Public Agency Inventory Process
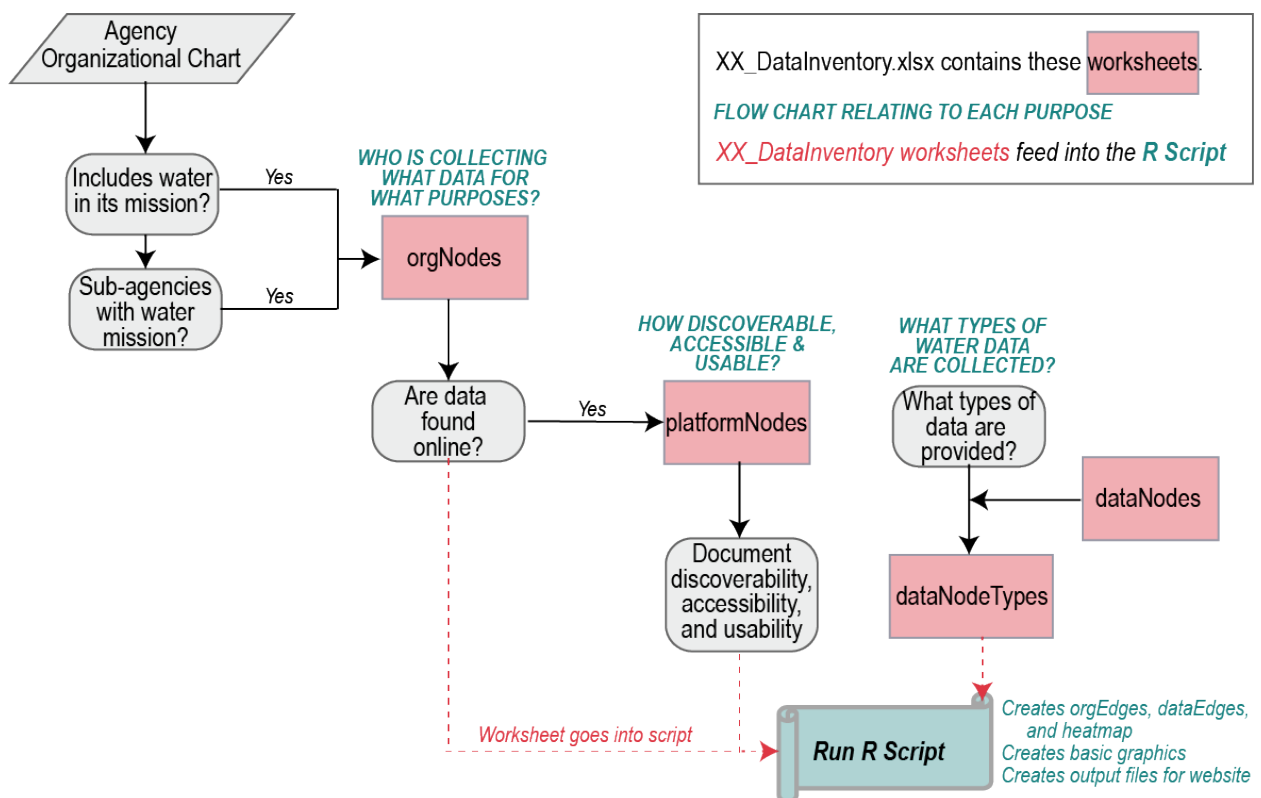
Last Updated: September 10, 2019

File names and worksheets are in red.
Column headers are italicized in teal.

Purpose of Inventory

From a general public, outsiders perspective we want to know:

- Who is collecting water data and for what purposes?
- What types of water data are collected?
- How discoverable, accessible, and usable are those data for secondary users?

Figure: Flow chart of the process to create a data inventory.

Templates and Metadata

The IoW has developed a template and provided metadata so that federal and state public agencies may create their own inventory and, if desired, share that inventory with the IoW online.

- XX_DataInventory.xlsx provides a template for inventorying water data held by other state agencies. A brief description of tabs is provided here. See metadata_DataInventory.xlsx for column descriptions.
- metadata_DataInventory.xlsx provides descriptions of the tabs, column names, and dropdown menus selections.

