Agenda

ITMD Project Annual Workshop

April 13, 2022

Confederated Tribes of the Warm Springs Indian Reservation of Oregon

8:30		Welcome ITMD: A Short History and Thoughts for the Future	Doug Hatch, CRITFC Fishery Science Dept Chair
8:40		Budget Brief	Sheryn Olson, ITMD Coordinator (CRITFC)
8:50		Columbia Basin Fish & Wildlife Library: Resources for Researchers	Tami Wilkerson, Library Program Lead/Librarian, CBFWL
9:00		Our biggest obstacles? Our possible solutions?	▪ YNF – Michelle, Bill, Anneliese, *et al.* ▪ CTUIR – Colette Coiner, Stacy Schumacher, *et al.* ▪ CTWSRO – Stefan Kelly, *et al.* ▪ NPT – Clark, Tyler, Sam, Ryan, *et al.*
9:45		**Data Management Maturity Model**	How do we score a similar dataset? – Spawning Ground Survey: Clark Watry discussion leader
10:00		**Assessing the DMMM tool:** **Sheryn Summarizing**	▪ YNF – Bill Bosch, Data Manager ▪ CTUIR – Colette Coiner, Data Management Coordinator ▪ CTWSRO – Stefan Kelly, Watershed Restoration Coordinator ▪ NPT – Clark Watry, Project Leader/Data Steward ▪ CRITFC – Joe Nowinski
10:15		**Break**	
Managing Large Datasets			
10:30		From Micro to Macro: Scaling-up a data management ecosystem	Benjamin Warren, Jacob Shapley and Danny Warren, Washington Department of Fish and Wildlife Biological Database Systems
11:00		Data management and visualization for large datasets at CMOP	Charles Seaton, Coordinator, CMOP | Center for Coastal Margin Observation & Prediction (stccmop.org)
11:20		Managing data from Unmanned Aerial Systems (UASs)	Connar Stone, IT and Drone Program Manager, Grande Ronde Model Watershed
12:00		**Lunch**	

Inter-Tribal Monitoring Data Project
Annual Meeting April 13, 2022

Data Management Maturity Model

How do we each score a similar dataset?
Use Case – Spawning Ground Surveys

	Combined Dataset Evaluation										
Data Mgmt. Stage	**Dataset Planning**		**Collect/Create Data**		**Dataset Quality**	**Data Analysis/Reporting**		**Data Archiving**		**Management for future accessibility, interoperability**	**1 - Performed** **2 - Managed** **3 - Defined** **4 - Measured** **5 - Optimized**
	Dataset Categories										
Evaluation Categories	**1** Data Mgmt. Plan	**2** Dataset Workflow	**3** Dataset Data Collection Methods/Protocol (Standards)	**4** Dataset Data Collection (Technology)	**5** Dataset QA/QC Processes	**6** Dataset Reporting/ Publishing	**7** Dataset Access	**8** Dataset Metadata	**9** Dataset Storage	**10** Dataset Life Cycle Management	Dataset Maturity (average level)
SGS datasets (average level)											

Removed one driver (category) from datasets: Dataset Governance

Clark Watry, NPT presenting, et al.

DMMM Dataset Criteria: QA/QC and Metadata Categories

	Ad Hoc Undocumented Individual Heroics (siloed by project) Dynamic Reactive	Some repeatable Some consistency Some disciplines	Documented Standards Consistency	Process metrics Controlled process Adaptable process Consistent results	Incremental Innovative Improvements
	Dataset Maturity Level				
	1 - Performed	2 - Managed	3 - Defined	4 - Measured	5 - Optimized
6) **Dataset QA/QC Processes**: Are QA/QC processes clearly defined, implemented, and trackable?	Project manually catches basic transcription errors, applies basic range limits and validation, but does not necessarily follow a QA/QC protocol Corrections are not documented	Clearly defined methodology for QA/QC using range limits and other qualifiers Data corrections are documented at project scale	Automated QA rules applied according to standardized methodologies Data corrections are tracked at data system level and versions are documented	Independent verification process, metrics and possible statistical analysis used to validate All corrections are documented and sent to dataset owner	Data collectors are trained in data collection protocol and quality assurance best practices pre-season. Automated quality control processes detect data quality issues. QC results are reviewed post-season and lessons learned incorporated into DM plan and data collection system. Software Tools are used to validate DES rules and check for inconsistent outliers; Automated rules applied include regional standards
8) **Dataset Metadata**: Measures metadata adequacy, consistency, and availability for accessibility.	Metadata documentation exists only in reports	A metadata management process is established and followed Metadata documentation captures data interdependencies Metadata categories, properties, and standards are established and followed	A metadata management strategy for the project is established with input from relevant stakeholders Metadata management efforts are centralized and overseen by data governance, including additions and changes Measures and metrics are used to evaluate the accuracy and adoption of metadata	Metadata types and data definitions support consistent import, subscription, and consumption practices The metadata repository extensions include exchange data representation standards used by the project New metadata management activities are guided by metadata metrics and historical information about metadata Quantitative objectives guide metadata management and support process performance	Root cause analysis is conducted to reduce the variations between the repository information and the data it describes Performance prediction models guide changes in metadata management processes Quantitative metadata improvement objectives are derived from the metadata strategy Planned data changes are evaluated for impact on the metadata repository; and metadata capture, change, and refinement processes are continuously improved.

