

Appendix II: CA Water Data Provider Interview / Questionnaire Revised Draft

1. Background Information

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.

2. Please indicate whether your answers to the following questions will be about the data collected by your entire agency or organization, your division, or your office. You may select all that apply.

<i>Agency</i>	<i>Division / Office / Program</i>
U.S. Environmental Protection Agency	Region 9, Water Division, Drinking Water Office Region 9, Standards and TMDL Office
U.S. Geological Survey	Agency-wide
CA Department of Water Resources	California Cooperative Snow Surveys Division of Statewide Integrated Water Management [groundwater, surface water, and water quality datasets and well construction information] Hydrology and Flood Operations Office
CA State Water Resources Control Board	Office of Information Management and Analysis [CIWQS, GeoTracker GAMA, and SWAMP datasets]
East Bay Municipal Utility District	Water Operations Division, Water Supply Office
Santa Cruz County	Water Resources Division
American Rivers	California Region
San Francisco Estuary Institute	Regional Monitoring Program
Multi-Institutional <sup>1</sup>	Yosemite hydroclimatology study

<sup>1</sup> USGS, Scripps Institute of Oceanography, California Cooperative Snow Surveys, the University of Washington, and others.

Additional Notes

- DWR [groundwater dataset]

*DWR has three types of groundwater data:*

- *Wells measured periodically (2 - 6) times per year*
  - *Wells measured by DWR*
  - *Wells measured by cooperators*
- *Wells monitored continuously (all by DWR, and only since 2002)*

- DWR [well construction information]- *Well Completion Reports is the name of the form drillers are required to submit to DWR.*

3. How would you categorize the data that your agency, division, or office collects? Please check all that

apply.

<i>Category</i>	<i>Agency / Division / Office / Program</i>
Watersheds / groundwater basins	EPA [Standards and TMDL Office], USGS, DWR [HFOO], SWRCB [GeoTracker GAMA], SWRCB [SWAMP], Santa Cruz County
Streamflow	USGS, DWR [surface water dataset], DWR [HFOO], SWRCB [SWAMP], EBMUD, Santa Cruz County, American Rivers
Discharges (wastewater, agricultural, industrial, stormwater)	EPA [Standards and TMDL Office], SWRCB [CIWQS], SWRCB [GeoTracker GAMA]
Groundwater elevation	USGS, DWR [groundwater dataset], SWRCB [GeoTracker GAMA], EBMUD, Santa Cruz County, American Rivers
Receiving water quality	EPA [Standards and TMDL Office], USGS, DWR [water quality dataset], SWRCB [CIWQS], SWRCB [SWAMP], Santa Cruz County
Water quality (wastewater, agricultural, industrial, stormwater)	EPA [Standards and TMDL Office], USGS, SWRCB [GeoTracker GAMA], SWRCB [SWAMP], Santa Cruz County, American Rivers, SFEI [RMP]
Drinking water quality	EPA [Drinking Water Office], SWRCB [GeoTracker GAMA]
Aquatic ecology	USGS, EPA [Standards and TMDL Office], SWRCB [SWAMP], Santa Cruz County, American Rivers, SFEI [RMP]
Meteorological	USGS, DWR [Snow Surveys], DWR [HFOO], SWRCB [SWAMP], EBMUD, Multi-Institutional [study]
Oceanographic	EPA [Standards and TMDL Office]
Water system operations	USGS, DWR [HFOO], EBMUD
Other	EPA [Drinking Water Office] – violations and enforcement actions, water system inventory information; DWR [Snow Surveys] – hydrologic; DWR [water quality dataset] – water quality in general, not just receiving waters; DWR [well construction information] – well construction information; DWR [HFOO] – snow; SWRCB [GeoTracker GAMA] – groundwater cleanup sites, landfill and other waste discharges to land, regulatory monitoring data, and well data; American Rivers – assessments, surveying; SFEI [RMP] – sediment quality; Multi-Institutional [study] – hydrologic

4. Do you (i.e., your agency, division, or office) provide access to your raw data or are datasets only publicly available after they have undergone QA/QC? Please explain.

Raw data only

- DWR [well construction information] – *DWR provides access to the forms it receives from drillers and does not perform a QA/QC check on this data.*
- DWR [HFOO]
- SWRCB [CIWQS] – *SWRCB does not perform a QA/QC check on the data provided.*
- EBMUD - Available via both CDEC and <http://www.ebmud.com>

Both raw and QA/QC

- USGS – Both are publicly available. For example, approximately 85% of streamflow data appears in real time, then reappears 3-9 months later after it's been through QA/QC (e.g. a daily average flow for streamflow rather than 15 minute measurements). The raw (real time) data gets overwritten in the process but is available by request and is also available in the Instantaneous Data Archive.
- DWR [Snow Surveys]
- SWRCB [GeoTracker GAMA] - Both the raw and QA/QC data are available on GeoTracker and GeoTracker GAMA.

