

# Geoconnex WRI Report

## Table of Contents

<b>1 Purpose</b>	<b>2</b>
<b>2 Executive Summary</b>	<b>2</b>
<b>3 Introduction</b>	<b>2</b>
3.1 Objectives . . . . .	2
3.2 Background . . . . .	3
3.3 Overview . . . . .	5
<b>4 Glossary</b>	<b>5</b>
<b>5 Use Cases</b>	<b>6</b>
5.0.1 . . . . .	7
5.0.2 Discovering and Distinguishing Hydrologically Relevant Observed and Modeled Datasets . . . . .	7
<b>6 User Research</b>	<b>7</b>
6.1 Outreach through Internet of Water network . . . . .	7
6.2 Domain Meetings and Conferences . . . . .	7
6.3 Interviews . . . . .	7
<b>7 Architecture and Implementation of Geoconnex Linked Data     System</b>	<b>7</b>
<b>8 Governance</b>	<b>7</b>
8.1 Functional Requirements Research . . . . .	7
8.1.1 Survey Questions and Results . . . . .	8
8.2 Proposed Governance Plan . . . . .	9
8.2.1 The Geoconnex Working Group . . . . .	9
8.2.2 USGS-EPA Joint Committee on Geoconnex . . . . .	10

# 1 Purpose

The purpose of this report is to document the activities, outputs, and outcomes of the NCSU Water Resources Research Institute award for the “Internet of Water: Research and Development toward a linked data system and foundational knowledge network for the Internet of Water”. The linked data system and foundational knowledge network, named Geoconnex, was conceptualized as an operationalization of the Open Geospatial Consortium Environmental Linked Features Interoperability Experiments for the United States in the domains of water science and management. The outcomes include a performant infrastructure leveraging semantic technology and open, modern API standards that allow data providers to independently publish metadata on the web in a manner that results in their data becoming linked to other providers’ data where spatially, hydrologically, and topically relevant. If adopted by data providers at wide scale, this would enable much improved discoverability of water-related datasets. Further work is needed to encourage participation in the system and use of the infrastructure. Potential future work could also establish best practices to use the same technologies to enable the automatic translation of observation and model data across data systems, fostering improved interoperability of data in addition to the improved discoverability already enabled by the currently implemented infrastructure.

## 2 Executive Summary

## 3 Introduction

### 3.1 Objectives

The main objective of this project is to establish a foundational framework for a contributor-based system that facilitates regular harvesting and cross-referencing of metadata through two key components:

1. Facilitating the Internet of Water community to publish detailed, machine-readable, and cross-referenced metadata (linked data)
2. Developing a centralized crawler/harvester to catalog all the linked data into a single knowledge graph, serving as a part of the index for eventual Internet of Water search utilities.

The objectives, therefore, encompassed both data-publisher-oriented research and development, as well as the establishment of a centralized knowledge network and services.

Specifically, the project aims to:

- Generate a demonstrative set of reference web content concerning various environmental features such as watershed boundaries, stream reaches, aquifers, monitoring locations, administrative geographies, and

water-related infrastructure (e.g., dams, bridges) for data-publishing organizations to link to.

- Investigate existing hydrologic and web ontologies pertinent to the publication of data relevant to water science and management, and develop guidance for embedding linked-data content into web resources about water data.
- Develop a use case regarding the discovery and use of modeled and observed data for the same real-world feature or system of interest, and to investigate the metadata and knowledge graph infrastructure requirements to realize that use case.
- Develop software tools that enable data providers to generate and publish linked data without imposing burdensome requirements on existing data systems. All documentation and code were made publicly accessible.
- Create an open source web crawler/harvester infrastructure tailored to water data that can navigate web pages about water data, follow embedded links to other water data web pages, and harvest and catalog the metadata and their linkages to construct a single knowledge graph linking all water data.
- Investigate what governance mechanisms might be appropriate to establish and maintain the system to account for the needs of the water data community, incentivise participation by data providers, and facilitate use by data users.

