

State Agency P2P Network Webinar Series

Ask the Data Architect: Presented by Kyle Onda

July 17, 2020





Program Updates

- **Welcome**
 - **Peter Colohan, Executive Director, Internet of Water**
- **Take note for today's webinar**
 - **We are recording!**
 - **Other administrative notes**
- **State Agency Peer-to-Peer (P2P) Network**



Webinar Presenter



Kyle Onda
Data Architect

Internet of Water P2P
Webinar Series Presents:

Ask the Data Architect

"Finding and integrating relevant data from various sources is a significant bottleneck in converting data into actionable information."



Ask the Data Architect

Outline

- Address Key Specific Questions
- Address General Interest Patterns
- Overview of Geoconnex





Question 1: Water Use Databases

“Within my organization we want to create a fully functional water-use database and we want to identify **what structure** and/or database **language** provides the greatest functionality. SQL (MySQL), other languages?

What database structure do you recommend to build a water-use database?

What database language provides the most functionality (to allow for easy input, to create future web applications, to **link with other water-use databases** (USGS, EPA, etc.), etc.)?”





Question 1: Water Use Databases

Short answers:

- **What structure?**
 - Relational (SQL) Database, unless dealing with high-frequency data (consider NoSQL Key-value data stores)
- **What language/database?**
 - Most SQL databases/languages are comparable for basic performance and easiness to learn.
 - Choose based on: particular needs for extensions, existing expertise and/or acquisition costs
- **How to allow for easy input?**
 - This always takes some custom IT and scripting work across database vendors
- **How to allow for web applications and linking to other water use databases?**
 - Invest in a RESTful API
 - Consider using WaDE 2.0 Schema as a starting point (<https://github.com/WSWCWaterDataExchange/WaDE2.0>)



Question 1: Types of Databases

Relational Database Management Systems (RDMBS)/SQL

A collection of linked tables, with strict rules (schema) for what kind of data go in each column of each table

Good for:

- Highly structured data with limited types and links between entities
- Efficient queries for the same information in the same format
- Data integrity, durability

Non-Relational Databases/No SQL

Key-Value stores with flexible schema

Good for:

- Heterogeneous data (including objects like images, binary files, etc)
- Creating unanticipated links between entities
- Adding unanticipated entity types and attributes
- Simple queries of large amounts of data
- Large amounts of data with distributed data storage across many servers



Question 1: Which Type of Database?

Relational Database Management Systems (RDMBS)/SQL

Typical use cases

- Accounting Systems
- Customer Management
- Regulatory Information
- Scientific Data

Non-Relational Databases/No SQL

Typical use cases

- Searching unstructured content like collections of text documents
- Searching over relationships between nodes in a network
- IoT applications with high-frequency time series, and varied sensors coming off and online or changing attributes frequently



Question 1: Which SQL?

Most RDBMS use some version of a Structured Query Language (SQL) for Create, Read, Update, Delete (CRUD) operations.

SQL is similar in code across database providers, with syntactical differences around:

- Date/time formats
- Quoting of names
- Null values
- Case sensitivity



Question 1: Which SQL?

What are your developers most comfortable with?



Question 1: Which SQL?

Popular RDBMS	License Cost (if your org does not already have one)	Support	Portability	read/write performance, security, etc.	High-Frequency Time Series	Geospatial Operations
Oracle	\$\$\$	Enterprise	Low	****		**** (Oracle Enterprise Only)
Db2 (IBM)	Free version or \$\$\$	Enterprise	Low	****		** (Db2 spatial extender)
MS SQL Server (Microsoft)	Free version or \$\$\$	Enterprise	Low	****		** (Performant with paid versions only)
MySQL (Oracle) /MariaDB	Free	Community	Medium	****		*
PostgreSQL	Free	Community	Medium	****	TimescaleDB	**** (PostGIS)
SQLite	Free	Community	High	**		*



Question 1: Which Database Structure?

Which structure? What Schema?

