

Integrated Data Management System for Critical Zone Observatories

A pilot for EarthCube?

Anthony K. Aufdenkampe, Stroud Water Research Center

Kerstin Lehnert, IEDA/Columbia.

Ilya Zaslavsky, SDSC.

Jeff Horsburgh, USU.

Emilio Mayorga, UW-APL

David Tarboton, USU

David Lubinski, UC-Boulder



Obama issues Executive Order in support of open data

Timothy Vollmer, May 10th, 2013

Yesterday President Barack Obama issued an [Executive Order](#) requiring federal government information to be open and machine-readable by default. This Order is the latest in a series of actions going back to 2009 in support of increasing access to and transparency of government information.

In addition to the Executive Order, the White House released a [Memorandum](#) (PDF) explaining how federal government agencies will comply with the new open data policy.

This Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities. This includes using machine readable and open formats, data standards, and common core and extensible metadata for all new information creation and collection efforts. It also includes agencies ensuring information stewardship through the use of open licenses and review of information for privacy, confidentiality, security, or other restrictions to release.

It provides a forward-thinking set of guidelines for open data to be released by U.S. federal agencies:

Open data: For the purposes of this Memorandum, the term “open data” refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users. In general, open data will be consistent with the following principles:



Executive Order on Open Data

[Seal Of The Executive Office Of The President](#) / Public Domain

Integrated CZO Data: 2009-2011

\$500k supplement to Boulder Creek CZO, PI: M. Williams

- Support Data Management at 1st Three CZOs
 - \$240k total, \$80k each for Boulder, Merced & PSU
- National CZO website & data workshops
 - \$60k total, \$40k to Boulder, \$10k each Merced & PSU
- CZO Central Repository Pilot for Hydro Data
 - \$100k to SDSC
- CZO Desktop Pilot
 - \$100k to Utah State University & Idaho State Univ.

Zaslavsky et al. 2011, *Environ. Info. Management*

<http://criticalzone.org/national/publications/pub/zaslavsky-et-al-2011-the-initial-design-of-data-sharing-infrastructure-for/>

CZOData: Overall Approach

- Do not reinvent the wheel! Build on
 - CUAHSI HIS, EarthChemDB, LTER, etc
- Consistent data presentation on web
 - Metadata
 - Data values
- Central data system for data discovery
 - Harvested by SDSC (pull system)

CZOData: Principles & Policies

- Each CZO will operate and be responsible for its own local data management system for collecting, organizing, quality controlling and publishing data through its web site.
 - Different philosophy than CUAHSI HIS
 - Each CZO is master of it's own data
 - We don't care what goes on under the hood (to an extent)
 - Each site uses it's own protocols, data bases, etc.
 - Allows CZO to honor site legacy data

CZOData: Principles & Policies

- Each CZO publishes its data on the web in ascii text format with sufficient metadata so that the data can be unambiguously interpreted
- Metadata follows a proscribed format
 - Data managers just need rules to follow
- Easy to harvest by central portal
- Makes it simple at the site level so scientists comply
 - Addresses the chokepoint that is getting data/metadata from the scientists to data managers

CZO Data Publication System

CZO Data Repository & Indexing (CZO Central)

External cross-project registries

DataNet, NEON

Standard CZO Services

CZO Metadata

Ontology

Shared vocabularies

Archive

Harvester

CZO Data Products

CZO Desktop Applications

CZO Desktop

Matlab

R

Excel

ArcGIS

Modeling

CZO Web-based Data Discovery System

Standard CZO data display formats

Web site

Web site

Web site

Local CZO DB

Local CZO DB

Local CZO DB

Spatial, hydrologic, geophysical, geochemical, imagery, spectral...



