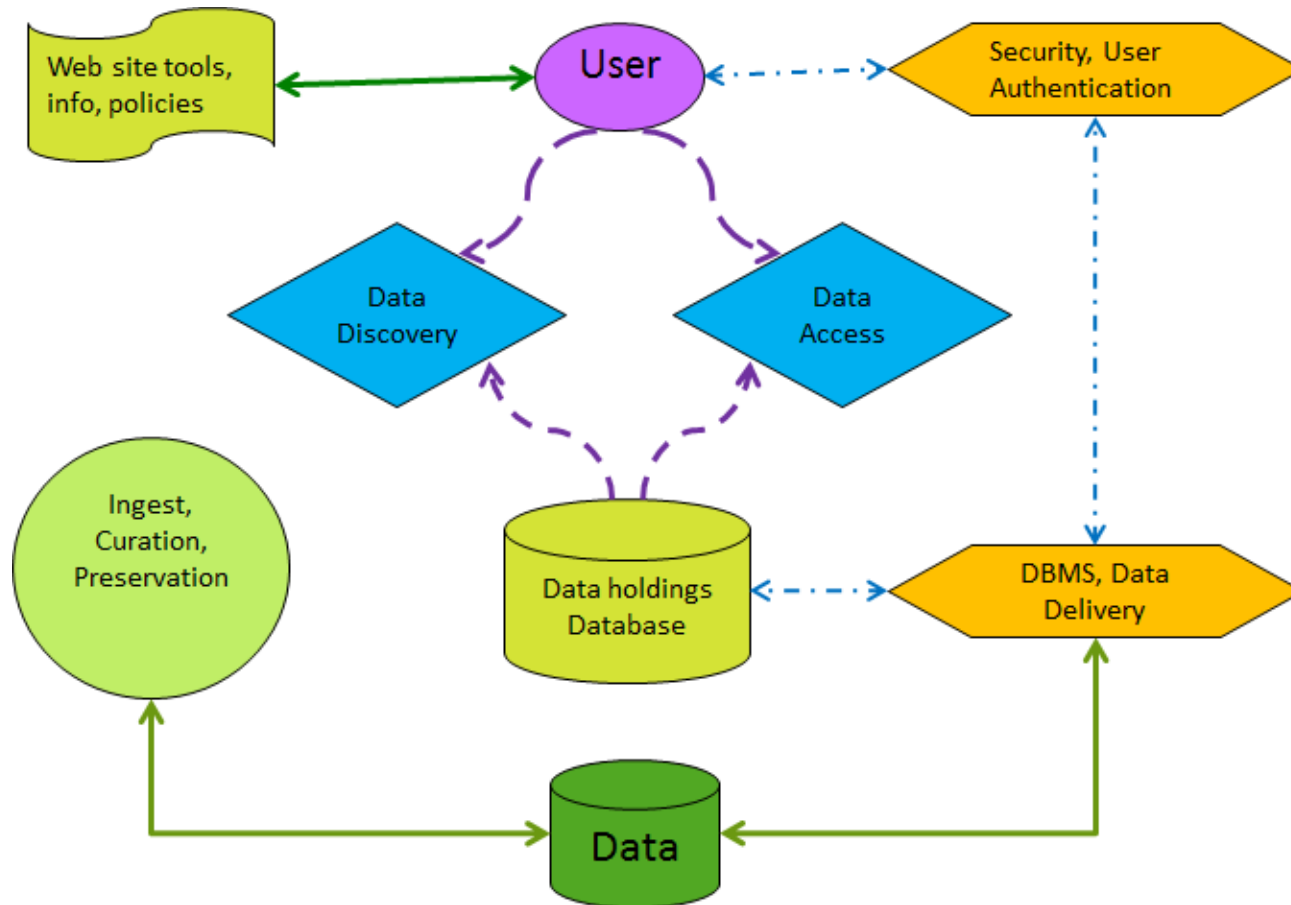
Workflow for existing datasets

15

- 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.