In general, this is driven by what you want to query.

In typical water-use/water rights databases, a key historical use case has been legal:

- Find all the water rights in a PLSS grid
- Look up a name or water right ID number and return the records about that water right

More modern uses include being able to summarize water use at multiple spatial and temporal scales.



Question 1: Which Database Schema?

A useful starting point is the Western States Water Council Water Data Exchange (WaDE) v2.0 Schema:

https://github.com/WSWCWaterDataExchange/WaDE2.0/tree/master/Design_docs

Designed as a “Least Common Denominator” among water rights and use databases in western states with prior appropriations water rights regimes.

- Includes SQL scripts to generate blank databases in:
 - MySQL
 - PostgreSQL
 - MS SQL
 - SQLite



Question 1: Which Database Schema?

https://github.com/WSWCWaterDataExchange/WaDE2.0/tree/master/Design_docs

The screenshot shows the AGRAM EXPLORER interface with a search bar on the left listing various tables. The main area displays three tables with their columns and data types:

OrganizationUUID	varchar (258)	PK
ReportingUnitNativeID	varchar (258)	
VariableSpecificCV	varchar (258)	
BeneficialUseCategory	varchar (500)	
PrimaryUseCategory	varchar (258)	NULL
AllocationUSGSCategoryCV	varchar (258)	NULL
WaterSourceNativeID	varchar (258)	
MethodName	varchar (258)	
TimeframeStartDate	date	
TimeframeEndDate	date	
DataPublicationDate	date	
ReportYear	varchar (4)	
Amount	float	NULL
PopulationServed	float	NULL
PowerGeneratedGWh	float	NULL
IrrigatedAcreage	float	NULL
InterbasinTransferFromID	varchar (258)	NULL
InterbasinTransferToID	varchar (258)	NULL

OrganizationUUID	varchar (258)	PK
AllocationUUID	varchar (258)	NULL
SiteNativeID	varchar (258)	
VariableSpecificCV	varchar (258)	
BeneficialUseCategory	varchar (500)	
PrimaryUseCategory	varchar (258)	NULL
AllocationUSGSCategoryCV	varchar (258)	NULL
WaterSourceNativeID	varchar (258)	
MethodName	varchar (258)	
TimeframeStartDate	date	
TimeframeEndDate	date	
DataPublicationDate	date	
ReportYear	varchar (4)	
Amount	float	NULL
PopulationServed	float	NULL
PowerGeneratedGWh	float	NULL
IrrigatedAcreage	float	NULL
AllocationCommunityWaterSupplySystem	varchar (258)	NULL
SWISIdentifier	varchar (258)	NULL
InterbasinTransferFromID	varchar (258)	NULL
InterbasinTransferToID	varchar (258)	NULL
Geometry	geometry	NULL
OrganizationUUID_1	varchar (258)	FK

RegulatoryOverlayUUID	varchar (258)	PK
OrganizationUUID	varchar (258)	
ReportingUnitNativeID	varchar (258)	
DataPublicationDate	date	
ReportYearCV	varchar (4)	

- Consider modifying or designing from scratch in cross-database Entity-Relationship Diagram (ERD) software (e.g., DBWrench, Lucidchart, sqlDBM)
- Graphically define tables, their columns, and the relationships between them
- Export SQL code that generates the blank preferred database



Q & A

Please submit your *content-related questions* in the webinar's **Q&A box**.
The moderator will read your question aloud.

Administrative questions can be placed in the **"Chat" box**.





Question 2: Utility IoT & Privacy Law

“As more water systems are beginning to incorporate **“smart” water tech** that use the Internet of Things to gather data to inform decision-making, e.g., smart irrigation controllers and consumer side of the meter leak detection devices.

What, if any, are some of the **key legal and policy challenges** water utilities are facing with respect to **privacy** and/or **4th Amendment issues**? How have these utilities navigated these challenges?”





Question 2: Utility IoT & Privacy Law

Privacy laws, where relevant, typically apply to “personal” or “customer” information.