Legend

Database
Encoding

XML
Schema
Encoding

Data and Metadata Transfer

<http://search.criticalzone.org/>

Catalog

Metadata
Catalog

Metadata
Transfer

Metadata
Transfer

Data
Storage

Data Server

*At each CZO
(primarily)*

Data Delivery

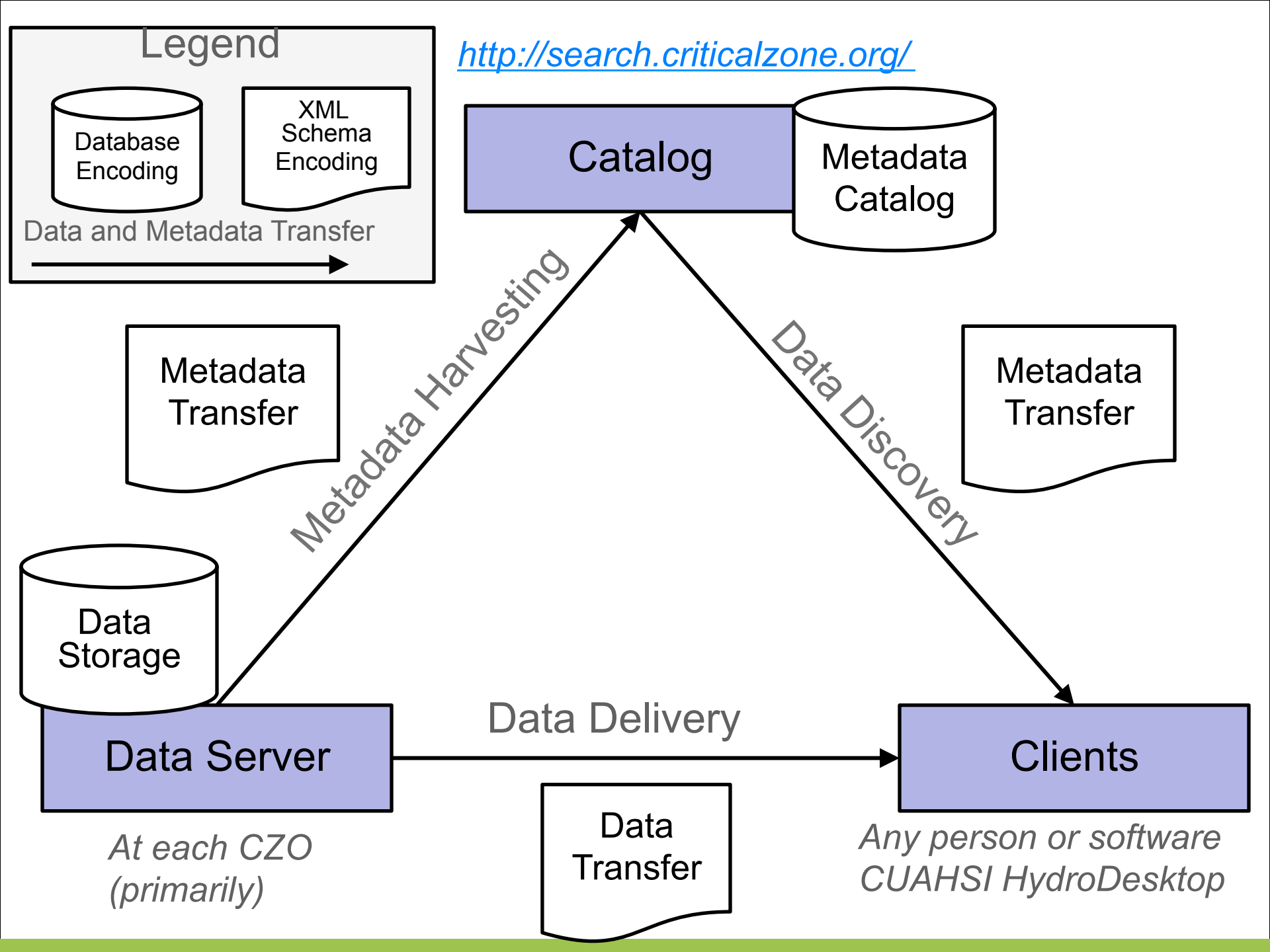
Data
Transfer

Clients

*Any person or software
CUAHSI HydroDesktop*

Metadata Harvesting

Data Discovery

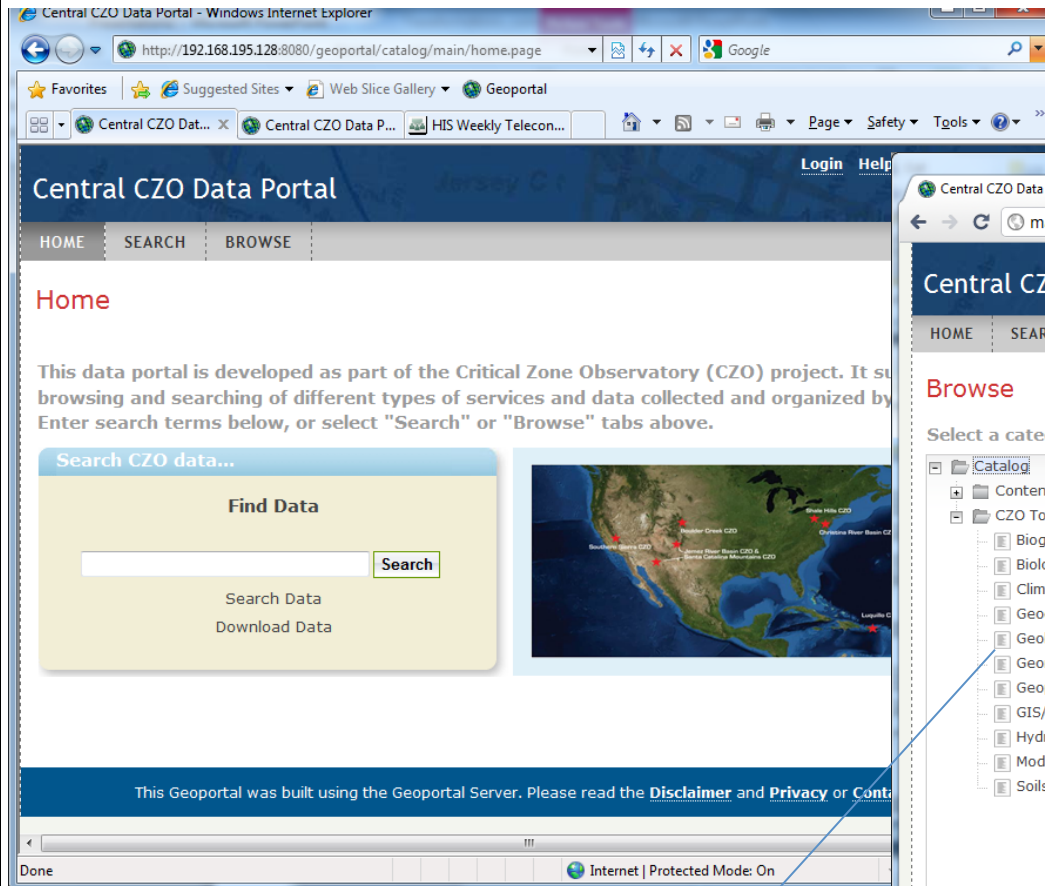


CZO Data Portal

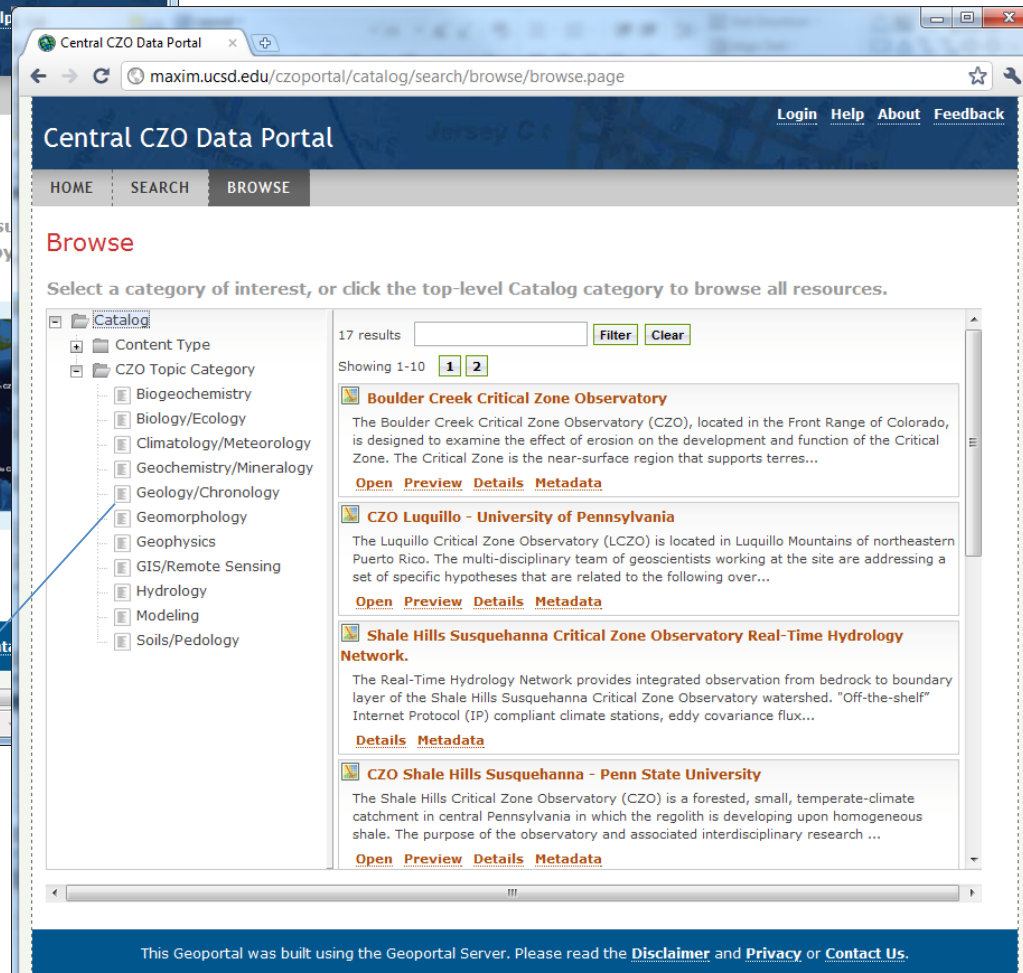
At <http://search.criticalzone.org/>

Built on Geoportal metadata catalog software:

<https://github.com/Esri/geoportal-server>



Registered data are organized by CZO thematic categories



The portal is CSW-compliant (CSW=Catalog Services for the Web): can be federated with other catalogs including *data.gov*

Supports search by location, resource type, thematic category, keywords, plus full-text abstract search

Display files from CZO web sites are registered to the data discovery portal automatically

In addition, display files of known types are expressed as data services, which are also registered in the portal

The screenshot shows the Central CZO Data Portal in a Windows Internet Explorer browser. The page title is "Central CZO Data Portal" and the URL is "http://maxim.ucsd.edu/czoportal/catalog/main/home.page". The navigation bar includes "HOME", "SEARCH", and "BROWSE". The "SEARCH" section shows a search for "geochemical" with 1 result. The result is from "CUAHSI HydroCatalog" and is titled "CZO Shale Hills Susquehanna - Penn State University". The description states: "The Shale Hills Critical Zone Observatory (CZO) is a forested, small, temperate-climate catchment in central Pennsylvania in which the regolith is developing upon homogeneous shale. The purpose of the observatory and associated interdisciplinary research ...". The "Additional Options" section includes "WHERE" with radio buttons for "Anywhere", "Intersecting", and "Fully within". A map of the Shale Hills area is shown with a red rectangle indicating the search area. A blue arrow points from the "CUAHSI HydroCatalog" link in the search results to a text box on the right.