Process

Fill out XX_DataInventory.xlsx

- Obtain an organization chart for the executive branch of government.

orgNodes worksheet

- Refer to the metadata_DataInventory.xlsx to better understand column headers described below. Snapshots of worksheets will also be displayed here.
- For each entity in the organizational chart – find on website, location vision/mission statement. If there is a water component to the entity, then add to the orgNodes worksheet. Use the vision/mission statement and the website to assess data purposes. Data purpose categories are described here: https://internetofwater.org/valuing-data/moving-towards-valuation-by-data-purpose/#Categorizing

entityID	entity	tier1and2entity	tier1and2entityID	entityIDAbove	mission	website	operational	decision	regulatory	research	level
n01	Executive Branch	Executive Branch	1	NA	Implement and enforce	https://www.whitehous	Yes	Yes	Yes	Yes	1
n02	Department of Agricult	Department of Agricultur	2	n01	Provide leadership on f	https://www.usda.gov/	Yes	Yes	No	Yes	2
n03	Agricultural Research S	Department of Agricultur	2	n02	Conduct research to de	https://www.ars.usda.gc	No	Yes	No	Yes	3
n04	National Agricultural S	Department of Agricultur	2	n02	Provide timely, accurate	https://www.nass.usda.g	No	Yes	No	No	3
n05	Natural Resources Con	Department of Agricultur	2	n02	Provide resources to fal	https://www.nrcs.usda.g	Yes	Yes	No	No	3
n06	Economic Research Ser	Department of Agricultur	2	n02	Anticipate trends and e	https://www.ers.usda.gc	No	Yes	No	Yes	3
n07	Department of Comme	Department of Commerc	3	n01	Promote job creation, e	https://www.commerce.	Yes	Yes	No	Yes	2
n08	National Oceanic and /	Department of Commerc	3	n07	To understand and predict	http://noaa.gov/	Yes	Yes	No	Yes	3
n09	National Weather Serv	Department of Commerc	3	n08	Provide weather, water	https://www.weather.go	Yes	Yes	No	Yes	4

- Use the organizational chart to assign *tier1and2entities* and *levels*. *tier1and2entities* refer to the executive branch (tier1) and the next level of government (tier2). The *tier1and2entityID* refers to a unique number given to each tier1 (should just be the executive level of government) and tier2 entities (often these are departments). *Level* refers to how far removed an entity is from the executive branch. The *entityIDAbove* refers to the entity organizationally located above the current entity. This column is used to create the orgEdges worksheet.

orgEdges worksheet (optional)

- Use the orgNodes worksheet to create the hierarchical relationships showing how these organizations are structured. Skip the first row and copy entityIDAbove and paste into the *from* column. Next, skip the first row and copy entityID and paste into the *to* column. The r-script will also create this worksheet.

from	to
n01	n02
n02	n03
n02	n04
n02	n05
n02	n06
n01	n07
n07	n08
n08	n09
n01	n10

platformNodes worksheet

- This worksheet lists the name of the platform or point of data access. The entity refers to the *orgNodes* entity who is collecting and/or hosting the data. Sometimes data are provided by a combination of organizations and will not have a direct correlation to the orgNodes worksheet. Create a web label that readily identifies the hosting entities. The *heatmapLabel* should also include the name of the platform.
- Using the metadata_DataInventory.xlsx, provide an assessment of how discoverable, accessible, and usable the data are as provided in the columnsDescription and metricsKey worksheets.

platformID	platform	entity	entityID	webLabel	heatmapLabel	tier1and2entity	tier1and2entityID	website
p01	HSIN-CI Dams Portal	Office of Infrastructu	n26	DHS-IP	DHS-IP: HSIN-CI Dam	Department of Homeland	6	https://www.dhs.gov/I
p02	Flood Map Service Center	Federal Emergency N	n24	DHS-FEMA	DHS-FEMA: Flood Ma	Department of Homeland	6	https://msc.fema.gov/
p03	IP Gateway	Office of Infrastructu	n26	DHS-IP	DHS-IP: IP Gateway	Department of Homeland	6	https://www.dhs.gov/i
p04	No Public Portal	National Geospatial	n11	DOD-NGIA	DOD-NGIA: No Publi	Department of Defense	4	https://www.nga.mil/F
p05	Corps CWMS	Army Corps of Engin	n12	DOD-USACI	DOD-USACE: Corps C	Department of Defense	4	http://water.usace.arr

