

Counting California Planning Meetings Notes June 7, 2004 (North) and July 13, 2004 (South)

Attendees

North: UCB: Harrison Dekker, Tim Dennis, Fred Gey, Glenn Gillespie, Ilona Einowski,

UCD: Marcia Meister, Juri Stratford

UCSF: Min-Ling Fang, Peggy Tahir

UCSC: Ken Lyons, Lucia Orlando

CDL: Patricia Cruse, Felicia Poe, Rosalie Lack, Margaret Low

South: UCLA: Cheryl Bartel, Kris Kasianovitz, Jan Goldsmith, Libbie Stephenson

UCI: Kay Collins, Dan Tsang

UCR: Ken Furuta, Lynne Reasoner

UCSD: Elizabeth Cowell, Megan Dreger, Jim Jacobs, James R. Jacobs

CDL: Rosalie Lack, Laine Farley, Harrison Dekker

Note: UCSB was not able to attend due to a scheduling mistake on my part; their input will be gathered by phone

What we learned

- Preservation is very important; we should continue to collect and organize data for preservation
- Few go to CoCA as their first stop, they primarily use FedStats, American FactFinder and Statistical Abstract
- CoCA has value, but...
 - needs more content
 - needs added functionality in order to better serve user needs
 - best serves novice data users (not necessarily novice in their domain area, but novice to the data world) so we need more user support tools (e.g., intuitive navigation, easier search options, glossary)
- We don't need to investigate GIS, yet; GIS on the web is not ready for prime time, but we should provide GIS output for others to take
- Micro and survey data are important to make available online to users
- Marketing--the first step is getting librarians using it
- Campus participation is important, in particular in content development
- Some proposed models for CoCA: CoCA as ready reference tool; as a starting point for research, as a "union catalog" for CA and UC data (OAC model)

I. CONTENT

CoCA needs more content to be valuable

WHAT KIND DO WE NEED?

- Focus on CA but offer other state totals and national totals, where possible
- CA State Agency data; CA EPA CERES
- Provide different levels of geography where available e.g., congressional district, school district. Note: in order to view these levels we will need to provide maps
- Provide county level data; comparing across counties is done often
- Historic data; no one is archiving so we should

- More "novel" content:
 - Hot topics that are of interest to CA such as same sex marriage, gasoline prices; this is real content, plus it could pull users in (great marketing tool); top 10 hot topics
 - Local data: there is a concern that local governments are not keeping their data long-term. Look at Associations of Governments site--Are we duplicating abag, sandag, mbag, skag if we put up local gov data? (Mary Martin is working on a book with locations of state and local agencies)
 - Provide county and city level data that are nowhere else
- UC statistics – enrollment, gender etc. produced by UCOP
- Crime data
- Projection data – e.g., population in 2010
- The most current data are important – even if they are estimates
- Voter registration records
- Faculty datasets; gathering and organizing them is a way to preserve them
- Micro and survey data
 - A growing number of users want to download the actual datasets
 - Microdata and survey data on the web is very new and not a lot has been done and it is much needed
 - Field polls: some are freely available; a good microdata to start with
 - Union Catalog of Microdata: (MelData)
 - Create an index of all data UC has access to (library-owned, at data centers/archives, free, licensed); create a union catalog
- Tract level data – we get a lot of questions about neighborhoods; local community. (ucsc, ucsd, uci, ucla) [Tracts means we need maps. CoCA should only tackle this if we can do something that American Factfinder cannot do. Zip codes access could be a compromise if we cannot do tract]
- NOTE: In the North, many conceded that data at the tract level is not necessary to include since American Factfinder already does it; in the South, many expressed the need for tract/neighborhood level data, but did recognize that we should not re-invent the wheel.
- Obscure data needs. Users can find the other stuff. CoCA has easy to find data. Add obscure data (e.g., citizens vs naturalized citizens; neighborhoods in LA with same sex couples)
- Want data not produced by government agencies (privacy issues)

HOW DO WE DO IT?

- Campus participation is key: Have campus folks look at the current CoCA topics page and identify obvious holes. CoCA team needs to get campus help to catalog what we add to CoCA. CDL could create a toolkit to help folks

II. USER EXPERIENCE

CoCA best serves novice data users (not necessarily novice in their domain area, but novice to the data world)

HOW CAN WE SUPPORT NOVICE USERS?

- At the table level (at point of use) provide a link to table description
- Include a glossary
- Make it google-like; students need it to be easy

- Use pull-down menus to help guide users through the search (e.g., Rand California – they provide pull-down menus; this allows you to see what is there)
- Yahoo-like directories are a good thing
- Intuitive navigation
- Provide geography search on the home page
- Improve search results display
- Like the model of the example searches, provide a link to top ten search terms.
- Create a zero results page
- Create new topic/subject organization; e.g. ethnic, gender
- If you add the functionality that allows users to create their own tables it is important to have documentation so users understand what they just did
- Integrate statistical literacy
- Add link to American Factfinder (if we don't do tract data)
- Add the something like the reference interview---“we don't have this, how about that instead”; E.g. They may want x data for 2000 for y county, but we have x data for 2001 for y county

III. FUNCTIONALITY

CoCA needs added functionality in order to serve user's needs

SEARCH

- Provide geography/place name searching and browsing
- Provide “search within search results”
- Provide “modify search” option for too many results
- Support natural language search, e.g., violence against women
- Implement a thesaurus
- Provide fielded searching via pull-downs

RESULTS

- Currently, the page is too long; investigate ways to have the “greatest hits” rise to the top of the list
- Offer sort options
- Include “dates covered” on the table labels

TABLES

- Support “Extraction”; the ability to extract individual variables and create custom tables; Cut and paste and create new tables on the fly. Of particular interest is time series data e.g., 1970, 1980, 1990 data in one table.
- Tables are currently too wide, hard to view and print

MAPS

- Improve Map feature; allow for mapping more than one vector/variable
- Put county names on the maps
- Provide scaling
- Provide maps and charts but choose meaningful variables and provide percents not raw numbers

OUTPUT

- Graphs
- Raw data
- Html pages
- Provide links: to reports/briefs; to the agency; to cited reports (UC-elinks model)
- We should use the footnotes in CoCA more as an index to help folks to launch from there to other things.