Central CZO Data Portal

Search

geochemical

Search

Results 1-1 of 1 record(s)

Expand results [Zoom To Results](#) [Zoom To Searched Area](#)

CZO Shale Hills Susquehanna - Penn State University

The Shale Hills Critical Zone Observatory (CZO) is a forested, small, temperate-climate catchment in central Pennsylvania in which the regolith is developing upon homogeneous shale. The purpose of the observatory and associated interdisciplinary research ...

[Open](#) [Preview](#) [Details](#) [Metadata](#) [Zoom To](#)

See results through REST
API: [GEORSS](#) [ATOM](#) [HTML](#) [FRAGMENT](#) [KML](#) [JSON](#)

Records shown from: This Site
Click here to select different site or configure search.

This Site 1 Results

[CUAHSI HydroCatalog](#) 1 Results

Select sites where the search will be distributed
You may select up to 5 sites.

☒ This Site
☐ ArcGIS.com
☒ CUAHSI HydroCatalog

Additional Options
[Clear](#)

WHERE

☒ Anywhere ☐ Intersecting ☐ Fully within

Map showing the Shale Hills area with a red rectangle indicating the search area.

Federation with CUAHSI HydroCatalog, to allow search of hydrologic data from ~70 networks

CZO Display File (version 1)

Requirements:

- Human readable
- Machine parsable
- Conform to ODM 1.1 & CZO Shared Vocabulary

Solution:

- ASCII text file, with comma-separated data
- Header with detailed, structured metadata
- Separate Sites and Methods files
- <http://criticalzone.org/national/publications/pub/whitenack-et-al-2011-czo-display-file-specification/>

CZO Shared Vocabulary System

<http://sv.criticalzone.org>

Purpose:
To promote
consistent
use of
terminology.

The screenshot shows the homepage of the CZO Shared Vocabulary System. The browser window title is "CZO | SV | Home - Mozilla Firefox". The address bar shows "http://sv.criticalzone.org/". The website header includes the "National CZO Program" logo and the text "CRITICAL ZONE OBSERVATORIES" with the tagline "studying the zone where rock meets life". It also mentions "funded by NSF". A navigation menu includes links for Home, Sites, Data, Research, Publications, People, and About Us. The main content area is titled "Shared Vocabulary Registry" and contains a "Home" section with a "Sign In" link. Below this is a paragraph explaining the system's purpose and a section titled "Shared Vocabularies" which lists various fields and their uses in a table. A sidebar on the right contains sections for "Navigate This Section", "This SubSection", "Highlights", "Other National CZOs", and "Related Programs".

Shared Vocabularies


CensorCode	Used to populate the CensorCode field of the DataValues table
DataLevel	Used to populate the DataLevel field in the DataValues table
DataType	Used to populate the DataType field of the Variables table
QualifierCode	Used to populate the QualifierCode field in the DataValues table
SampleType	Used to populate the SampleType field in the Samples table
SpatialReferences	Defines the coordinate systems used in the Sites table
Units	Defines the units used in the Variables and OffsetTypes tables
ValueType	Used to populate the ValueType field in the Variables table
VariableName	Used to populate the VariableName field in the Variables table
VerticalDatum	Used to populate the VerticalDatum field in the Sites table

Builds on CUAHSI Master CV Registry, <http://his.cuahsi.org/mastercvreg/>

CZO Data Management Web Admin. Interface

CZO data managers use this web-based system to register display files, edit service metadata, initiate data retrieval, validate the data against shared vocabularies, and update hydrologic time series services

The administration system will be extended to geochemical samples and other data



The screenshot shows the CZO Data Management Web Admin. Interface. At the top, there is a header with the CZO logo (National CZO Program) and a navigation bar. The navigation bar includes links for Home, My Data Services, Add Data Service, and All Data Services. On the right side of the header, there is a user greeting "Welcome, twhitenack" and a "Log out" link. Below the navigation bar, there is a section titled "Critical Zone Observatory Central Web Service Catalog". Under this section, there are three main categories of services: POINT OBSERVATION DATA SERVICES, GEOCHEMICAL SAMPLE DATA SERVICES, and SPATIAL DATA SERVICES. The POINT OBSERVATION DATA SERVICES section includes a description of the website's purpose and a list of five steps for the data publication model. The GEOCHEMICAL SAMPLE DATA SERVICES and SPATIAL DATA SERVICES sections are currently empty.