QA/QC data only

- EPA [Standards and TMDL Office]
- DWR [groundwater dataset]
- DWR [surface water dataset] – Some provisional data may be available during the year and is marked with a QA flag. Final / QA'd data is available after the end of the water year. Historical data was supposedly QA'd, but there is no documentation on how it was done.
- DWR [water quality dataset]
- SWRCB [SWAMP] – While the public does not have access to the raw data, the regional boards and state board representatives that manage the programs do.
- Santa Cruz County – Once their data is generated according to their lab procedures, they input it into the database. They will be releasing a website that will display some of this data and that will update daily. Their data is also made available upon request.
- American Rivers – Data goes through QA/QC before being submitted to SWAMP. Although raw data would be available upon request. Watershed groups that American Rivers funds upload the data to their own web sites.
- SFEI [RMP]
- Multi-Institutional [study] – available upon request

Other

- EPA [Drinking Water Office] – They do not receive the raw drinking water quality data but only the violations of the federal regulations reported to them by the states plus the 98<sup>th</sup> percentile lead and copper levels for water systems.

5. How much of your data are currently available online?

	All	Most	Some	None
--	-----	------	------	------

Data collected in the past year	EPA [DWO] DWR [ground] DWR [surface]	USGS [Raw: 85%] EBMUD	EPA [TMDL] DWR [Snow Surveys] <sup>1</sup> DWR [water quality] <sup>2</sup> SWRCB [GG] Santa Cruz	USGS [QA/QC] DRW [well info] SWRCB [SWAMP] SWRCB [CIWQS] SFEI [RMP] Multi [study]
Data collected since 2000	EPA [DWO] USGS [QA/QC] DWR [ground] DWR [surface] SFEI [RMP] <sup>3</sup>	EPA [TMDL] USGS [Raw: 70-80%] SWRCB [SWAMP] EBMUD	DWR [Snow Surveys] <sup>1</sup> DWR [water quality] SWRCB [GG] Multi [study] <sup>4</sup>	DRW [well info] SWRCB [CIWQS] Santa Cruz
Data collected prior to 2000	EPA [DWO] <sup>5</sup> USGS [QA/QC] DWR [ground] SWRCB [SWAMP] <sup>6</sup> SFEI [RMP]	EPA [TMDL] DWR [surface] SWRCB [GG] EBMUD	USGS [Raw: 10-15%] <sup>7</sup> DWR [Snow Surveys] <sup>1</sup> DWR [water quality]	DRW [well info] SWRCB [CIWQS] Santa Cruz

Other

- American Rivers – Data are submitted to SWAMP at the end of a project, not on an annual basis. Unknown prior to 2000.
- DWR [HFOO] - Most (some archived data are in the process of being digitized and placed online).
- Multi-Institutional [study] – Study began in 2001.

<sup>1</sup> Approximately half of this data is currently available online (much of the online data are averages). The goal to put everything up, including the raw data (the raw data records are handwritten).

<sup>2</sup> There are two sets of water quality data: pre-1992 and post-1997 (there is a gap in between with no data). Pre-1992 data need to be QA'd and matched with current methods and nomenclature. Post-1997 data are complete with QA. The issue with post-1997 data is that it needs to be made available to a wider audience.

<sup>3</sup> All that they intend to post is online (annual status and trends monitoring data get posted, special studies reports get posted while the data do not).

<sup>4</sup> Data are also available upon request.

<sup>5</sup> All data for the past 20 years are available on EPA's website. More data are available in the SDWIS database. (There are data as far back as 1979.)

<sup>6</sup> This is not a lot because the SWAMP program began in 2001, although some regions have given them some of their historic data. BDAT has a lot of this data right now and the rest would be available on the CEDEN network.

<sup>7</sup> It's very difficult to find and access much of the raw data prior to 2000 as the systems that were used to collect and manipulate the raw data prior to then (or a bit earlier) were different enough from the systems that they use now, so there's a technological gulf. Although the hardcopy still exists, it's increasingly difficult to get your hands on.

6. Of the data that are not currently available online, how much are in accessible digital formats (i.e.,

recent file formats or, if older formats, the necessary hardware / software to access the data are still available)?