- In California, this is “information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.”



Question 2: Utility IoT & Privacy Law

- The legal landscape is ever evolving and has generally been more controversial in energy than water.
- **Government/Law Enforcement/4th Amendment:**
 - In general, utilities must comply with court ordered, warrantless requests of smart meter data
 - **United States v. Miller (1976)** - Individuals cannot have a right to privacy under the Fourth Amendment with respect to information voluntarily given to a third party.
 - **Carpenter v. United States (2018)** - Need a warrant for cell phone location data. Narrow ruling, does not apply to much else (yet).



Question 2: Utility IoT & Privacy Law

Other Privacy Law

- **CA AB 1274 (2013):** Energy and gas consumption records can only be shared with third parties with a formal contract (e.g., academics, data analytics companies, the public) with the consent of the customer, and requires shared data to be destroyed after the contract period.
- **CA AB 1520 (2015):** Water consumption of *non-residential customers* of public utilities are subject to public records requests. Water consumption of *residential customers* are not subject to public records requests.
- **CA AB 375 (2018):** Customers must provide permission for businesses to share “personal information” with third parties, and can request any records be destroyed.



Question 2: Utility IoT & Privacy Law

Upshot

The legal implications of smart metering for water currently is unclear and varies by jurisdiction, with more legislation and litigation directly addressing telecommunications and energy information rather than water.

For many water utilities, the **“safe” choice** is to assume that any general or electricity-specific legislation applies to them, meaning:

- Internal metering and billing data systems must have “reasonable” IT security
- Customers must be notified, and be allowed to opt-out of having their information shared with any third party (other than law enforcement)
- Contracts for data sharing involving “personal information” require the destruction of copies of data held by third parties after the contract has been fulfilled
- In some cases, allowing customers to opt-out of smart metering in favor of traditional manual reads



Question 2: Utility IoT & Privacy Law

Other strategies:

- Create agreements with partnerships to securely host data in common and mediate data sharing contracts with third parties.
 - Strategic California Urban water Analytics (SCUBA) warehouses data from southern California water utilities, streamlines secure data sharing for academic and policy research.
- Deidentify and aggregate data so that it is no longer “personal” and can be shared more freely without liability.



Q & A

Please submit your *content-related questions* in the webinar's **Q&A box**.
The moderator will read your question aloud.

Administrative questions can be placed in the **"Chat" box**.





Question 3: DB Development Outsourcing

“What are some options for state agencies **outsourcing the development of web/cloud-based databases** for monthly and annual data submission and reporting?”





Question 3: DB Development Outsourcing

Commercial Off-the-Shelf Software (e.g., EarthSoft EQuiS, Kisters WISKI Water Quality, Aquatic Informatics AQUARIUS, Ribeka, Locus Tec, etc.)

IT Contractor/
Custom Database

Self-Managed
(License + Support Model)

Software as a Service (SaaS)
(Subscription Model)

Completely customized to
fit your needs

Retain control over physical
IT assets and data storage
and security

No physical IT assets,
shared responsibility for
data security

Force your data into
existing data model

Force your data into
existing data model



Question 3: DB Development Outsourcing

Cost Structure Analogies

Commercial Off-the-Shelf Software (e.g., EarthSoft EQuiS, Kisters WISKI Water Quality, Aquatic Informatics AQUARIUS, Ribeka, Locus Tec, etc.)