- From the table level, provide links to full datasets and extraction functionality
- More extensible to external software – e.g., GIS software
- Download options to Excel and Ascii

OTHER

- GIS not needed (ESRI has a project/partnership with census to use arcinfo/view)
- CSA (California Statistical Abstract) model in current CoCA is problematic – CSA is not fully integrated, but you can search it.
- Provide the option for the user to create their own “profiles”; provide counties and variables with pick boxes and they create custom tables/reports. A good example is the Profile Warehouse at SANDAG.

IV. PRESERVATION

- It is critical that CoCA play a role in preserving local and state government data and historical data
- Collection, organization and preservation of data should be one of the top objectives; Focus attention on collections rather than on services
- Consider collaboration with regional agencies e.g., ABAG
- Find data that others are not collecting; get and preserve them.

V. MARKETING

- We have got to get librarians to use CoCA first; how? conduct training for librarians
- Get faculty on board -- send brochures to faculty members
- Revive CoCA user group
- ALA award – really promote this. Ask UCOP to include on their site; write a press release
- Do a short presentation at student library orientation
- Map of state showing contributors, to encourage others to donate
- Collaboration is key

VI. CAMPUS PARTICIPATION

- Content development
- Creating “hot topics”
- Cataloging datasets
- UC data archives; get their data into one place
- UCSD could play a role in creating DDIs

VII. CURRENT PRACTICES

HOW DO YOU ANSWER QUESTIONS NOW?

- Usually the first sources are: Statistical Abstract; Historical Abstract, FedStats, American Factfinder (for population queries); go directly to the agency site; Statistical Universe (soon to be canceled)
- CoCA is “not a source for federal data; why would I go to CoCA for data produced by federal agencies when I can go directly to the agency that produces the data?”
- Use ICPSR and OSHPD data (UCSF)
- We use CoCA for ready reference questions from the general public; we don’t use it to help UC patrons
- Encyclopedia of census (e.g., for homelessness issues)

- Users need help with complex, detailed sources
- Statistical Abstracts is a good model, you find what you need or you see where it came from and can go directly there.

WHAT TYPES OF QUESTIONS DO YOU GET?

- Compare variables across counties and across states
- Some want to compare across counties nationally
- Geographic area of concern is CA and area around SC campus (UCSC)
- Get a lot of historical census questions (UCSC)
- Want demographic info for a city (UCD)
- Get CA, nat'l and internat'l questions (UCB)
- Users who need data are not necessarily hard core data users, but they have hard core data questions (UCSF)
- At UCB, they observe that there is a general trend in the increase in folks who can use SPSS and they want the data. At UCSC many folks still can't use SPSS, e.g., social science users.
- Get a lot of questions about neighborhoods; local community. (UCSD, UCI, UCLA)

VIII. AUDIENCES

UC RESEARCHER BEHAVIOR

- They are willing to spend the time, hours and hours to find what they want
- They are domain experts in their field not necessarily in data
- They value preservation
- They are the ones who could promote CoCA for students to use

UC UNDERGRADUATE

- Immediate results; instant gratification
- Just want a number
- Like graphs, maps
- Sexy presentation e.g., Flash, PPT

GENERAL PUBLIC

- They would like "novel" local data
- Immediate results; instant gratification
- They want free content
- Some uses: Business/management/market research (county, city level); Community groups (grants)

IX. COUNTING CALIFORNIA

CoCA's STRENGTH

- CoCA is best for users who need "publications" (vs raw data)
- Multiple ways into the data
- Data integration; bringing data from multiple agencies into one site
- Good for the novice (American FactFinder is hard, you must know the dataset in order to effectively use it)
- Simplicity of CoCA is good, but there are improvements that can be made
- Offering excel as an output format is good for novices
- Providing the citations is very helpful

- Preservation and historical value do not fit in with the mission of the agencies while it is important to the University of CA.

MODELS FOR CoCA

- CoCA as starting point for research; provide links to sources for further research
- CoCA as data dump for CA data (OAC model – provide contributors with the tools to contribute their data)
- CoCA as an index; Statistical Abstract model – you go there to learn where you can go;
- As an index to sources, like SRI
- CoCA as ready reference model vs research tool
- CoCA as “Publications” vs. raw data

COMPARISON TO OTHER SERVICES

- American Factfinder
 - They have some good examples of trying to explain tables, but still takes user away from the table
 - Hard to search
 - Preservation is not their goal
 - CoCA can make finding just CA data easier
- Rand California
 - A lot of data
 - Interface and search is nice; the choices are up front and visible
 - Save as html option is good
- Other state CoCA -like examples? North Carolina State government
- FastStats - <<http://www.cdc.gov/nchs/fastats/Default.htm>>: health stats; provides list of sources, points users to “see also/see more”
- Statistical Abstract; it provides data and provides pointers to agencies.
- Neighborhood Knowledge and Knowledge CA – at UCLA. Census data and county assessor data. Output is maps
- Ciesin < <http://www.ciesin.org/>> The Center for International Earth Science Information Network