	All	Most	Some	None	N/A
Data collected in the past year	DWR [wq] Santa Cruz American Rivers SFEI [RMP] Multi [study]	SWRCB [SWAMP] EBMUD	EPA [TMDL] DWR [Snow Surveys] <sup>1</sup> DRW [well info] <sup>2</sup> SWRCB [CIWQS]	SWRCB [GG]	EPA [DWO] USGS[QA/QC] DWR [ground] DWR [surface]
Data collected since 2000	DWR [wq] Santa Cruz American Rivers <sup>3</sup> SFEI [RMP]	EPA [TMDL] SWRCB [SWAMP] EBMUD	DWR [Snow Surveys] <sup>1</sup>	SWRCB [GG] SWRCB [CIWQS]	EPA [DWO] USGS[QA/QC] DWR [ground] DWR [surface] DRW [well info]
Data collected prior to 2000	DWR [wq] Santa Cruz SFEI [RMP]	EPA [TMDL] EBMUD	DWR [Snow Surveys] <sup>1</sup> DWR [surface] DRW [well info]	SWRCB [GG] SWRCB [CIWQS]	USGS[QA/QC] DWR [ground] SWRCB [SWAMP]

Other

- DWR [HFOO] - *The data that is not digitized is in written form on paper in file cabinets.*
- USGS [raw data] - *It's difficult to access the raw data prior to the mid-90s. Most of this older raw data is either in digital form but in formats that they aren't able to access anymore (e.g., tapes) or exists as hardcopy.*

<sup>1</sup> *Approximately half.*

<sup>2</sup> *Most of the time, complete information is not submitted. Data is collected by four regional offices. Different offices have some data available in digital format (in a data base), but not the entire form.*

<sup>3</sup> *All as far back as 2001, unknown prior to that.*

7. How much of your digital data are available in parsable formats (e.g., Excel spreadsheets rather than embedded in .pdf files)?

	All	Most	Some	None	N/A
Data collected in the past year	EPA [DWO] USGS [QA/raw]  DWR [Snow Surveys] SWRCB [GG] American Rivers <sup>1</sup> SFEI [RMP] <sup>2</sup> Multi [study]	EPA [TMDL] DRW [well info] SWRCB [SWAMP] EBMUD Santa Cruz	SWRCB [CIWQS]	DWR [wq]	DWR [ground] DWR [surface]
Data collected since 2000	EPA [DWO] USGS [QA/raw] DWR [Snow Surveys] SWRCB [GG] American Rivers SFEI [RMP] Multi [study]	EPA [TMDL] DRW [well info] SWRCB [SWAMP] EBMUD Santa Cruz		DWR [wq] SWRCB [CIWQS]	DWR [ground] DWR [surface]
Data	EPA [DWO] <sup>3</sup>	EPA [TMDL]	DWR [surface]	DWR [wq]	DWR [ground]

collected prior to 2000	USGS [QA/raw] DWR [Snow Surveys] SWRCB [GG] SFEI [RMP]	DRW [well info] EBMUD Santa Cruz		SWRCB [CIWQS]	SWRCB [SWAMP] Multi [study]
-------------------------	---	--	--	---------------	--------------------------------

Other

- DWR [HFOO] – *All online digital data are in parsable formats.*

<sup>1</sup> *Data is kept in Excel but is probably submitted to SWAMP as .pdf files, so might need to request the data directly from American Rivers to receive it in a parsable format.*

<sup>2</sup> *The exception would be any non in-house special studies for which they did not receive the data in a parsable digital format.*

<sup>3</sup> *All digital data back to 1990 are in parsable formats.*

8. Do you have plans to make your offline digital data available online? If so, to what extent and in what time frame? If no, why not? (E.g., lack of funding, lack of staff resources, not relevant, etc.) Enter N/A if not applicable.

Yes

- DWR [water quality dataset] – *Yes, this process is very slow. Data prior to 1992 needs to be QA'd. They do not have the resources to do this quickly. Making post-1997 data available requires an education process of program managers. Some readily make this data available, others need to agree to make this data available to a wider audience.*
- DWR [HFOO]
- SWRCB [SWAMP] – *Yes, in the next two years.*
- Santa Cruz County – *They are working towards this now and hope to have it up in the next 3 years. They are short funding and technical expertise to migrate their database from MS Access to MySQL or SQL and to become more SWAMP compliant.*
- American Rivers – *Yes, data is submitted to SWAMP at the end of each project.*
- Multi-Institutional [study] – *Yes, in the next one to two years.*