Data Management Maturity Model

How do we each score a similar dataset?
Use Case – Spawning Ground Surveys

Discussion

Confederated Tribes of the Warm Springs
Indian Reservation of Oregon

Data Management Maturity Model

Yakama			Last year comparison
# Data Sets	13		8/10 received better score
# Data Systems	3		All same score as last year

Comments referred to the data systems:

- Michelle and Binh might be better for this exercise than I am. {STAR/123}
- System uses CDMS as back-end; Joe Nowinski might be better evaluator. {123/CDMS}
- A long way to go to get there {referring to project level management}. Could use 1 or 2 staff who have these types of things as primary job responsibility.

Data Management Maturity Model

CRITFC		
# Data Sets	4	
# Data Systems	1 (CDMS)	

Comments applied to most of the datasets:

1. Appoint data steward
2. Define dataset lifecycle tasks
3. Define roles and responsibilities
4. Allocate time/funding for data management
5. Develop data collection, QA/QC process, reporting, and metadata documentation
6. Automate reporting to make summaries easily reproducible
7. Develop a long-term plan for this (one) dataset

Data Management Maturity Model

NPT		
# Data Sets	4	
# Data Systems	2	

Comments:

Comments About Datasets' category levels:

For 'Defined' {level 3}, what does it mean for a dataset design to be reusable?

Confusing because CDMS inherently "re-uses" dataset designs for projects using the same datastore.

Would like an example of level of detail for "DM tasks and strategy defined in project reporting".

'#4-Measured', how would the "performance of a dataset design" be measured?

'#5-Optimized', bullets referencing Data Governance seem like they should be part of next category. What are "industry standards"?

'#3-Defined', request example of dataset governance documentation. Who is documentation intended for (internal only??)?

….{and many more comments, will make available in final slide deck}

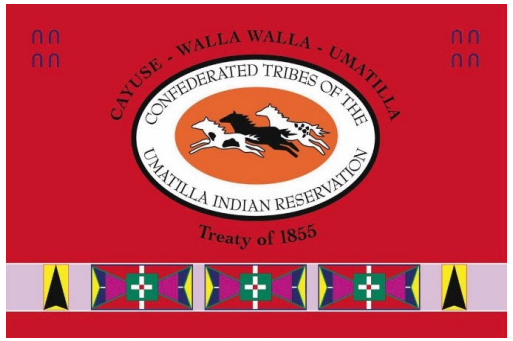
Confederated Tribes of the Warm Springs
Indian Reservation of Oregon

Data Management Maturity Model

CTWSRO		
# Data Sets	3	
# Data Systems	1 (CDMS)	

Comments:

- Filled these out to the best of my understanding.
- The huge caveat is that I don’t have access to the CDMS being that I'm in a remote office - I think someone in Warm Springs who has actually interacted with the CDMS would provide a different perspective.
- Many of these things are under development too, so I'm optimistic that a lot of these scores improve over the next 12 months or so.
- I think for all three of these datasets the next steps involve continuing to develop the workflow for integrating them into the CDMS and defining roles and definitions for things like QA/QC, reporting, and life cycle management. I think we're trending in that direction, but it's a process.



Data Management Maturity Model

CTWSRO		
# Data Sets	4	
# Data Systems	1 (CDMS)	

Comments:

Current defined Criteria are confusing and hard to distinguish the higher levels (3 and up)

- Need to refine to very definitive steps:
- Paper > Electronic>Auto Import

Original intent of the DMMM was for CTUIR to evaluate our own movement

- Discussion points with other departments within CTUIR
- Not meant to be a grading system just a visual representation of where things are
- Some datasets will never be above a certain level due to their needs
- Not meant to compare the 4 tribes to each other