IT Contractor/
Custom Database

Self-Managed
(License + Support Model)

Software as a Service (SaaS)
(Subscription Model)

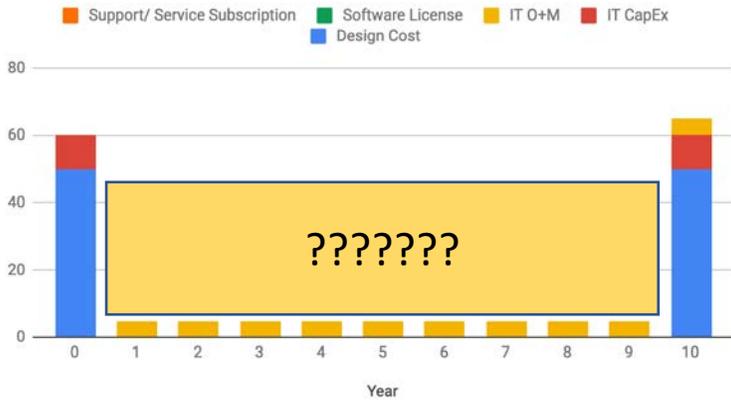




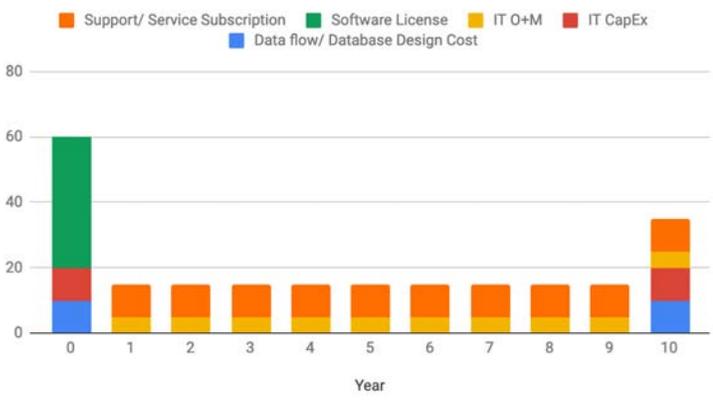
Question 3: DB Development Outsourcing

Commercial Off-the-Shelf Software (e.g., EarthSoft EQuiS, Kisters WISKI Water Quality, Aquatic Informatics AQUARIUS, Ribeka, Locus Tec, etc.)

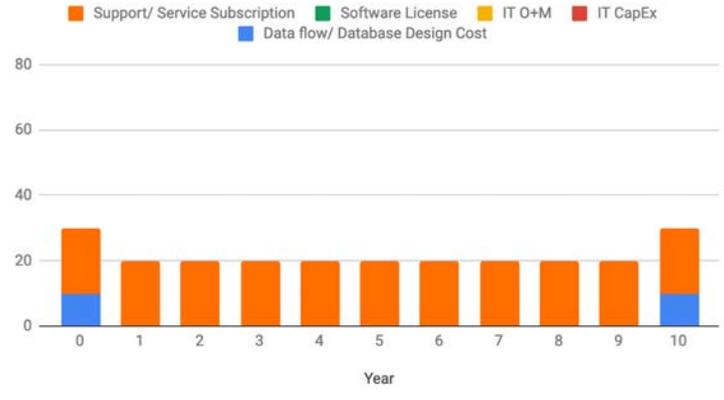
IT Contractor/
Custom Database



Self-Managed
(License + Support Model)



Software as a Service (SaaS)
(Subscription Model)



Uncertainty around support costs/database revision costs

Tradeoffs between recurring vs "lumpy" outlays



Q & A

Please submit your *content-related questions* in the webinar's **Q&A box**.
The moderator will read your question aloud.

Administrative questions can be placed in the **"Chat" box**.





Question 4: Agriculture Supply Chains

“With regard to water reporting in supply chains, how can **digital technologies** like the Internet of Things (IoT) **assist** with this in **agriculture supply chains?**”





Question 4: Agriculture Supply Chains

- “Smart farming” - e.g., optimize input (including water) efficiency through use of solar radiation, soil moisture, air temperature, precipitation, humidity sensors and water meters on withdrawal points and applicators.
- AgGateway produces software and standards to help Ag industry sectors integrate supply chain data, and has many working groups for data linkages across the sector for product traceability. <https://www.aggateway.org/>



Q & A

Please submit your *content-related questions* in the webinar's **Q&A box**.
The moderator will read your question aloud.