Interested, but lack the resources

- DWR [well construction information] – *They would be very interested in making this information available on the web, but do not have the resources (money and staff). A web application would help manage the data set and update the well completion report form. In addition, the application would need to be able to locate a well through both geo-coding and GIS location on a map; relate it to the groundwater level data (there is no rigorous correspondence now); work for drillers that are not particularly tech savvy, and work for the permitting agencies in all 58 counties (different counties permit under different departments).*

No

- EPA [Standards and TMDL Office] – *No, only for very specific and small data set regulatory projects (e.g., for TMDL or Total Maximum Daily Load projects).*
- DWR [surface water dataset] – *The data not currently made available is either of questionable quality or was published in various DWR and USGS reports. There are no plans to make it available due to lack of funding. Even with funding, this would be a fairly low priority.*
- SFEI [RMP] – *No, although though may revisit this in the future for particular special studies.*

Not applicable

- DWR [Snow Surveys]

- DWR [groundwater dataset]
- SWRCB [GeoTracker GAMA]
- EBMUD

9. If you do not intend to make your offline digital data available online due to resource constraints, would you embark upon this effort if resources became available? Please explain. Enter N/A if not applicable.

- DWR [surface water dataset] – *The Statewide Data Program - Surface Water - would make the data available if/as they re-QA the data. This would take approximately six full-time PYs one year who were knowledgeable of their system and procedures.*
- DWR [water quality dataset] – *Yes. For data prior to 1992, this would be akin to matching samples for N up with total N, N-soluble, N-NH3, or some other form of N; checking for duplicate records; and trying to assign a detection limit for the analysis.*
- DWR [well construction information] – *Yes. This would be a multi-year, large budget project. Coordination with drillers and counties will be a significant part of the project.*
- SWRCB [CIWQS] – *Their rollout of the online system is hindered in part by lack of resources, but this isn't the only problem.*
- Santa Cruz County – *They could do a lot more with additional funding for database and GIS technical support.*

10. Do you have plans to migrate your data in older digital formats to current formats? If so, to what extent and in what time frame? If no, why not? Enter N/A if not applicable.

Yes

- DWR [surface water dataset] – *This is being done very slowly on a site by site basis. The data includes monthly and annual minimum and maximum flows for flood projects.*
- DWR [well construction information] – *They are slowly scanning old paper documents. They only digitize six fields from each form, not the complete form. Three of their four offices are good about this, while the fourth is way behind on scanning documents.*
- SWRCB [SWAMP] – *Yes, if they receive anything they do the migration themselves.*
- Santa Cruz County – *Yes, they are working on this now.*

No

- USGS – *The older raw digital data is in formats that they aren't able to access anymore (e.g., tapes).*

Other

- DWR [water quality dataset] – *They do not have a time frame to do this.*

Not applicable

- EPA [Standards and TMDL Office]
- DWR [Snow Surveys]
- DWR [groundwater dataset]
- DWR [HFOO]
- SWRCB [CIWQS]
- SWRCB [GeoTracker GAMA]
- EBMUD
- American Rivers
- SFEI [RMP]
- Multi-Institutional [study]

11. If you do not intend to migrate your data to current formats due to resource constraints, would you embark upon this effort if resources became available? Please explain. Enter N/A if not applicable.

Yes

- DWR [water quality dataset] – *Yes. For data prior to 1992, this would be akin to matching samples for N up with total N, N-soluble, N-NH3, or some other form of N; checking for duplicate records; and trying to assign a detection limit for the analysis.*
- DWR [well construction information] – *Yes. For the office that is behind on scanning documents, it would require someone to travel to that office and organize the reports to ensure there are no duplicate copies, all pages in the report are present and in order, etc.*
- Santa Cruz County – *Their work is limited by resources. They could do a lot more with additional funding for database and GIS technical support.*

Perhaps

- DWR [surface water dataset] – *Yes. Although, if given the choice, they would spend the money on higher priorities. These include documenting their procedures and developing metadata for what they have.*

12. Do you have plans to digitize your non-digital data and make it available online? If so, to what extent and in what time frame? If no, why not? Enter N/A if not applicable.

