DRoP Research Findings Section – Draft

Bowdle, Julie. DRoP Research findings section: draft. Includes Appendix 1: CA Water DRoP user Needs Survey; Appendix II: CA water Data Provider Interview / Questionnaire Revised Draft.

Berkeley, CA: Berkeley Water Center, Water Resources Center Archives, 2010.

We designed two information gathering tools to aid in our evaluation of the water resources data management needs within California and inform our project design decisions going forward. The CA Water DRoP User Needs Survey gathered information on participants' need for CA water data, their experiences with existing methods for finding and accessing data, and their opinions on current unmet needs and recommendations for future services. The CA Water DRoP Data Provider Interview gathered information on the availability, extent, and current format of CA water data across a selection of federal, state, and local government agencies as well as non-governmental organizations and also solicited participants' opinions on current unmet needs and recommendations for future services. Please refer to Appendices I and II for the lists of survey and interview questions, respectively.

Invitations to participate in the CA Water DRoP User Needs Survey were sent out to two email lists: the WRCA Colloquium and the Water Resources Center (a UC multi-campus research unit) email lists.   The WRCA Colloquium and the Water Resources Center email lists, containing approximately 1000 and 5000 emails respectively, consist of UC Faculty, staff, and students along with federal, state, and local agency staff, non-governmental organization staff, water and irrigation district staff, consultants, water attorneys, and laypersons. 241 responses were received by the submission deadline (three weeks from date of invitation). Appendix I contains a statistical summary of responses.

Invitations to participate in the CA Water DRoP Data Provider Interview were sent out to a selection of staff at federal, state, and local agencies and non-governmental organizations that collect CA water data. Agencies that participated include: U.S. Geological Survey, U.S. Environmental Protection Agency, CA Department of Water Resources, CA State Water Resources Control Board, East Bay Municipal Utility District, Santa Cruz County, American Rivers, and San Francisco Estuary Institute.

The CA Water DRoP User Needs Survey

Most participants categorized their current occupation or interest in CA water data as that of researcher / scientist, engineer, consultant, professor / educator, manager / decision maker, or some combination of the above (multiple selections were allowed) affiliated with universities, engineering firms / consultants, or federal, state, or local governments.

Top 5 Occupations / Interests	Percentage
Researcher / scientist	43.6%
Engineer	23.7%
Consultant	22.0%
Professor / educator	21.6%
Manager / decision maker	14.5%

Top 5 Employers / Affiliations	Percentage
University	26.7%
State government	22.5%
Consultant	16.3%
Local government	8.3%
Federal government	7.1%

When categorizing their need for CA water data, the majority of participants (75.5%) specified research / analysis as one of their needs. Less than 50% of participants selected each of the other reasons, which included environmental interests (49.4%), decision-making, policy (41.5%), decision-making, engineering (41.1%), regulation (33.6%), decision-making, operations (33.2%), education support (33.2%), and litigation support (15.4%).

Of the top five occupations, consultants, professors / educators, and policy makers have a greater need for data collected by other organizations, on average, with the average percentage of outside data used by participants close to or greater than 50% for all occupations / interests and affiliations (see Appendix I, Tables 4.1 and 4.2). Data from state agencies appear to be in greatest demand, followed closely by data from federal and then local agencies, with a higher percentage of participants needing non-digitized data from local agencies than from the other agency / organization types.

Data from:	Characterization of Needed Data			
	Digital	Digital, older format	Not digitized	N/A
State agencies	82.3%	34.1%	20.4%	5.8%
Federal agencies	80.5%	28.3%	19.5%	9.3%
Local agencies	68.6%	27.4%	26.5%	14.2%
Universities	65.9%	23.0%	14.6%	24.3%
NGOs	55.3%	20.4%	14.6%	33.6%

Participants spend an average of 10 hours per month searching for data (median: 5 hours per month) with close to 50% finding it somewhat difficult to find data outside of their organization and another 10% finding it very difficult. Although, it was noted by several participants that the ease of access depends upon the organization and the type of data sought. (Of note, 26% of participants find it somewhat difficult to find data within their own organization and another 4% very difficult. Of those, the majority are affiliated with universities (30.8%) or state government (26.2%). Nearly all participants who said it is very difficult to find data within their own organization are affiliated with state government (75%).) Based upon their experience searching for data, nearly 60% of participants said that the outside data that they need is only sometimes available online and another 7% said that is it rarely available online.

Google (or other search engine) searches and specific agency or program web site searches are the most often used methods when searching for CA water data with 35% of participants using Google (or another search engine) daily. Colleague inquiries, agency / program inquiries, and databases of water data are the next most utilized while indices / catalogs of water data and libraries / archives are the least utilized tools or methods of the choices provided (see Appendix I, Table 16.1). The success rate of Google (or other search engine) searches is high or very high for less than 40% of participants (less than 45% for agency or program web site searches) and low or very low for 12-14% of participants. These two methods along with colleague inquiries have the highest success rate while indices / catalogs of water data and libraries / archives have the lowest (see Appendix I, Table 16.2).

Less than 1% of participants said that the currently available methods for finding and accessing CA water data are always sufficient to meet their needs, 46% of participants said that they are sufficient most of the time, while 48% said they are only sometimes sufficient. Some of the issues raised regarding the effectiveness of existing tools and methods include not always knowing what data is available or where to look for it, the number and lack of comprehensiveness of existing archives or "fragmentation of data archiving within and between organizations," inconsistency in data management and data formats within and between organizations, inadequate or undocumented metadata, limited accessibility to historic data, difficulty in using data that is posted in non-parsable formats (e.g. pdf files), and the lack of a data sharing mandate.