Administrative questions can be placed in the **"Chat" box**.





General Topics

Future Webinar Topics

- Historical data digitization
- Data integration
- Data sharing



Poll: Please participate in a quick poll!



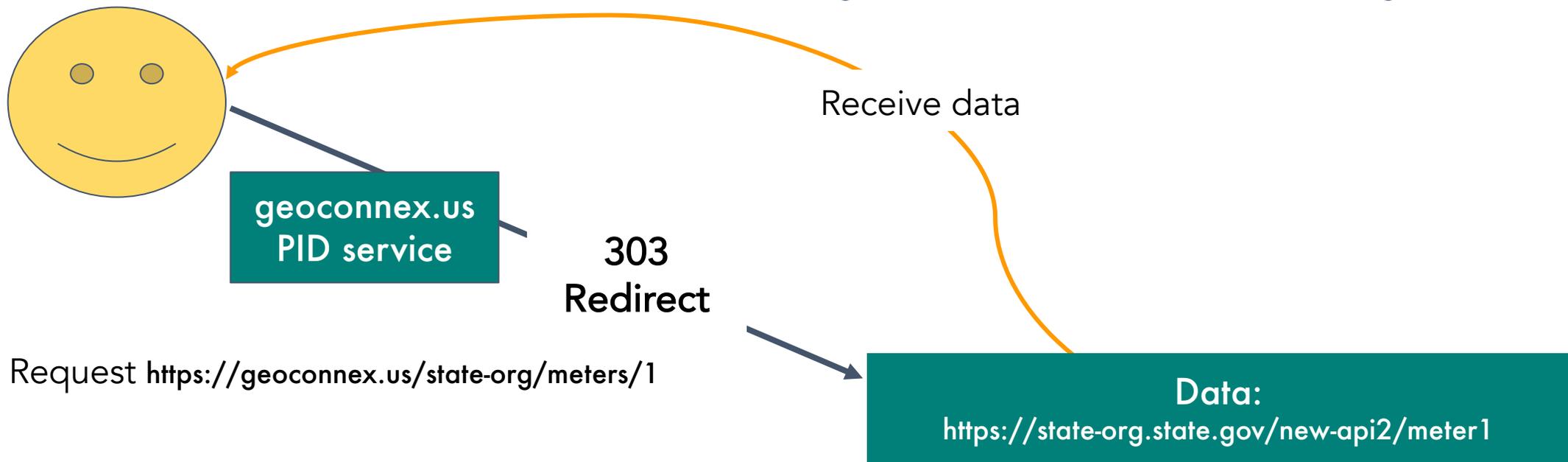


Geoconnex

Are you a data producer, publisher, or republisher, who wants your data to...

→ be persistently identified?

(make your web resources robust to website reorganizations, bureaucratic renaming)





Geoconnex

Are you a data producer, publisher, or republisher, who wants your data to...

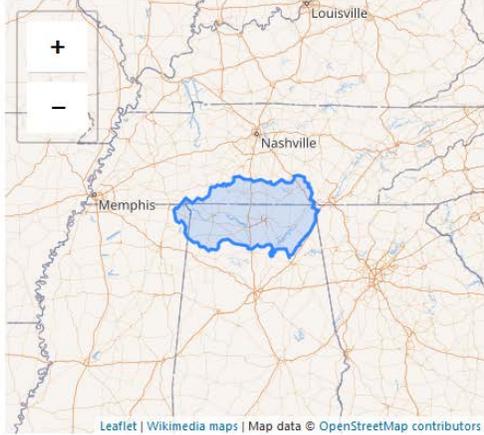
➔ be persistently identified?