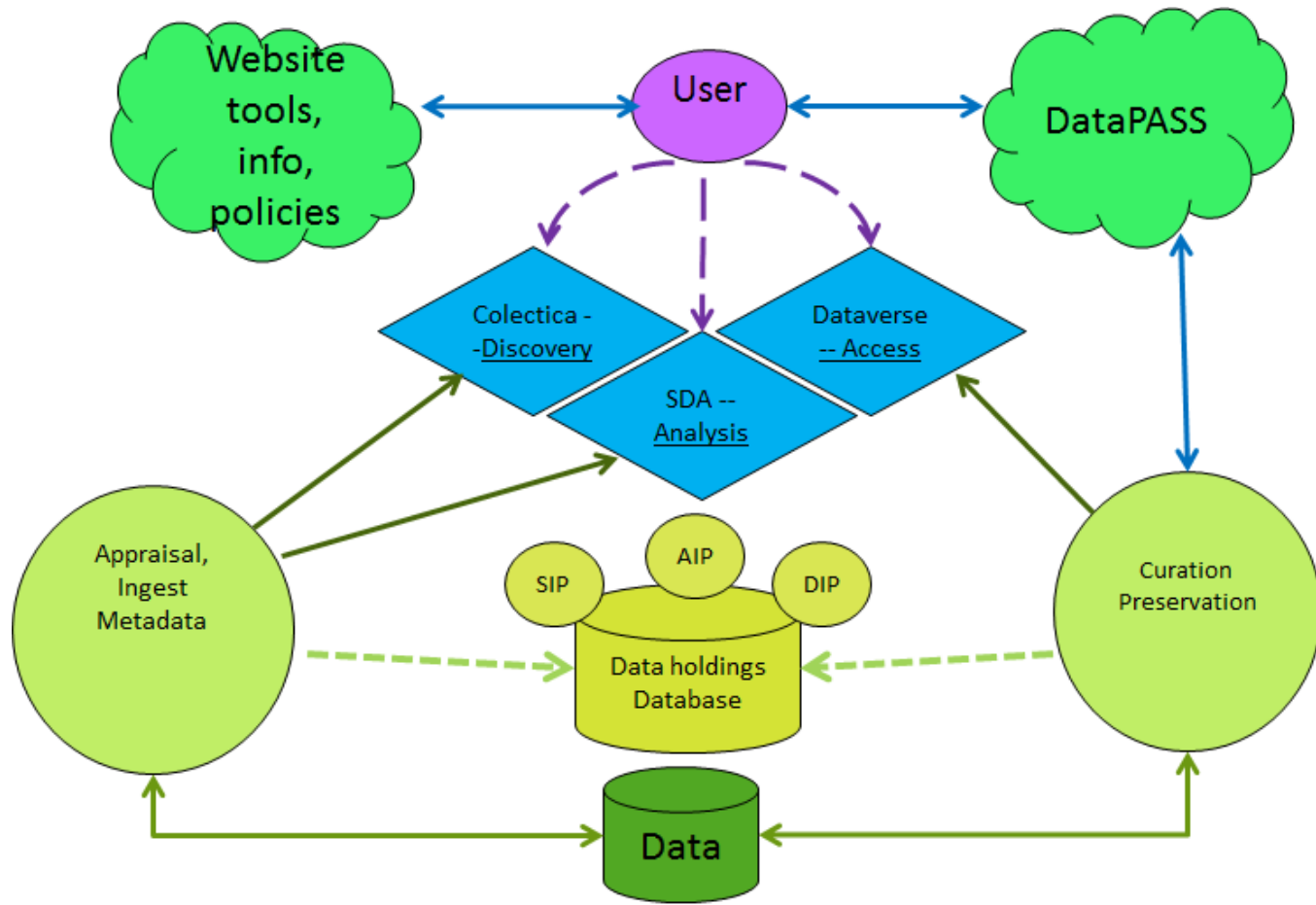
Operational Schematic without Collaboration Tools

16



Operational Schematic Using Collaboration Tools

17



Colectica Repository Public Display

18

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. (Univ

Publisher

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

The information in this public display is created by using the Colectica desktop tool and the uploading to the cloud-based Repository.