easeDisco	methodDi	easeAcces	methodAc	fileFormat	metadata	metadataI	dataDefini	metaAttril	metadata!	timeliness	lengthAva
No	Unknown	Permissio	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Current O
Low	Website	Software F	Individual	pdf; softw	No	xml	Yes	Administr	No	Irregular	Unknown
No	Unknown	Training R	Unknown	Unknown	No	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
No	Unknown	No	Unknown	Unknown	No	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
Medium	Map	Yes	Individual	csv; png; i	Yes	html	No	Hybrid	No	Daily or Hi	Limited Re

dataNodes worksheet

- This tab is static and refers to the categories of water data used in this data inventory

dataID	dataGroup	dataCategory
d01	Built	Infrastructure
d02	Natural	Infrastructure
d03	Quality	Quality
d04	Regulatory	Quality
d05	Evapotranspiration	Quantity
d06	Extreme Events	Quantity
d07	Glacial and Snow	Quantity
d08	Groundwater	Quantity
d09	Meteorology	Quantity
d10	Precipitation	Quantity
d11	Reservoir	Quantity
d12	Soil	Quantity
d13	Surface Water	Quantity
d14	Hydropower	Use
d15	Irrigation	Use
d16	Use	Use
d17	Utilities	Use
d18	Management Plans	Use

dataNodeTypes worksheet

- For each platform, list the data being collected and categorize by data group and category. For instance if a platform collects streamflow, gage height, groundwater levels, and water quality parameters then it should have three rows:
- The *dataID* can be calculated using the vlookup command.

platfo	entity	tier1and2entity	platform	website	data	dataCategory	dataGroup	dataID
p51	National Weather Servi	Department of Comme	Advanced Hydrologic Pred	https://water.w	Precipitation as F	Water Quantity	Precipitation	d10
p51	National Weather Servi	Department of Comme	Advanced Hydrologic Pred	https://water.w	River Stage	Water Quantity	Surface Water	d13
p34	Earth Science Division	National Aeronautics a	Alaska Satellite Facility DA	https://www.as	Wetland Extent	Infrastructure	Natural	d02
p34	Earth Science Division	National Aeronautics a	Alaska Satellite Facility DA	https://www.as	Soil Moisture	Water Quantity	Soil	d12
p24	Environmental Protecti	Environmental Protecti	ATTAINS	https://www.ep	303d Assessment	Water Quality	Regulatory	d04
p24	Environmental Protecti	Environmental Protecti	ATTAINS	https://www.ep	Source of Impairr	Water Quality	Regulatory	d04
p24	Environmental Protecti	Environmental Protecti	ATTAINS	https://www.ep	TMDLs	Water Quality	Regulatory	d04

dataEdges worksheet (optional)

- The *to* column is the *platformID* on the dataNodeTypes worksheet. The *from* column is the *dataID* on the dataNodesTypes worksheet. The r-script will also create this worksheet.

from	to
p51	d10
p51	d13
p34	d02
p34	d12
p24	d04
p24	d16
p06	d07

Run R Script

- Scripts assumes the following filing structure
- Data Folder
  - xx (state abbreviation) folder
    - XX_DataInventory.xlsx file
- The code and a description of what each step does is found in: 2019_08 R script.pdf.
- The script can be downloaded here: https://github.com/internetofwater/DataInventory

Inventory_markdown.rmd

- Open inventory_markdown.rmd
- change the directories, folderName, and filename to obtain spreadsheet of interest. For example:
  - swd <- "c:\\Users\\Username\\Documents\\dataInventory\\data\\"
  - folderName <- "federal"
  - filename <- "Federal"
  - filledFolders <- c("federal", "ca", "nc", "tx")
    - as hub inventories are completed, add the state to the list.
- change the html directory to point to the web tool
  - *Note that at some point the directory steps can be largely automated to loop through all states if necessary.
  - *make sure you add the new state folder to save files to

- The file will do the following:
  - Read in the excel spreadsheets
  - Create the orgEdges and dataEdges spreadsheet
  - Combine platformNodes and dataNodes worksheets into a single file for the website
  - Create the heatmap spreadsheet
  - Create the combined scorecard for all completed hub inventories
  - Create basic graphics similar to what is on the website