Yes

- DWR [Snow Surveys] – *Yes, as they can get to it. They don't have any resources dedicated to this undertaking.*
- DWR [well construction info] – *They will make this data available online when they develop a web application; and when they have the resources.*
- DWR [HFOO] – *Yes, pending time and funding.*

No

- USGS – *Unaware of any such plans (for the older raw data prior to the mid-90s), although it could come up in the future for particular projects. If it occurs, it's likely to be local or need driven rather than across the board.*
- DWR [surface water dataset] – *No due to lack of resources.*
- SWRCB [CIWQS] – *No, they do not plan to digitize the monitoring reports which contain the data that have been submitted by dischargers over the past 30 years in non-digital formats. There are tens of thousands of reports stored in different locations within different offices across the state. They do not see a worthwhile benefit to this effort.*
- EBMUD – *No due to lack of resources.*

N/A

- EPA [Standards and TMDL Office]
- DWR [groundwater dataset]
- DWR [water quality dataset]
- SWRCB [GeoTracker GAMA]
- SWRCB [SWAMP]
- Santa Cruz County
- SFEI [RMP]
- Multi-Institutional [study]

13. If you do not intend to digitize your non-digital data due to resource constraints, would you embark upon this effort if resources became available? Please explain. Enter N/A if not applicable.

Yes

- DWR [Snow Surveys]
- DWR [well construction instruction] – *Yes. This would require staff for coordination as much as it would require funding.*

- EBMUD

Perhaps

- DWR [surface water dataset]

No

- USGS – *If so, this effort would likely be local or need driven.*
- SWRCB [CIWQS] – *This would not be a worthwhile effort. It would still require staff time and resources even if resources were available to do the main portion of the effort.*

14. Do you have data that you feel are particularly vulnerable to loss (due to staff retirement, older file formats, decay of media, etc.)? If so, please describe.

Yes

- EPA [Drinking Water Office] – *There's a possibility of not receiving data from smaller district or county offices (for example, run by one individual) in case of retirement or lay off until employee(s) can be replaced.*
- DWR [Snow Surveys] – *Much of the data is hard copy and there's just the one copy of it.*
- DWR [surface water dataset] – *People have retired who knew what they did and the process wasn't documented, so much of the knowledge has already been lost.*
- DWR [well construction information] – *DWR is legally required to maintain these documents but could lose much of this information in a disaster.*

No

- EPA [Standards and TMDL Office]
- USGS
- DWR [groundwater dataset] – *All of the data is stored in a database and also available on the web.*
- DWR [water quality dataset] – *Post-1997 data is secure. (Pre-1992 data had the staff loss and so for that data they have what they have.)*
- DWR [HFOO]
- SWRCB [CIWQS]
- SWRCB [GeoTracker GAMA] – *This is addressed by their Software as a Service contractor.*
- SWRCB [SWAMP] – *Between SWAMP and CEDEN the data seems secure for the long term.*
- EBMUD
- Santa Cruz County – *While a lot of the anecdotal knowledge will leave, they have good track of the data.*

- American Rivers – *Electronic files are backed up daily.*
- SFEI [RMP]
- Multi-Institutional [study]

15. Do you have a long term data management plan in place that includes preservation of current data and migration to future formats when necessary? If so, please describe it briefly.

Yes

- EPA [Drinking Water Office] – *Quarterly snapshots of all the data (with data going back to the 70s) migrated into EPA’s Drinking Water Data Warehouse.*
- USGS
- DWR [groundwater dataset] – *Enterprise processes exist for change management and technology upgrades.*
- DWR [surface water dataset] – *This system will be maintained and converted as necessary in the future so that data is not lost.*
- DWR [water quality dataset] – *Current data has daily backups, and enterprise processes for change management.*
- DWR [HFOO] – *Program plans include funding and time to keep the system current.*
- SWRCB [GeoTracker GAMA] – *This is done by their Software as a Service contractor.*
- SWRCB [SWAMP]
- EBMUD – *They plan to continue maintaining current data they are storing electronically.*
- American Rivers – *Includes preservation of current data. Migration to future formats on an as-needed basis.*
- SFEI [RMP]
- Multi-Institutional [study]

No

- EPA [Standards and TMDL Office]
- DWR [Snow Surveys] – *Nothing that formal.*
- DWR [well construction information] – *Other than scanning, no.*
- SWRCB [CIWQS]

Other

- Santa Cruz County – *They are working to develop a program that will house water quality, water resources, habitat, and management data and relate that information in a GIS environment. They will use the system to detect resource trends and assess management effectiveness.*