Most participants said it was extremely important that data is findable and accessible online and available in usable digital form. Having well documented metadata was also listed as important or extremely important for most participants. Other factors listed as important to participants when using data include the accuracy and reliability or QA/QC of the data, the data's output / download format (with participants wanting common data formats compatible with standard software packages without requiring major format changes), and the continuity / completeness of the data. Documented site location information and collection methodologies, subsets of metadata, were also listed by participants.

Factors when using data	Extremely Important	Important	Somewhat Important	Not Important	Total Responses
are findable and accessible online	75.4%	19.7%	3.9%	0.9%	228
are in usable digital form	66.7%	26.2%	6.2%	0.9%	225
have well documented metadata	45.0%	31.7%	19.7%	3.7%	218

Specific agencies whose data participants would like greater access to include CA Department of Water Resources, U.S. Geological Survey, water and irrigation districts, Regional Water Quality Control Boards, U.S. Environmental Protection Agency, CA Department of Public Health, and State Water Resources Control Board, along with a number of additional federal, state, and local agencies (see Appendix I, Table 7 for the complete list).

Participants search most frequently for water quality, streamflow, watersheds / groundwater basin, discharge, and water system operations data and would benefit most from improved access to watersheds / groundwater basin, water quality, and discharge data (see Appendix I, Table 8 for complete results). The majority of participants also rated their need for improved access to data that are available online as extremely important and similarly for improved access to data collected in the past 10 years. In addition, more than 75% of participants rated their need for improved access to digital data not currently available online, data that have undergone QA/QC, data available in reports, and data collected prior to 2000 as either extremely important or important (see Appendix I, Tables 9 and 10).

The service that would provide the greatest benefit to and generated the greatest interest among participants is making online data more findable / accessible via index and/or catalog. Making offline digital data more findable / accessible via index and/or catalog came in second and connecting users to hard-to-find data by request came in third (see Appendix I, Table 21 and below). Additional services of

interest suggested by participants included a "digital well inventory for subsurface geology," a "centralized catalog of storm-drain data," "developing a network group [for] expert exchange and collaboration," guidelines for metadata development, and standardizing similar data types across agencies. Particular data that participants listed as a priority for cataloging and/or archiving included water system operations, groundwater levels and quality, and well logs, and groundwater pumping / use.

Services	Extremely interested	Interested	Somewhat interested	Not interested	Total Responses
Making online data more findable / accessible via index and/or catalog	59.4%	34.6%	5.1%	0.9%	217
Making offline digital data more findable / accessible via index and/or catalog	42.6%	38.8%	16.7%	1.9%	209
Connecting users to hard-to-find data by request	33.0%	44.3%	19.3%	3.3%	212
Archiving / hosting data in danger of being lost (through staff retirement, outdated formats, etc.)	34.5 %	30.1%	25.2%	10.2%	206
Providing agencies with guidance for digitizing and/or posting data	21.4%	38.8%	17.4%	22.4%	201
Archiving / hosting data at the request of a data generator	16.1%	33.7%	34.2%	16.1%	199

The search feature that generated the greatest interest among participants is the ability to search by geographic location / region (this is assuming the ability to search for records by keyword, category, title, author, agency, program, project, and data range is already in place). A map interface to records came in second and the ability to search by watershed a close third (see Appendix I, Table 24 and below). Additional features of interest suggested by participants included the ability to search by public agency or water management boundary, legislative district, and groundwater basin.

Search Features	Extremely interested	Interested	Somewhat interested	Not interested	Total Responses
Search by geographic location / region	70.5%	23.0%	5.1%	1.4%	217
Map interface to records	59.7%	26.5%	10.9%	2.8%	211
Search by watershed	59.6%	26.8%	7.5%	6.1%	213
Version history	21.5%	44.4%	26.3%	7.8%	205
Data use history	16.7%	40.5%	33.8%	9.0%	210

Community features generated the least amount of interest among participants with less than 20% being extremely interested in any of the proposed community features. Of those proposed, the ability for users to submit suggestions for additional data sets to acquire generated the most interest and the ability of users to comment on data came in second (see Appendix I, Table 26 and below). Additionally, a few participants raised concerns with a community or wiki approach to data acquisition and management citing the need for QA/QC and stating that such approaches rarely have adequate QA/QC.

Community Features	Extremely interested	Interested	Somewhat interested	Not interested	Total Responses
Users can submit suggestions for additional data sets to acquire	18.0%	44.9%	25.4%	11.7%	205
Users can comment on data	12.1%	41.3%	30.6%	16.0%	206
Users can post modified versions of data sets	7.3%	28.2%	34.0%	30.6%	206
User moderated discussion forums	7.7%	25.0%	39.3%	28.1%	196
Users determine priority of acquisitions	5.0%	29.4%	40.8%	24.9%	201

When asked what additional suggestions participants have for the CA Water DRoP, responses included the coordination and sharing of information with other data management efforts, the need for greater involvement by local governments and municipalities, centralizing data or data linkages in order to streamline searches, and a reiteration of the importance of QA/QC, access to complete and well documented metadata, the ability to search by location, and digitizing data. One participant (affiliated with a municipal water district) raised the difficulty this would impose on smaller agencies: "Some water districts are so small that this is a monumental task. What can be done to assist them becoming a part of the digital age?"

The CA Water DRoP Data Provider Interview