Colectica workspace

19

Tool bar options

The screenshot displays the Colectica Designer interface. At the top, a menu bar includes 'Project', 'Import', 'Export', and 'Help'. Below it is a tool bar with various icons for actions like 'Save', 'Languages', 'Create a Checkout', 'Add a Study', 'Add a Series', 'Add a Resource Package', 'Add a Concept Set', 'Add a Universe Set', 'Add an Organization Set', 'Add a Category Set', 'Add a Code Set', 'Add a Data Collection Wave', 'Add a Variable Set', 'Add a Dataset Definition', and 'Modules'. A red box labeled 'Tool bar options' points to this bar.

The main workspace is divided into several panes. On the left is a 'Project' pane with a tree view showing 'Studies', 'Resource Packages', 'Metadata Citation', 'Conceptual', 'Classifications', 'Data Collection', 'Questions', 'Sequences', 'Survey Instruments', and 'Data'. The 'Conceptual' pane is expanded, showing 'Concepts', 'Universes', and 'Organizations'. The 'Classifications' pane is also expanded, showing 'Categories' and 'Codes'. The 'Data Collection' pane is expanded, showing 'Data Collection Waves'. The 'Questions' pane is expanded, showing 'Questions'.

The central pane displays the 'Items' list for the 'Los Angeles County Social Survey, 1997'. The list includes the following questions:

Name	Question Text
R10C	Random number for ordering Q
CON2	I would like to interview someone in your household. In order to determine who to interview, I need to know how many people
CON3	Of the people you just included in your count of household members, what is the first name of the one who will have the next
CON4	What is the first name of the person in your household who had the last birthday?
CON5	In that case, it will take me a moment to select a person to interview based on a random selection table I have. What is the fi
CLIK	Before we start, I would like to assure you that this interview is confidential and completely voluntary. If we should come to a
SEX	Are you male or female?
LA_LAST5	I have a couple questions about how things are in the Los Angeles area. In the last five years do you think that Los Angeles h
ETHGRPLA	Has the wide variety of ethnic groups in Los Angeles helped or hurt the quality of life here?
CONFLICT	Now, I'll ask you some questions about ethnic relations in Los Angeles. Are ethnic groups in conflict in Los Angeles/Southern C

A red box labeled 'This work space is used when entering info for DDI modules' points to the 'CON2' question in the list. Below the list, there are buttons for 'View DDI', 'Create Documentation', and 'Revision History'. The bottom pane shows the 'Question Text' editor for the selected question, with a text area containing 'Case ID'.

Example codebook from Colectica (PDF)

20

Bookmarks

- Cover
- Higher Education General Information Survey (HEGIS), 1968-1969: Financial Statistics [Computer file].
 - Overview
 - Abstract**
 - Purpose
 - Sample
 - Study Citation
 - fin_merge_1969
 - Citation
 - Record: fin_merge_1969
 - FICE - FICE CODE
 - OESTATE - OE STATE CODE
 - SEQUENCE - SEQUENCE NUMBER
 - SYSCODE - SYSTEM CODE
 - SEQCODE - SEQUENCE CODE
 - STATCOD - STATUS CODE
 - OECODE - OLD CODE PRIOR ASSIGN FICE
 - GEOGCODE - GEOGRAPHIC CODE

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

Creator: United States Department of Education. National Center for Education Statistics.

Publisher: Inter-university Consortium for Political and Social Research [distributor].

Copyright: doi:10.3886/ICPSR02094.v1

Survey Design

21

The screenshot displays a survey design tool interface. At the top, there are two tabs: "Design" (selected) and "Basic Information". Below the tabs, there is a toolbar with a green arrow icon and a blue icon with a grid. A tooltip labeled "Toggle Display Style" is visible over the blue icon. The main area contains a list of survey questions, each with a question mark icon and an expand/collapse arrow:

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

Below these questions is a section for conditional logic, labeled "PersonCounter < Q1". It includes a "Conditional" button and a "Branch" button. The condition is set to "PersonCounter != 1". Below this, there are four more questions, each with a question mark icon and an expand/collapse arrow:

- 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

22

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

23

Thank you



Jeremy Iverson

<http://www.colectica.com/>



Elizabeth Stephenson

<http://dataarchives.ss.ucla.edu/>