16. What is your understanding of your data’s audience?

<i>Agency / Division / Office / Program</i>	<i>Audience / Understanding</i>
EPA [Drinking Water Office]	Congress (reports sent to congress once a year), the public, water systems, manufacturing companies, companies that produce water treatment equipment, groups like the environmental working group and national resource defense council, companies with sales interests like bottled water companies or those selling water home treatment units, etc.
EPA [Standards and TMDL Office]	States, watershed managers, dischargers, the public, etc.
USGS	Broad range including people doing resource management, energy production, public health, recreation, etc.
DWR [Snow Surveys]	A broad audience including water resource managers, climatology, virtually any natural resource investigation, plants, animals, etc.
DWR [groundwater dataset]	Good understanding. Specific consultants, academic institutions, etc.
DWR [surface water dataset]	Pretty good understanding. Specific consultants, academic institutions, etc.
DWR [water quality dataset]	Good understanding but not great. Others in DWR would have a better understanding of this dataset’s audience.
DWR [well construction information]	Good understanding. Audience ranges from consultants to the SWRCB.
DWR [HFOO]	A wide audience of data users from water system operators to the general public
SWRCB [CIWQS]	Primary audience is state regulatory staff (permit writers and compliance staff)
SCRCB [GeoTracker GAMA]	Public, legislators, state agency staff, technical consultants, scientists, schools, academia
SWRCB [SWAMP]	A broad range, from scientists to managers to the public – the environmental community
EBMUD	Audience is mostly internal, some external water agencies
Santa Cruz County	Fairly good. Multiple audiences, the general public and agency folks
American Rivers	Funders, decision makers, internal, the general public, water master and diverters
SFEI [RMP]	Pretty broad, including regulatory agencies, academia, consultants, their permit holders, etc.
Multi-Institutional [study]	Researchers

17. Would your audience benefit from your data (or a subset of your data) being catalogued in a searchable database of CA water data? If yes, approximately how much data (i.e. number of data sets to catalog) are there? If not currently available online, where does the data reside (e.g., central location on site / offsite, distributed devices on site / offsite, some combination of the former)? Do metadata exist for these data?

Yes

- EPA [Drinking Water Office] – *Yes, provided the data is available from either EPA or the state.*
- USGS – *“It’s always useful to make things more readily discovered.”*
- DWR [Snow Surveys] – *Yes, the snowpack dataset. Portions of this dataset are available online. The data that’s not currently available online resides in one central location. Metadata exists for both the online and offline data.*
- DWR [well construction information] – *There are approximately 750K well completion reports (varying quality, some multiple pages) residing in four offices. These reports are confidential under CA law and not available to the public without well owner consent.*
- DWR [water quality dataset] – *This dataset is located on Water Data Library (unsure of size).*
- Santa Cruz County – *They are currently working on this.*
- SWRCB [SWAMP] – *Yes, if our effort comes up with another data integration system (already in CEDEN, but CEDEN, in general, wants to share and exchange data with other partners). Data includes chemistry results, toxicity test results, discrete field measures, laboratory results, tissue chemistry results, benthic invertebrate results, and observational data. Note: There are a couple of cataloging efforts coming out of the Central Valley office focused on the San Joaquin River and Delta water quality information that they’re interested in possibly expanding state-wide.*
- Multi-Institutional [study] – *Yes, not knowing all the researchers who might be interested in it.*

No

- EPA [Standards and TMDL Office] – *Most of the data relevant to California is already in a state database (State Water Resources Control Board, DWR).*
- DWR [groundwater dataset] – *This data is being cataloged by Google and Microsoft (assumedly BING).*
- DWR [surface water dataset] – *This data is being cataloged by Google and Microsoft assumedly BING).*
- DWR [HFOO] – *Part of this process is in progress with the Western Regional Climate Center.*
- SWRCB [GeoTracker GAMA] – *The data are already provided in this manner.*

- EBMUD – *Their audience is provided data directly via e-mail reports or accessing their website.*

Unsure

- SWRCB [CIWQS] – *Unsure if there would be an added benefit.*
- American Rivers
- SFEI [RPM]

18. Would your audience benefit from your data (or a subset of your data) being archived and hosted in a central repository of CA water data? If yes, approximately how much data (i.e., number / size of data sets) are there? What format(s) are they currently in? If not currently available online, where does the data reside? Do metadata exist for these data?

Yes

- DWR [Snow Surveys] – *Yes, the snowpack dataset. Portions of this dataset are available online. The data that's not currently available online resides in one central location. Metadata exists for both the online and offline data.*
- DWR [well construction information] – *There are approximately 750K well completion reports (varying quality, some multiple pages) residing in four offices. These reports are confidential under CA law and not available to the public without well owner consent.*
- DWR [water quality dataset] – *This dataset is located on Water Data Library (unsure of size).*
- Santa Cruz County – *If it were combined with other datasets, then yes. They are hoping to accomplish the standalone effort with their own data.*