Welcome, twhitenack
Log out

Administration

Home My Data Services Add Data Service All Data Services

Critical Zone Observatory Central Web Service Catalog

POINT OBSERVATION DATA SERVICES

This website is for use by CZO data managers. It supports the sharing of point time series data collected and published by CZO sites.

The CZO version of the data publication model includes the following steps:

1. CZO data manager places CZO display files on a web server.
2. For each CZO site, the Central repository application regularly reads the configuration display file and harvests the newly uploaded display files into a CZO database on the central server.
3. Water data web services for each CZO site are updated with the new data, and become available to client applications.
4. Optionally, CZO data managers [login](#) to CZO Central and associate the newly uploaded variable names with a [Hydrologic Concept Ontology](#) or edit other web service characteristics.
5. Datasets are subsequently registered at the [CZO portal](#) site.

GEOCHEMICAL SAMPLE DATA SERVICES

SPATIAL DATA SERVICES

- Each registered data service is published as a WFS service and registered with the [CZO portal](#)

<http://central.criticalzone.org>

Services registered by CZO data managers

Data Service Details:

[View Public Details Page](#)[CZO harvest configuration](#)

Service Title: Luquillo Critical Zone Observatory

Network Vocab: czo_luquillo

Service WSDL: http://192.31.21.100/czo_luquillo/cuahsi_1_0.asmx?WSDL

Source Info:

Organization: University of Pennsylvania

URL: http://www.sas.upenn.edu/lczo/
What organization is publishing this data?

Contact Info:

Name: Miguel Leon

Email: leonmi@sas.upenn.edu

Phone:

Who is the primary contact?

☒ Is service public?
Service must be public to be accessible through this portal.

Update **Cancel**

Citation:

How do you want your data be cited when downloaded?

Abstract:

The Luquillo Critical Zone Observatory (LCZO) is located in the northeastern Puerto Rico. A multidisciplinary team of scientists at the site are addressing specific hypotheses

Editable service definitions and management interface for each CZO data service

[Request Data Harvest](#)

Send a message to the site administrator requesting approval or an additional harvest. (If you just submitted a new service, a message

[Change Image](#)

Upload a custom image and max icon

Data managers control how their data is annotated.

Ingesting of Display files is triggered on the server by the Data manager.

Welcome, twhtenack
Log out

Administration

Home My Data Services Add Data Service All Data Services

CZO Custom Config File URL:

https://oni.sas.upenn.edu/lczo-test/sites/oni.sas.upenn.edu/lczo-test/files/config.txt

Email: leonmi@sas.upenn.edu
Receive logs via email

Schedule: Daily

[Edit](#)

[Ingest Data](#)

[Harvest Service Metadata](#)

Ingest Logs

Logfile	started	finished
viewlog	3/24/2011 4:23:09 AM	3/24/2011 5:03:12 AM
viewlog	3/23/2011 7:04:56 AM	3/23/2011 9:28:57 AM
viewlog	3/22/2011 3:25:58 PM	
viewlog	3/22/2011 3:23:11 PM	
viewlog	3/22/2011 3:20:27 PM	
viewlog	3/22/2011 3:13:16 PM	
viewlog	3/22/2011 3:11:24 PM	
viewlog	3/22/2011 3:08:50 PM	
viewlog	3/22/2011 3:07:42 PM	
viewlog	3/18/2011 6:16:23 AM	

1 2 3 4 5 6 7 8 9 10 ...

Display file ingestion log

Data Publication Process

(for hydrologic time series)

\doc

TITLE. Streamflow data for Jemez River Basin.

ABSTRACT. 30 minute streamflow data measured and computed for National Preserve).

INVESTIGATOR. Peter Troch, Department of Hydrology and Water Tucson, ptroch@arizona.edu
CZOL: www.czo.gov

VARIABLE NAMES. Streamflow.

KEYWORDS. hydrology, streamflow, Valle Caldera, New Mexico.

CITATION. [1] Broxton P.D., Troch P.A., and Lyon S.W. (2009) small mountainous catchments. Water Resources Research 45: W0
[2] The following acknowledgment should accompany any public data were provided by the NSF-supported Jemez River Basin and

CZO Display File

Raw Display file metadata
Is registered with the CZO
data portal, to assure original
data is discoverable and
downloadable.



WaterML
Service



Catalog
Search
Service

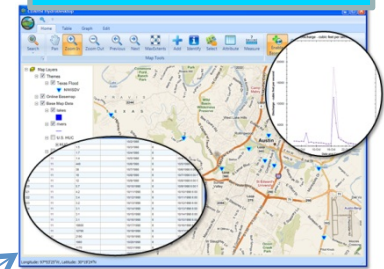
OGC WFS
Service



CZO Portal utilizes the OGC CSW
(catalog services for the web)

OGC CSW
Service

CZO Desktop



WFS Service Is
registered with the
CZO data portal

Broader internet
community
accessing data
using standard
protocols.