### 3.2 Background

The Internet of Water is an initiative that aims to create a network of interconnected water data systems, modeled after the organizational structure of the Internet, as recommended by the Aspen Institute Dialog Series on Water Data Internet of Water Report. This aspiration is shared by an emerging community of water data producers and publishers, including but not limited to the USGS, USEPA, CUAHSI, the Western States Water Council, the Water Data Collaborative, and various state water resources and environmental quality agencies. This initiative now has an associated Coalition, with a steering committee composed of Duke University, CUAHSI, the Water Data Collaborative, the Western States Water Council, and the Lincoln Institute of Land Policy. There are several members of the coalition outside the steering committee, including representation from academia, philanthropy, the private sector, government agencies and intergovernmental associations, and professional societies. The Center for Geospatial Solutions at the Lincoln Institute, offers technology development, educational resources, and technical and social coordination for the wider Internet of Water community.

The U.S. Geological Survey Water Mission Area is actively working on the development of the National Hydrologic Geospatial Fabric (NHGF), a signifi-

cant contributor to the Internet of Water. The NHGF is designed to establish a spatial-temporal framework to support water resources data and modeling across the United States. Currently, the best available water resources data on topics such as water availability, quality, and use for a specific feature of interest are collected, published, and in some instances, aggregated and republished by a diverse range of federal, state, tribal, local government, academic, and community-science organizations. This fragmented approach makes it extremely challenging for the general public, government, and scientific communities to locate all pertinent water data about a specific environmental feature.

To address this challenge, the Internet of Water is seeking to establish a collaborative partnership with the USGS and other researchers and technologists to develop and test metadata web publishing approaches, technologies, and communities. This collaboration aims to enable and incentivize all data producers to make their data discoverable through common internet-based spatial-temporal queries. The objective of this system, named Geoconnex, is to make a maximum amount of water information accessible via user-friendly search applications, without centralizing data governance and storage. Ideally, a user should be able to query a single web interface about a location of interest and receive enough metadata to quickly locate all places on the internet where water data about that location from all relevant organizations can be found.

The technical approach taken in this project leverages prior work, including:

1. The Second Environmental Linked Features Interoperability Experiment ([SELFIE])(<http://www.opengis.net/doc/PER/SELFIE-ER>), which conceptualized a web architecture consisting of persistent identifiers for real-world water features of interest that direct to landing pages with structured metadata that includes links to data relevant to the given feature of interest.
2. [W3C Web standards and best practices](#), including for [data on the web](#) and [spatial data on the web](#)
3. Open Geospatial Consortium (OGC) [API Standards](#) that provide specifications for interoperable data sharing and processing services.
4. The OGC [WaterML2](#) family of information models for water data
5. [science-on-schema.org](http://science-on-schema.org) guidance for publishing metadata about scientific datasets
6. Several open source software projects. In particular, contributions were made for this project to:
  - [pygeoapi](#), a server that implements OGC API Standards
  - [gleaner](#), a metadata harvester that implements W3C best practices

### 3.3 Overview

The rest of the report is organized as follows:

[Glossary](#) provides definitions for specific terms and abbreviations used throughout the report

[User Engagement](#) describes how users were identified and engaged throughout the project

[Use Cases](#) describe the general data discovery and publication use cases, and domain use cases that were developed and how they were addressed by Geoconnex implementation efforts

[Architect]

## 4 Glossary

**API** Application Programming Interface, a set of rules for how machines can exchange information

**Data Content** A document accessible by URL that presents information about an NIR.

**GeoSPARQL** An OGC standard for representing and querying geospatial data in RDF

**HTML** HyperText Markup Language, a text format for web content

**HY\_Features** [Surface Hydrologic Features Conceptual Model](#)

**JSON** JavaScript Object Notation, a data format common for web development and data transfer

**JSON-LD** JSON for Linking Data, a type of JSON designed to map JSON from different sources onto common vocabularies and data models to facilitate interoperability and automated data integration. It is a format of RDF.

