Stream Management Plan Workshop

Step 5 - Existing Data Collection

Data as a Platform

Colorado Water Congress Conference Workshop, August 23-24, 2016

Steve Malers, Open Water Foundation
steve.malers@openwaterfoundation.org
www.openwaterfoundation.org
Open Water Foundation
Social enterprise 501(c)3 nonprofit focusing on developing open source software tools to help make better decisions about water resources. Water is a public resource, and water data and software tools should also be public.

open data | open software | open decisions
openwaterfoundation.org
Stream Management Plan

“(5) collecting and synthesizing existing data describing flows for river ecosystems, boating, or other needs in the watershed”

- Colorado Water Plan

Translation: How much wet water do you need, where, and when for environment and recreation?

Also: Water touches everything...and so does data.
Some Initial Thoughts

• Draw a picture of what you want to do
  ▪ It will guide what you need as input, process, and output
  ▪ Sooner or later you will need to explain what you are doing with a picture

• Good ideas come from anywhere...you just need to find them (or have someone show you)

• Create a culture of true innovation... “the art of the possible”

• Create and leverage building blocks
“There are two goals when presenting data: convey your story and establish credibility.”

-Edward Tufte

Main activities around data:
- collect
- manage
- enhance/modify/quality control
- use to analyze/model
- visualize
- attribute/document/archive/publish

Do you need a data management plan in your stream management plan?
- Perhaps, but keep it simple via conventions (who decides on the conventions?)
Diffusion of Innovation

Crossing the Chasm (Geoffrey Moore)

- Plan
- Do
- Check
- Act


http://en.wikipedia.org/wiki/Crossing_the_Chasm
Time is REALLY Important

• Time series of data
  ▪ Raw time series data
  ▪ Time series of spatial data layers
  ▪ Animations?
  ▪ What scale/precision? real-time? daily/monthly/annual historical?

• Timestamping/versioning datasets
  ▪ Evolving datasets must be tracked
  ▪ Timestamps need standards: YYYY-MM-DD, YYYYMMDD

• The meaning of time:
  ▪ Observation date/time (when measured)
  ▪ Effective date/time (when something goes into effect)
  ▪ Publish date/time, including software versions

Things change so timestamp is important!
Location is **REALLY Important**

- Spatial coordinates (geographic or other coordinate system)
- Relationship (upstream, downstream, network)
- Stream mile (from state line, confluence)
- How important are accuracy and precision?

*Things change*!

*Things change in time... and space!*
Data schema is **REALLY** Important

- “schema” = organization or structure
- Consistent definitions
- “conceptual integrity” = things fit together
- Unique data identifiers
  - ideally NOT internal database identifiers
  - best to use original source identifiers used in business processes if possible
- Relationships between datasets
  - direct often better (example=stream ID)
  - inferred or calculated may be more difficult to populate and may not be as good (spatial intersect)
- Transportability (works with different computers, tools)

**Does dataset have a time element?**

**Does dataset have a spatial element?**
Be a “Systems Thinker”

Urban Water Provider System Components

Storage

Downstream

Upstream

Self-Supplied Industry

Irrigated Agriculture

Environment & Recreation

Storage

Surface water diversions

Groundwater withdrawals

Springs

Effluent treatment plant

Water treatment plant

Wholesale provider (raw/treated)

Distributed potable water

Distributed non-potable water

Resale entity

Urban Water Provider System Components

System losses

Treated wastewater effluent

Sewage collection system

Collection / Detention / Runoff

Wastewater treatment plant

Treated wastewater

Storage

Surface and subsurface (delayed) flows

Irrigated Agriculture

Environment & Recreation

Environment & Recreation

Self-Supplied Industry

How are data uniquely identified in a larger system?
How might data be used outside of specific analysis/process?
Open Data Definition

- **Accessible to all** - the data becomes accessible outside of the organization that generated or collected it
- **Machine-readable** - data must be usable, which means it must be made available in formats that are easily used by third-party applications
- **Free** - zero or low costs for data access and openness
- **Unrestricted rights to use** - data that is unencumbered by contractual or other restrictions leads to the maximum potential of innovation

- “Generating Economic Value through Open Data” in “Beyond Transparency”
The Role of Automation

• Automation is fundamental to software/process testing and validating tools.
• Automation requires documentation, simplifies sharing.
• Manual processes are error-prone.
• Manual processes do not scale.
• “Automation frees people to do their best work”.