Legend

Database
Encoding

XML
Schema
Encoding

Data and Metadata Transfer

<http://search.criticalzone.org/>

Catalog

Metadata
Catalog

Metadata
Transfer

Metadata
Transfer

Data
Storage

Data Server

At each CZO

Data Delivery

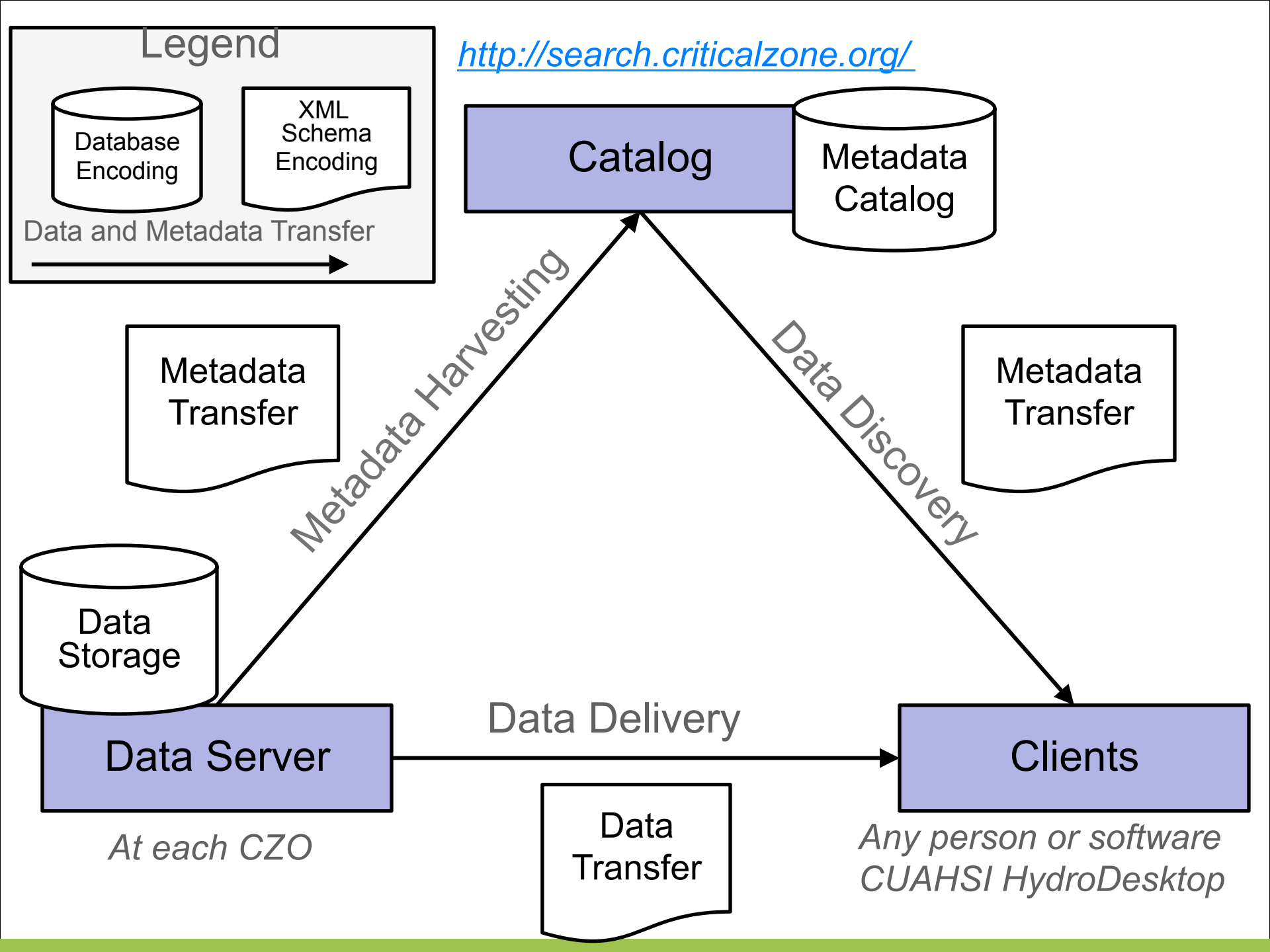
Data
Transfer

Clients

*Any person or software
CUAHSI HydroDesktop*

Metadata Harvesting

Data Discovery

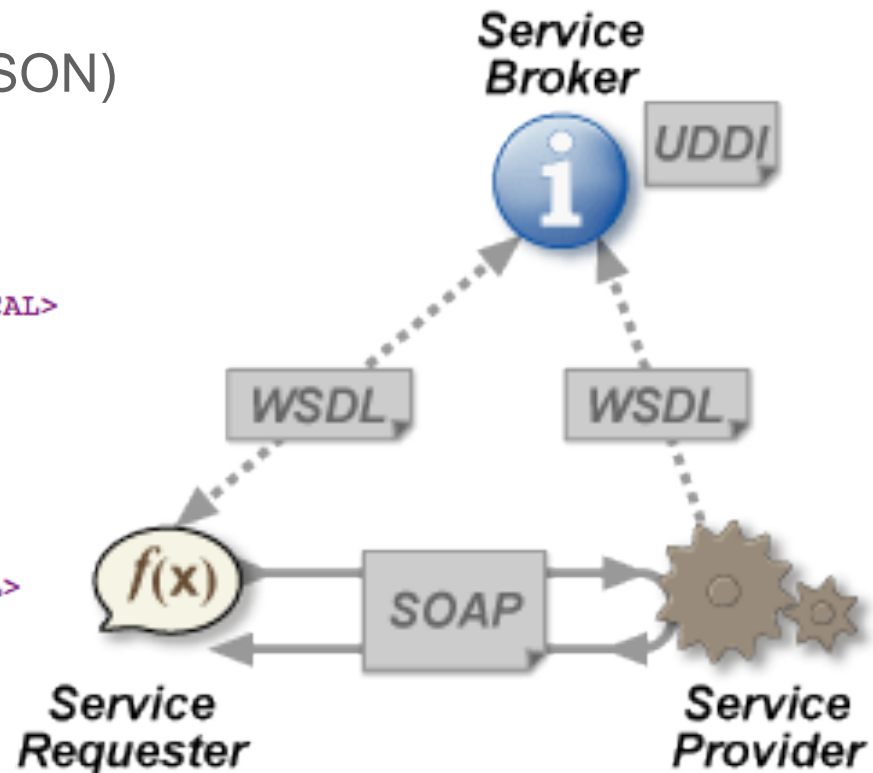


Web Services

a method of communication between two electronic devices over the web

- Service Requester/Consumer (clients) request information from Service Providers (servers) using standard operations like GET, POST, PUT, DELETE for HTTP
- Data is exchange via XML (or JSON)

```
<!-- Edited by XMLSpy® -->
▼<CATALOG>
  ▼<PLANT>
    <COMMON>Bloodroot</COMMON>
    <BOTANICAL>Sanguinaria canadensis</BOTANICAL>
    <ZONE>4</ZONE>
    <LIGHT>Mostly Shady</LIGHT>
    <PRICE>$2.44</PRICE>
