California Water Quality Monitoring Council

Data Management Workgroup Notes

February 1, 2017 — 1:00pm to 4:00pm

Department of Water Resources, Conference Room 121
3500 Industrial Blvd
West Sacramento, CA 95691

IMPORTANT INFORMATION!

Conference Web Conferencing information is provided on the last page of this agenda.

For security purposes, all visitors are required to sign in at the entry of the building.

For information and resources concerning this meeting, we welcome you to visit the Data Management Workgroup collaboration site: **http://sfei.li/dmwg**

ITEM: #	1	Approx.Time:
Title of Topic:	**Introductions, announcements, agenda review, and minutes approval**	5 mins
Purpose:	Decision: Share information and review materials.	
Attachment Link:	Meeting Notes:	
Contact Person:	Gary Darling	gary.darling@water.ca.gov
	Tony Hale	tonyh@sfei.org, (510) 746-7381

ITEM: #	2	Approx.Time:
Title of Topic:	**Launching a new Open Data Platform**	70 mins
Purpose:	Coordination and Information: David Harris is leading the development and launch of a new open data platform in the California Natural Resources Agency. He will be demonstrating its proposed uses, architecture, and decisions regarding its design.	
	The DMWG will be offered the opportunity to learn about the decision-making process that led to the adoption of this new technology infrastructure, which might influence future decisions for the CNRA and DWR.	

Background:	The new CNRA Data Platform will serve as a CNRA is doing a digital data library for the natural resources organizations. It will be the one stop for CNRA data, and will federate with others, including Data.ca.gov. At its core, the system is largely based in open source CKAN software, operating in the Amazon Web Services cloud. It also leverages many of the advanced features that AWS offers.
	The open-source technology ecosystem in which the Open Data Platform was built is very well supported by a robust software development community. Instead of being locked into a vendor, the tool's stakeholders are able to drive technology decisions. The platform is focused on natural resources entities and hosting/serving their data types (time series, spatial data etc.).
	The platform, in many ways, aligns with the recent goals of the DMWG and the Environmental Data Summit. The system supports uploads, data curation, data discovery with a hub for ETL (extract, transform, and load) operations.
	A formal launch is expected in December or January.
Contact Person:	David Harris David.harris@water.ca.gov
Notes:	David Harris presentation on launching a new open data platform

- Open data is available to the public, it's accessible, it's reusable, and it's freely available. Open data is data that can be published, shared, and be freely used, re-used, and redistributed.

- Tony-how we define objectives for the services we are providing, think about our audience, where does data science fit in? David – the critical stakeholder group is program driven, support decision makers, designed to be flexible and adaptable and powerful to organizations.

- Tony – Is there any thinking or established communication pathways to pierce silos? David – work on supporting the programs first. If it's not improving the outcome, there is no reason to pierce the silos.

- USEFUL, USEABLE, and USED – the approach and goals. One size will not fit all. Where can we afford to consolidate and standardize in a centralized way? Ultimately, it's a business intelligence platform and using it will find the procedures and policies needed.

- Tony – does ADP fully encapsulate the ETL process or is the ETL process spread out? David - there are multiple ETL solutions, and we can add more. The current tool is cattle, it's open source and popular.

- Tony – does the ODP manage the front end, and ADP manage the back end? David – it's hard to diagram the front and back end because ETL can be in the front (pulling data) or back. it's more of a web than a straight-line flow.

- Unifying set of components, there's 3+1 fundamental functions of service 1. Need to be able to find stuff 2. Be able to make data useable and have it interoperate (ETL like function) 3. Visualize it data to create answers 4. Balance (find coherence)

- Tony – if another agency wants to collaborate, would it be governance?