No

- EPA [Drinking Water Office] – *Only if our effort were to offer some level of integration of datasets or unique tools not currently available on EPA's website would this be potentially of interest.*
- EPA [Standards and TMDL Office] – *This data is already archived on CEDEN.*
- USGS
- DWR [groundwater dataset] – *Probably not. (The dataset includes approximately 10,000 - 11,000 wells per year since 1956, with more than 42,000 wells total. Wells are measured 2 - 6 times per year.)*
- DWR [surface water dataset]
- DWR [HFOO] – *Part of this process is in progress with the Western Regional Climate Center.*
- SWRCB [CIWQS]
- SWRCB [GeoTracker GAMA] – *This is done by their Software as a Service contractor.*
- SWRCB [SWAMP] – *No, seems duplicative (data already in CEDEN, BDAT).*
- EBMUD – *Their audience is provided data directly via e-mail reports or accessing their website.*
- SFEI [RMP] – *They're already in the process of linking into CEDEN.*
- Multi-Institutional [study] – *They're trying to accomplish this by working with another archive and will be hosting the dataset there.*

Unsure

- American Rivers – *Unsure if beneficial in addition to being available in SWAMP.*

19. What types of data would you have concerns about making available to the public?

Confidential information was listed by most participants as the type of data they would have concerns about making available to the public. Specifically, participants listed issues under the California Public Records Act (homeland security, draft, etc.), confidential information under California law (well completion reports and private domestic well owner information), GPS-level location data for wells and places with supply or infrastructure implications, other data generated on private lands that could potentially create problems for land owners, certain data related to ongoing enforcement cases, and proprietary data. Raw data was also listed by a couple of participants (expressing either that it should not be made available to the public or concerns with it being made available without analysis and interpretation) and screening level data intended for planning but not regulatory purposes was listed by another participant.

20. What metadata standard(s) do you currently use to describe your data, if any? Please check all that apply.

<i>Metadata Standard</i>	<i>Agency / Division / Office / Program</i>
EML (Ecological Metadata Language)	DWR [groundwater dataset], DWR [well construction information]
EPA standards for water quality	DWR [water quality dataset]
FGDC (Federal Geographic Data Committee)	EPA [Standards and TMDL Office]
SWAMP comparable format	SFEI [RMP]
Internal	EPA [Drinking Water Office] <sup>1</sup> , DWR [Snow Surveys], SWRCB [SWAMP] <sup>2</sup> , Multi-Institutional [study]
Unknown	EPA [Standards and TMDL Office], USGS, DWR [HFOO], SCRCB [GeoTracker GAMA], EBMUD, Santa Cruz County
None <sup>3</sup>	DWR [groundwater dataset], DWR [surface water dataset], DWR [well construction information], SWRCB [CIWQS], American Rivers <sup>4</sup>

<sup>1</sup> XML tags used to identify data elements passed from SDWIS/State to SDWIS/ODS and eventually to the data warehouse, not enough graphic data to claim adherence to the FGDC standard

<sup>2</sup> although considering using the CERES metadata system which is FGDC compliant

<sup>3</sup> for questionnaire responses “None” may mean they’re using an internal format as “Internal” wasn’t listed as an option

<sup>4</sup> no methodology for metadata, although submit metadata to SWAMP as well

21. What form(s) are your metadata currently in? Please check all that apply.

<i>Dataset(s)</i>	<i>Digital</i>	<i>Print</i>	<i>Personal Memories</i>	<i>N/A</i>
EPA [Drinking Water Office]	X			

EPA [Standards and TMDL Office]	X	X		
USGS	X	X	X	
DWR [Snow Surveys] <sup>1</sup>	X			
DWR [groundwater]		X	X	
DWR [surface water]			X	
DWR [water quality]	X			
DWR [well construction info]		X	X	
DWR [HFOO]	X	X	X	
SWRCB [CIWQS]				X
SWRCB [GeoTracker GAMA]	X			
SWRCB [SWAMP]	X			
EBMUD				X
Santa Cruz County		X	X	
American Rivers	X			
SFEI [RMP]	X			
Multi-Institutional [study]	X	X	X	

<sup>1</sup> *Digital metadata for both online and offline hardcopy data*

22. As a data provider, what water resources data management services would provide the greatest benefit to your agency, division, or office? E.g., archiving / hosting data in danger of being lost (through staff retirement, outdated formats, etc.), providing agencies with guidance for digitizing and / or posting data, etc.

The services that participants said would provide the greatest benefit to their agencies or organizations included guidance and support, statewide standards (i.e., standardized methods and formats) to promote compatibility and comparability across datasets, a site listing all thresholds, a metadata catalog of water data, and an exchange system for water data.