**Landing Resource** A document accessible by URL that presents a default set of metadata –principally, links to Data Resources about a NIR.

**NIR** Non-Information Resource. A physical (e.g. a river) or conceptual (e.g. institution, jurisdictional area) object

**OGC** [Open Geospatial Consortium](#), an international consensus standards organization for geospatial and sensor data and data processing and sharing services

**OAFeat** [OGC API-Features](#), an OGC API Standard designed to provide vector geospatial data in a variety of formats

**PID** Persistent Identifier. An identifier that never changes for a given resource. In the Geoconnex context, refers to Geoconnex PIDs minted at the Geoconnex Persistent Identifier Registry

**pygeoapi** An open-source python server that implements OGC API standards

**RDF** Resource Description Framework, a generalized data model for knowledge graphs and cross-dataset interoperability.

**Registry** An information system that manages files containing identifiers. In the context of Geoconnex, the Geoconnex Persistent Identifier Registry at <https://geoconnex.internetofwater.dev>

**Resolver** A system that redirects URIs to URLs. In the context of Geoconnex, refers to the Geoconnex Resolver that redirects URIs that begin with <https://geoconnex.us/>

[schema.org](https://schema.org/): A vocabulary for use in structured data embedded into websites for search engine optimization and cross-website data interoperability

**SELFIE** [Second Environmental Linked Features Interoperability Experiment](#)

**SPARQL** A standard query language for RDF data.

**URI** Uniform Resource Identifier, a unique set of characters that identifies a resource. Within the geoconnex context, URIs should be HTTP URIs, structured like URLs, that identify a real-world resource (NIR), but direct via HTTP code 303 ('See Other') to a Landing Resource about the NIR

**URL** Uniform Resource Locator, or web address for any kind of web resource. Within Geoconnex, URLs are distinguished from URIs in that URLs point to or perhaps identify information resources, but not NIRs/real-world objects. URIs identify NIRs but direct to web resources that have information about NIRs.

## 5 Use Cases

General use case:

### 5.0.1

### 5.0.2 Discovering and Distinguishing Hydrologically Relevant Observed and Modeled Datasets

## 6 User Research

### 6.1 Outreach through Internet of Water network

### 6.2 Domain Meetings and Conferences

### 6.3 Interviews

## 7 Architecture and Implementation of Geoconnex Linked Data System

## 8 Governance

### 8.1 Functional Requirements Research

From November 2022 to June 2023, the team engaged Federal, state, tribal, local, and NGO data providers through a variety of channels, including personal communications, conference presentations, webinars, and Internet of Water Coalition activities to solicit advice regarding what governance structures would be appropriate to improve the participation of data providers in the Geoconnex system. This advice was synthesized into a set of key preliminary functions for a Geoconnex governance framework:

- Define and refine the identification and stewardship of essential reference features.
- Establish metadata requirements for various data types, including : time series, : discrete sample data, : summaries of remotely sensed data; : statistical and administrative records data.
- Establish location metadata requirements for key data sets, including but not limited to: : surface water monitoring locations (quantity and quality); : ground water monitoring locations; : water use and diversion locations; : hydrologic cataloging features, e.g. watershed boundaries, HUCs, aquifers, : interface with coastal data; : administrative boundaries for PWS and irrigation districts, groundwater management areas, conservation districts, census data, federal regions for EPA, USGS, USBR, USACE, NOAA, and other agencies.
- Create and oversee a data submission process. : Review system for structural bias and other questions related to diversity, equity and inclusion (DEI), in collaboration with external groups. : Encourage participation in the system.

To validate these requirements and inform a draft governance plan, survey of

technical experts and Geoconnex superusers identified through the Internet of Water Coalition, followed by a virtual convening of the IoW Coalition’s Geoconnex Working Group. This feedback was synthesized into a proposed governance plan to help guide future work to ensure that the Geoconnex system serves the needs of the Internet of Water community.