	David - it's a natural resource agency platform and it's built to serve those objectives. Any partner organization can participate or be brought in if it will be useful.
	• Tony - Governance used to establish common control vocabularies or translations among vocabularies, or set up guardrails on what's possible, so it would be good to know how strong of a voice partner agencies may have to determine vocabularies. David – shared objectives and standards on how work is done, all that can be supported by the platform and meta-data extensions.
	• Tony - need to tackle multiple formats, multiple masters and servants, if there is a lack of agreement, have 2 different data sets versions, which one is authoritative in decision making, looking at legality, which dataset is paramount, which standard rules them all, there is an authority question to be tacked at some point. David - Different standards can be interoperable, some not so much. It's not going to assure everyone does the same, but it does remove all the technical obstacles.
	• Supports organizational branding, the departments will get pages that have just their data, individual departmental portals are built into the system, important to preserve the departments authority, demand will be able to be meet, it can receive data sets from partners
	• Tony – use case of storm water BMP from 2 labs that may have different formats D- the template can be created with drop zone folders and automate harvest of data, or they could write python to convert those as well
	• Solar is great for text search, there is also meta-data searching
	• Very open platform, can use many different tools
	• Completing final development activities, implementing proof of concept projects, working on data federation, need to streamline on-boarding process, should be launched soon
	• DOI's will make federation easier, and keep us out of data loops, and 2. Digital preservation and capacity management (data sets in multiple sites)
	• Tony – DOIs will happen at some point, it's better to try to set up the system now.
	• RP – QC'd vs. non-QC'd data, will that create dual data set? David – it's already resolved in that you can label the data set, and reference the QC'd vs. not QC'd data. DOI's will help.
	• RP – QAPP's associated with data, is there a way to link to QAPP? David - there are a ton of fields in the metadata, it could be an optional field which has a quality standard. Tony - it's an actual document and wouldn't fit into a field. Maybe the solution would be to use a customizable tab as a link to the QAPP document.
	• RP – library of documents? Lawrence Livermore Laboratory will be working on preserving data. There is a need to preserve our data and reports.
	• Tony – feedback mechanism? David – we will be collecting feedback.

	There is a feedback link at the bottom of every page. Change control process is established.	

ITEM: #	3		Approx.Time:
Title of Topic:	**AB 1755 Updates and Developments**		45 mins
Purpose:	Information: In a year's time, the implementation of AB 1755 has been proceeding swiftly. The DMWG will be reviewing progress to understand how best we can continue to assist in the process.		
Background:	In September 2016, the legislation AB 1755 passed into law. It mandates that DWR, in close consultation with the Monitoring Council among other entities, implement a data-sharing platform to promote sound decisions about water-related data – both water supply and water quality – throughout California. Since its September passage, the Act has received a great deal of attention and support from public, private, non-profit, and academic data providers and users. Soon to be released for public review are the "protocols" and "strategic plan" drafts, which we will discuss. We will also provide updates on some developments that have taken place recently regarding the formulation of an Open Water Information Architecture document and several other related publications. Questions remain, but swift progress is the mandate of the legislation and the goal of the agencies.		
Desired Outcome:	Everyone should share the same understanding of the current state of the efforts in various quarters.		
Attachment Link:	Assembly Bill No. 1755: https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB1755		
Contact Person:	Gary Darling	gary.darling@water.ca.gov	
Notes:	• Open to everyone with platforms, protocols more general case, more procedures less specifics, how can we federate systems • Policy vs technical people • Used use-cases and who, what, what, what – who needs, what data, in what format, for what decisions, (and where are the data gaps) • OWIA – Open Water Information Architecture – interoperable data, interoperable meta-data, open source tools, set standards and conventions, SOP's in QA/QC. Interoperability is fundamental, and the data needs QA/QC. Formal versioning to track data. • Tony – 2 different workgroups, once governance is finished, what		

	happens? Gary – not sure yet
	• QC – reproducibility is a big deal, check-point data would be OWIA data set, anomaly detection and correction section, curation protocols, intellectual property rights policy, i.p. rights, multiple access points, will there be a master catalog, open source platforms, open source code repository so anyone can use, decision support tools
	• https://eesa.lbl.gov/open-transparent-water-data-ca-workshop-documents access points to draft documents
	• 20 use cases, 20 queries, has groundwater, recharge, water rights, ecological cases as benchmark, there is an analysis of the use cases,
	• Rachel Pisor to be included in discussions with the 2 groups on QA. Rachel to help with interagency QA/QC work, including sharing how difficult it is starting new program.
	• Tony – key challenge will be how do you come up with translations of different measures into common tiers, 0 is no QA, but what is level 1 and 2, will need to establish equivalency
	• Tony - Dissuasions that cut across topics with closely related topics like documentation, versions, and DOIs. Gary – how to handle this by committees? Tony – relatively small specialized committees, feeding up to larger groups, and know where the connections are located
	• Tony – are David and Gary thinking about merging or federating their efforts? David is producing the initial ODP from a DWR perspective. The partners are not going for that. The gov-ops and CNRA platforms will confederate eventually. APIs currently don't work.
	• OWIA incompatible with CCAN in any way? Not that can be seen.
	• Local water agency needs? It's a sigma issue, but what about AB issue – we are working with those who have approached us.
	• The next big discussion with the sigma set, there is a big water conference in January, how do you do remote sensing?