In terms of guidance, responses included guidance on research design (i.e., when, where, and how often to collect data and when to stop collecting data), analyzing data (e.g., the types of statistics that can be used and how to use them), generating metadata (advice on what the best options are), and digitizing hardcopy data with regard to retrieval capabilities. In terms of support, responses included “technical support in database development and management,” technical support with GIS development, technical assistance with training data providers to collect data in SWAMP comparable format, and assistance with digitizing historical hardcopy data.

With regard to developing an exchange system for water data, participants suggested “facilitating data transfer between agencies,” “build[ing] a common language dictionary across organizations,” “the ability to translate or merge datasets” from different agencies (whether creating a site that is able to obtain and translate data from different sources and present this data in a common language or hosting data in a particular location where datasets must adhere to a specified format) so that users can easily compute across datasets without needing to retool data with each change in source, integration of different types of datasets (e.g., groundwater levels with water quality and well construction information) to promote better planning, and hosting data where users can do larger retrievals than are currently possible in CDEC.

One caution raised was that the responsibility for archiving all California water data should not fall on any one agency.

23. What institutional concerns should we be aware of when sharing data?

Institutional concerns that we should be aware of when sharing data include confidential data and information (see response to question 19), using or referencing the main data source, having and highlighting the appropriate metadata (including specifying whether the data is provisional or validated and the purpose for which the data was collected), keeping multiple copies of datasets in synch if hosting data (e.g., propagating changes when agencies find and correct errors in a dataset), having comparable data, and “the lack of data exchange standards.” One participant said it depends upon the data source.

24. What additional suggestions do you have for the CA Water DRoP in this water resources data management effort?

Suggestions for the CA Water DRoP ranged from the general to the specific.

General suggestions included not attempting to be a “one-stop shop for CA water data,” discovering and providing assistance when key datasets are being lost or not archived, finding discrete areas to focus on where assistance is needed rather than trying to address everything California-wide, coordinating this effort with other data management efforts (and one participant specifically suggested coordinating with the Western Regional Climate Center), settling on data standards and assisting data providers with meeting those standards, translating data of the same type (e.g. streamflow) into a common format if pulling into a composite database, and creating a catalog of existing CA water data that specifies where users can locate and access the data.

Specific suggestions included creating a catalog of groundwater data (identifying where groundwater data is being collected and by whom), designing a system that supports the interchangeability of streamflow data (whether hosting the data in a comprehensive database where datasets must adhere to a specified format or translating the data into a common format on the fly), designing a system that supports the interchangeability of meteorological data, and harmonizing all data within a watershed (by working through how to address different formats from various sources as well as different spatial and temporal scales).

Participants elaborated upon and provided a rationale for their suggestions. A catalog of groundwater data was suggested due to how greatly dispersed groundwater data is (across cities, counties, water districts, etc.) along with a lack of knowledge of who is collecting this data and where it is being collected. It was stated that even a blurred level of location information would be helpful. While the eventual goal would be to design a system that supports the interchangeability of groundwater data, the first step would be to create this metadata resource or catalog of groundwater data. A system supporting the interchangeability of streamflow data was suggested because locating the data isn't the major problem in this case (with approximately 70% of this data either being collected for or handed off to and vetted by USGS as well as a lot of communication between the data collectors) but instead being able to compute across datasets without needing to retool data with each change in source. A system supporting the interchangeability of meteorological data was suggested because there are several state-scale, co-existing networks rather than one large network where users can discover and access this data all in one place. It was stated that the Western Regional Climate Center has been working on this but that there is still a lot of work to be done. Harmonizing all data within a watershed (beginning with a FERQ, priority, or representative watershed and adding additional watersheds from there) was suggested for the purposes of improved analysis.

One additional suggestion was that if we decide to create a composite database, federal datasets and particularly surface water information from USGS and COOP precipitation observer records would be useful additions since neither are easy to search or download.

25. Please feel free to share any additional comments you may have here.

Additional comments from participants pertained to data unification and the funding and responsibility of data collection and maintenance. With regard to data collection and maintenance, one participant raised the issue of determining who is responsible for collecting the data over the long term, who is responsible for paying for this data collection and analysis, and finally who is responsible for maintaining the data. With regard to data unification, another participant suggested looking into CalEPA's National Information Exchange Network as "the EPA has a series of XML schemas for different data flows, different business rules for doing the exchanges, authentications, etc.," to which CEDEN will be subscribing.

Additional information was also provided on DWR's well completion reports: State well numbers (used for GW level monitoring) are rarely assigned on well completion reports. Locating wells is a problem. Often they are only given the address on the form. The address is not one of the fields DWR digitizes. So the field has to be read from the TIFF file, and the location plotted on map. Correspondence with groundwater level data was also stated as being an issue.