...need to understand the opacity/transparency of automation
One-off vs Framework/Platform

- Just do it!
- I don’t have time to figure out a better way!
- Don’t expect to do task again.
- Manual effort.
- Effort and cost?
- Who is responsible?

- Invest in tools and processes.
- They can and should be flexible to a degree.
- Assume that a task will need to be repeated later.
- Build a prototype first.
- Effort and cost?
- Who is responsible?

Is there a need for some consistent baseline data, processes, and tools across basins?
(some of you) Embrace the Command Line

Windows 7: Accessories / Command Prompt

Windows 10: All Apps / Windows System / Command Prompt

Cygwin, Linux, probably a “bash” shell:
(some of you) Embrace the API

API = Application Programming Interface
• Documented open standard for requesting and receiving data
• Integrates with software and allows for automation
• Forces publishers and consumers of data to understand the content

You should be asking: “Does it have a REST API?”

Input:
http://someserver/some/path/to/data?query1=A&query2=B&format=excel

Output:
Data in a format that you can use: text, csv, JSON, XML, Excel, map layer

Simplest example: a URL that provides a data file
Wouldn’t it be nice if....?

• All R2CROSS datasets are georeferenced and available for download?...even if no instream flow right resulted?
• All useful spatial datasets are linearly referenced to streams with stream mile?
• All useful datasets include attributes that allowed them to be linked spatially and temporally for analysis?
• Datasets are not always distributed as the “latest” but have version dates and a history so that we can understand trends?
• Datasets are publicly available for download?
Useful Concept: The “Snake” Diagram

STATEWIDE
Summary of Observed Wet & Dry Surface Water Hydrology

MAP KEY
- Gages
- Typical Wet Year
- Typical Dry Year
- Streams
- Reservoirs (> 10,000 AF)
- Major Hydrologic Basins in Colorado

NOTE: Wet and dry typical hydrology years determined separately for each basin.
Google Traffic Map

Red=bad, orange=a little bad, green=OK, but need to understand context of “bad”

So how does Google do it? Smart phone crowdsourced data. Resolution to ~300 feet.

People are now familiar with these types of products. Can stream conditions be similarly visualized?
National Hydrography Dataset (USGS)

NHD = complicated


Simpler alternative: Division of Water Resource’s Source Water Route Framework (SWRF)

http://cdss.state.co.us/GIS/Rages/AllGISData.aspx
## Stream Mile Representation Framework from South Platte/Metro Basin Implementation Plan

**E&R Attributes**
- Instream Flow Reaches
- Projects & Methods
- Stream Gages
- Diversion Headgates
- Organization Extent
- Master Plans
- and more...

**Stream mile + Excel = accessible data**
Spatial Scale/Scope: Region/Stream Reach/Focus Area

- **Basin** = Water District/Division, HUC
- **Stream Reach** = Stream mile or landmark range in basin
- **Focus Areas that were used in SWSI & BIPs**
  - Are they needed?
  - Focus area could instead be a basin (HUC)
  - Focus area could instead be a stream reach indicated by stream miles or landmark (many are already doing this).
Some Data Sources

- CDSS: cdss.state.co.us
- CDSS map viewer: see above
- CWCB data viewer: https://www.coloradodnr.info/H5V/?Viewer=cwcbmap
- DWR: water.state.co.us
- Open data portal: data.colorado.gov
- USGS NWIS: http://waterdata.usgs.gov/nwis
- RCC ACIS (climate): http://www.rcc-acis.org/
- EPA: https://www3.epa.gov/storet/web_services.html
- CDSN (water quality): http://www.coloradowaterdata.org/
- CNHP: http://www.cnhp.colostate.edu/
- DataBasin.org
- Watershed coalitions, Northern Water, river districts, etc.
- data.openwaterfoundation.org
CWCB Data

How would I find data for CWCB environmental and recreational, and stream restoration projects?