ITEM: #	4	Approx.Time:
Title of Topic:	**Flash Round! 5-minute updates on topics related to the workgroup**	30 mins
Purpose:	Informational: What is happening in the community? Examples might include progress updates on portal development efforts, potential connections to related projects, upcoming presentations and conferences, and new innovations we should assess.	
Desired Outcome:	Information and planning: This item allows for anyone to contribute news about developments that haven't been covered earlier in the meeting. We are looking for news related to water quality, ecosystem health, and data management. Please confine your update to 5 minutes.	

Contact Person:	Tony Hale	tonyh@sfei.org, (510) 746-7381
Notes:	• Tony - Southern California coastal water research project and sfei and collaborating on a common library of trash monitoring standards. Stricter monitoring compliance, need for programs to monitor trash much more consistently, and have a state-wide vision, currently there is a hodgepodge of methods, working to address the gap in standards which will map to morning questions which will then map to management questions	
	• Homeland security is concerned with drone maker DGI is using big data to determine those that might fly drones over critical infrastructure and could take your drone video and send it back to a server in China, which is amassing maps of critical infrastructure and takes your phone contacts.	
	• DOIs and data management plans should be items for the future	
	• 34 North – presented estuary portal to monitoring council, working on data dashboards, working with NASA for remote sensing data, Oregon State is generating turbidity maps,	

ITEM: #	5		Approx.Time:
Title of Topic:	**Assignments from the Steering Committee**		45 mins
Purpose:	Information: The Steering Committee tasked the DMWG with a number of deliverables, and the workgroup has made progress. We would like to review the documents and approaches toward completion.		

Background:	The Steering Committee last met on August 11, 2017 with representation from several divisions of DWR, SWRCB, Delta Stewardship Council (Delta Science Program), OPR, and DFW. Representatives shared ideas about future directions for their respective interests in data sharing and the work related to AB 1755 implementation.
	
	Since then, the DMWG has made progress on a number of assigned tasks that are complementary to the Open and Transparent Water Data Act. For this meeting, we will review some of the latest developments on the following:
	
	• Outline on guidance for an Open Data Handbook
	
	    • Initial effort at the departmental level
	    • Merge with data management plan guidance
	    • Customize for Prop 1 grant agreements
	
	• Developing a template for Data Management Plans
	
	    • Approach to protocols per AB 1755		

Desired Outcome:	After this meeting, we should have an improved focus on how we might		

	productively implement the guidance of the Steering Committee and what work lies ahead in developing materials.	
Contact Person:	Tony Hale	tonyh@sfei.org, (510) 746-7381
Notes:	- Protocols guidance document – be as specific as possible, how do we apply them across the board, it is better to mention the item i.e. version control system, than to go into too much detail or omit it all together. Describe the (be specific in) functions not the technical specifications, the objectives. If it isn't more specific, how is it actionable or what is it. Look for sweet spot between technical and non-technical audience - Protocols OWIA informing DWR, 2 looping paths, documents produced with partners, do we incorporate that product? On protocols just go for platform needs, or do we go beyond that to the OIWA? Tony - The 2nd most importing things is the interconnection among the different protocols and how they are interdependent. Ex version control and DOIs. The most important thing is the emphasis on the data sharing protocols and focusing on machine to machine communication.	

ITEM: #	**6**		**Approx.Time:**
Title of Topic:	**Next meeting and items to agendize**		5 minutes
Purpose:	Identify topics and responsibilities for inclusion in our next workgroup meeting.		
Background:	Previously the DMWG selected the 1st Wednesday of even months for its meeting time.		
Desired Outcome:	Agendize topics for next meeting as appropriate.		
Contact Person:	Tony Hale	tonyh@sfei.org, (510) 746-7381	
	Gary Darling	gary.darling@water.ca.gov , (916) 651-9207	
Notes:	- Put 34 North on the agenda for next meeting - Next meeting will be Thursday February 1st, venue to be determined, from 1-4 p.m. - Review steering committee document - Review open data handbook guidance document		