➔ be linked to persistent, sector-wide reference geographies & NHDPlusV2?

info.geoconnex.us   Contact

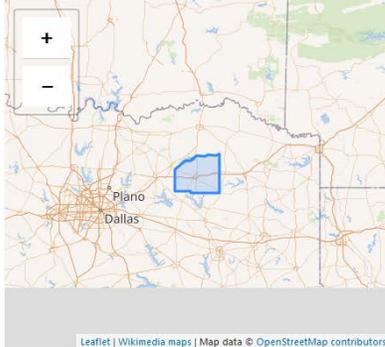
Home / Collections / HU04 / Items / Item 0603 JSON JSON-LD

Item 0603



Property	Value
id	0603
fid	4
uri	https://geoconnex.us/ref/hu04/0603
NAME	Middle Tennessee-Elk
gnis_url	
GNIS_ID	None

Item 48223



Property	Value
id	48223
fid	4
STATEFP	48
COUNTYFP	223
COUNTYNS	01383897
AFFGEOID	0500000US48223
NAME	Hopkins
LSAD	06
uri	https://geoconnex.us/ref/counties/48223
census_profile	https://data.census.gov/cedsci/profile?g=0500000US4...



Geoconnex

Are you a data producer, publisher, or republisher, who wants your data to...

➔ be persistently identified?

➔ be linked to persistent, sector-wide reference geographies & NHDPlusV2?

➔ be easily discoverable via Google Search?

Google Structured Data Testing Tool

https://info.geoconnex.us/collections/gages/items/1000002

```

1 <!doctype html>
2 <html lang="en">
3 <head>
4 <meta charset="utf-8">
5 <title>geoconnex reference features - Reference Gages - 1000002</title>
6 <meta name="viewport" content="width=device-width, initial-scale=1">
7 <meta name="language" content="en-US">
8 <meta name="description" content="geoconnex reference features">
9 <meta name="keywords" content="geospatial,data,api">
10 <link rel="shortcut icon" href="https://info.geoconnex.us/static/info-ico.png">
11 <link rel="stylesheet" href="https://unpkg.com/mini.css@3.0.1/dist/mini.css">
12 <link rel="stylesheet" href="https://info.geoconnex.us/static/css/main.css">
13 <!--[if lt IE 9]>
14 <script src="https://cdnjs.cloudflare.com/ajax/libs/html5shiv/3.7.3/html5shiv.min.js"></script>
15 <![endif]-->
16
17 <link rel="alternate" type="application/geo+json" title="This document as GeoJSON" />
18
19 <link rel="alternate" type="application/ld+json" title="This document as JSON-LD" />
20
21 <link rel="self" type="text/html" title="This document as HTML" />
22
23 <link rel="canonical" href="https://info.geoconnex.us/collections/gages/items/1000002" />
24
25 <link rel="collection" type="application/json" title="Reference Gages" href="https://info.geoconnex.us/collections/gages" />
26
27 <link rel="prev" type="application/geo+json" title="" href="https://info.geoconnex.us/collections/gages/items/1000001" />
28
29
30
31
32
33

```

https://purl.org/geojson/vocab#Feature

ID: https://geoconnex.us/ref/gages/1000002

@type	https://purl.org/geojson/vocab#Feature
@id	https://geoconnex.us/ref/gages/1000002
fid	2
https://schema.org/name	Big Black River near Depot Mtn, Maine
https://schema.org/description	USGS NWIS Stream/River Site 01010070: Big Black River near Depot Mtn, Maine
https://schema.org/subjectOf	https://waterdata.usgs.gov/monitoring-location/01010070
provider	https://waterdata.usgs.gov
provider_id	01010070
nhdpv2_REACHCODE	01010007007951
nhdpv2_REACH_measure	60.89532
nhdpv2_COMID	4289553
https://purl.org/geojson/vocab#properties	
@type	https://purl.org/geojson/vocab#Feature
@id	https://geoconnex.us/ref/gages/1000002