    <AVAILABILITY>031599</AVAILABILITY>
  </PLANT>
  ▼<PLANT>
    <COMMON>Columbine</COMMON>
    <BOTANICAL>Aquilegia canadensis</BOTANICAL>
    <ZONE>3</ZONE>
    <LIGHT>Mostly Shady</LIGHT>
    <PRICE>$9.37</PRICE>
    <AVAILABILITY>030699</AVAILABILITY>
  </PLANT>
```



WaterOneFlow and WaterML 1.1

- WaterOneFlow = web service interface for publishing hydrologic point time series
- WaterML = XML schema used by WaterOneFlow to encode data stored in an ODM database for transport over the Internet

WaterML 2.0

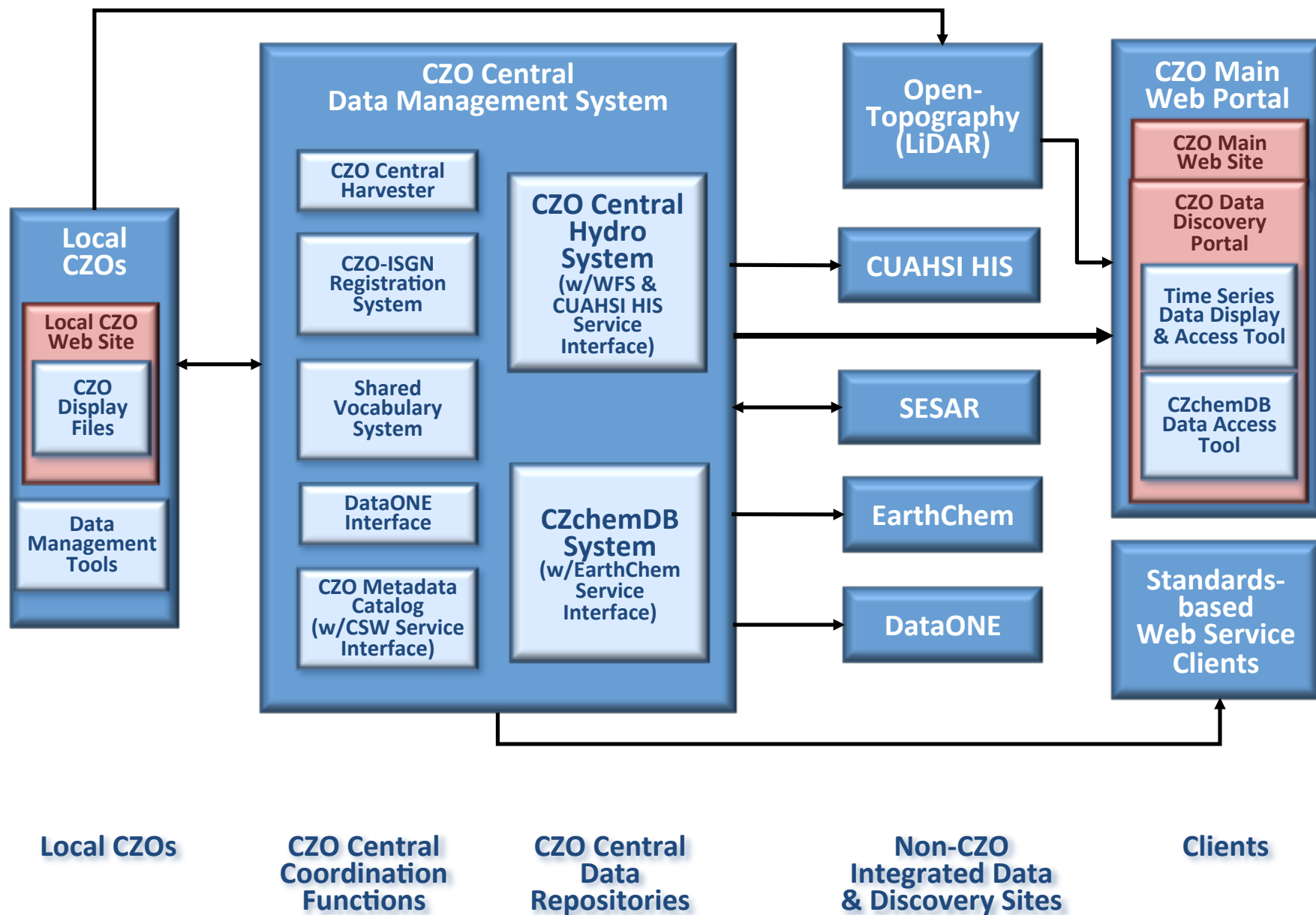
- An international metadata standard governed by the Open Geospatial Consortium (OGC)
 - To capture the semantics of hydrological observational data for data exchange
- An XML schema as a profile of OGC Observations and Measurements (O&M)
 - Part 1 – specifically focused on hydrologic time series
 - Part 2 – ratings, gaugings, and sections
 - Part 3 – water quality samples?????