“I recommend directing them to CWCB for data related to the Non-Consumptive Needs Assessments and Basin priorities by stream or watershed. Beyond that, I recommend that watershed groups work within their own collaborative to assemble data important to their objectives in Stream Management and/or Watershed planning. This may include water quality data, USFS inventories (roads, culverts), CPW fish data, ecological data generated by local assessments, burn severity data, topography (survey),...”

http://cwcb.state.co.us/environment/non-consumptive-needs/Paes/main.aspx
http://cwcb.state.co.us/ENVIRONMENT/WATERSHED-PROTECTION-RESTORATION/Pages/main.aspx
data.colorado.gov

Source for many “flat” datasets and provides an open data platform for distributing datasets
CDSS and CWCB Map Viewers

Good for locating data, ad hoc queries

Limitation - limited data download, lack of automation
Like many web tools, it can be difficult to find documents if you don't know what to request, some scans are large and slow to download.
Conservation Easements in Colorado’s Front Range

By Gregory Stavish

Mar 25, 2014 (Last modified Apr 3, 2014)

A GIS-Based Assessment to Determine the Extent Conservation Easements Protect Biodiversity

Biodiversity in the Colorado Front Range is a critical component for the sustainable development of the region (Chovan and Bernstein, 2010; Pague, et. al, 1996; Pague, et al, 1993). Modern land conservation efforts, which are often a result of donated gifts or opportunistic purchases, risk ignoring areas of high biodiversity (Whittaker, 1999; Daniels and Daniels, 2003). The nature of establishing conservation easements is, at times, unpredictable (Axel-Lute, 1999). Planning and executing conservation easements requires suitable geographic conditions, local outreach, and long-term negotiations with landowners.

A GIS-based analysis was conducted to determine whether conservation easements have protected high biodiversity areas from development. Land trusts, as well as other conservation organizations, can use this information to target outreach campaigns, gain public support, acquire new funding opportunities, and improve public education of the benefits of biodiversity conservation.

Potential Conservation Areas (PCAs) are “land areas that can provide the habitat and ecological processes upon which a particular element (i.e., rare plant and animal species and significant plant communities) occurrence, or suite of element occurrences, depends for its continued existence” (Lyon, et. al. 2001). PCA boundaries are determined by county-wide biological assessments conducted by the Colorado Natural Heritage Program. PCAs are ranked according to their biodiversity significance, ranging from 1 (greatest) to 5 (least). This study only analyzed categories 1 to 3. Each biodiversity category was delineated, analyzed, and categorized.
OWF is experimenting with providing value-added datasets to facilitate complex analysis.
HUC 10190007 (Cache La Poudre) Surface Water Supply

Monthly component volumes

HUC 10190007 (Cache La Poudre) SWSI

Historical paired SWSI values establish the SWSI distribution to lookup recent and current SWSI values.

TSTool software

Time Series Analysis and Automation
Summary

• Keep it simple
• Don’t let technologies abuse you (but they will)
• Understand your data and process limitations
• Prototyping is OK, dead ends happen, but have a roadmap so you know your destination
• Collaboration is easier with open standards and open data
• Learn from others
• Try to be agile

“A river is more than an amenity. It is a treasure. It offers a necessity of life that must be rationed among those who have power over it.” – Supreme Court Justice Oliver Wendell Holmes, Jr. (quote from 1931)
The content of this presentation is licensed under the Creative Commons Attribution NonCommercial License: http://creativecommons.org/licenses/by-nc/4.0