### 8.1.1 Survey Questions and Results

The following survey questions were posed to 7 experts who agreed to participate in the survey. The consensus syntheses of their responses following the convened discussion are summarized below each question, including some excerpted quotes from survey respondents.

1. **Is a governance mechanism for Geoconnex.us required?** Respondents recommended “Yes” to this question. “It is always good to have a mechanism for internal guidance and interfacing with the larger community, especially if the effort is cross-jurisdictional.”
2. **What stakeholders should be involved in the leadership of this governance mechanism?** Respondents suggested that stakeholders should include a cross-section of large, medium, and small data providers, technical experts, API users, and end users, across the full data life cycle. “Data providers (regulated and unregulated data providers, local water systems, etc.), data stewards (agencies), and data users (academia, NGO advocacy groups, public, govt).”
3. **Should the governance mechanism be voluntary and informal with links to government agencies, or should it be formalized and led by a government agency?** Respondents generally favored a voluntary and informal structure, with links to government. “Perhaps there needs to be a core governance team and some offshoots that allow nimble communication on technical matters, innovation, etc.”
4. **How large a governance mechanism is required for decision-making, and should it have an expert character or a representative character?** Respondents expressed a preference for an inclusive, but small, mechanism for the core business of setting up and defining Geoconnex. Respondents also suggested that results be reviewed by a wider audience. “If you think about the larger goal which in my mind that all water data has a Geoconnex PID then I think you need a somewhat larger governance with clear structures for how decisions are made within the system.”
5. **What mechanisms are necessary to engage a broader range of stakeholders?** Respondents suggested funding support, strong individual leadership, working examples by sector and agency, virtual webinars, and targeted outreach to specific audiences that would have an interest in certain specific reference features.



6. **What steps should be taken to ensure diversity, equity, and inclusion in this governance mechanism?** Respondents answered varied widely. Some suggested centering the user needs of a set of under-represented, over-burdened communities in the context of a larger DEI strategy and ensure representation in the IoW Coalition of organizations that represent those communities.
7. **Is a new governance mechanism(s) necessary, or can existing community governance mechanisms be used for this purpose?** If so, which ones?: Most respondents agreed that a new mechanism is needed because of the uniqueness of the system.
8. **Should the governance mechanism sunset after its work is complete, or will there be an ongoing need for system governance?** If the latter, at what time interval should the governance mechanism be reviewed? Most respondents recommended a period of 1 to 3 years for the new governance mechanism, with a review after that period.

## 8.2 Proposed Governance Plan

The following two-part framework is proposed to support Geoconnex governance:

### 8.2.1 The Geoconnex Working Group

A voluntary, informal, technical working group of experts convened by the Center for Geospatial Solutions (CGS) through the Internet of Water (IoW) Coalition under the auspices of a cooperative agreement with the U.S. Geological Survey (USGS). The working group will consist of representatives or liaisons from the non-profit, academic, and private sector to develop recommendations concerning the functional questions of Geoconnex governance over a period of three years. Public agency representatives from federal, state, local, tribal, and territorial public agencies may participate as liaisons and contribute to discussions but will not contribute to the consensus of the working group. The recommendations of the working group will be synthesized and published as a draft technical report of the Center for Geospatial Solutions at the Lincoln Institute of Land Policy, and submitted to the full IoW Coalition and other forums for review and comment. The forums invited to review the recommendations will include the Earth Science Information Partnership (ESIP), the federal roundtable on water data coordination sponsored by DOE, the CUAHSI network, and professional societies of water data users including AWRA, NAQWA, NWQMC. Following the comment period, CGS will reconcile the community comments for final publication and submission to USGS.

### **8.2.2 USGS-EPA Joint Committee on Geoconnex**

A joint committee of 3 representatives of USGS and 3 representatives of EPA Office of Water, convened and co-chaired by CGS under the auspices of its cooperative agreement with USGS. The Joint Committee will consider the recommendations of the Geoconnex Working Group and finalize decisions concerning governance of the system, to be published by the U.S. Geological Survey.