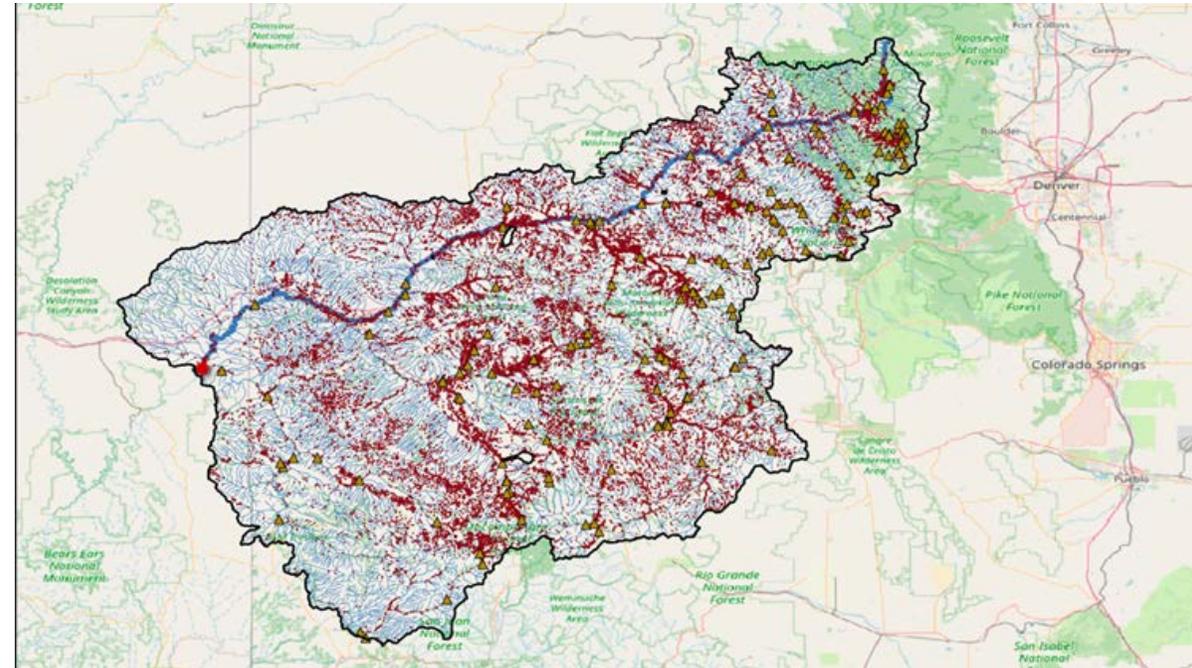




Geoconnex

Are you a data producer, publisher, or republisher, who wants your data to...

- ➔ be persistently identified?
- ➔ be linked to persistent, sector-wide reference geographies & NHDPlusV2?
- ➔ be easily discoverable via Google Search?
- ➔ be discoverable alongside data from other organizations?



- ▲ USGS NWIS Sites
- WaDE Water Points of Diversion
- NHDPlus V.2 Flowlines





Geoconnex

Are you a data producer, publisher, or republisher, who wants your data to...

- be persistently identified?
- be linked to persistent, sector-wide reference geographies & NHDPlusV2?
- be easily discoverable via Google Search?
- be discoverable alongside data from other organizations?

Join us on:



GitHub

<https://github.com/internetofwater/geoconnex.us>



ESIP-ALL Slack

<https://esip-slack-invite.herokuapp.com/>

Channels #geoconnex and #hydro-indexing-people



Q & A

Please submit your *content-related questions* in the webinar's **Q&A box**.
The moderator will read your question aloud.

Administrative questions can be placed in the **"Chat" box**.





Final Notes

Presenter Contact information:

Kyle Onda, IoW Data Architect, kyle.onda@duke.edu

IoW Contact Information:

internetofwater@duke.edu

<https://internetofwater.org/>

SAP2P Network:

Website: <https://internetofwater.org/peer-to-peer-network/>

Webinars: <https://internetofwater.org/webinars/>

Forum: <https://stateagencyforum.internetofwater.org/>



Poll: Please participate in a quick poll!

Follow-Up Information

- Links to the webinar recording and slides will be distributed once posted

Thank you for your participation!