Integrated CZO Data: 2012-2014

\$1.5M starting Dec. 1, 2012, PI: Aufdenkampe

- **Community:** extensive and iterative interaction and feedback from CZO PIs, scientists and data managers
- **Web:** uniform web portal appearance for the CZO sites and the national CZO program
- **Metadata:** development of a consistent metadata strategy for CZO data, ensuring that the data descriptions follow consistent semantics
- **Workflows:** enhancing publication and data discovery workflows
 - collection of data submission forms and tools
 - integrating with the EarthChem system
- **Discovery:** creating a uniform data discovery portal
- **Visualization:** developing a consistent online data visualization interface for CZO time series data

CZOData II Architecture



CZOData Challenges

1. Community Engagement

CZOData solution:

- Information Management Committee (IMC)
 - 1-2 investigators/CZO + site data managers
 - Weekly to monthly web conference meetings
 - Annual face-to-face meeting
 - Feedback to CZOData team
 - Data use scenarios
 - Meta-data requirements, shared vocabulary, etc.
- Workshops & Webinars to greater CZO community

CZO Data Challenges

1. Community Engagement
2. Web: Get small “long tail” data out of the dark!

CZOData solution:

- Leverage website to provide uniform data presentation & publication to CZO Central
 - Document repository (library)
 - Central database

[Dataset](#)

Christina River Basin - Stream Suspended Sediment (1993-2012)

TSS concentrations and elemental/isotopic composition (starting 2005) of baseflow and stormflow.


Variables: Solids_ total suspended, carbon to nitrogen molar ratio, carbon_ particulate organic, nitrogen_ particulate organic, nitrogen-15 stable isotope ratio delta

Date Range: (1993-2012)

Dataset Creators/Authors: Aufdenkampe, A.K.; Newbold, J.D.; Anderson, B. A.; Richardson, D.; Damiano, S.G.

Contact: Sara Geleskie Damiano, Stroud Water Research Center, 970 Spencer Road, Avondale, PA 19311, sgeleskie@stroudcenter.org

Field Area: [Boulton Run](#) | [Christina River Basin](#) | [Forest Endmember: Spring Brook](#) | [White Clay Creek @ SWRC](#) | [Construction Endmember: White Clay Creek below landfill](#) | [Lower White Clay Creek](#) | [Agricultural Endmember: South Branch Doe Run](#)

 [Water Chemistry](#)
 [Geomorphology](#)
 [Biogeochemistry](#)
 [Hydrology](#)
 [Geochemistry / Mineralogy](#)
[Christina](#)
[Description](#)
[Keywords](#)
[Citation](#)
[Publications](#)
[Acknowledgements](#)

Description

Total suspended solids (TSS) and Volatile Suspended Solids (VSS) from White Clay Creek near the Stroud Water Research Center, Avondale, PA, USA. The purpose is to quantify export of inorganic and organic particulate matter from the 725-hectare watershed. Samples consist of those taken at monthly intervals, normally the first Wednesday of each month regardless of weather or flow conditions and those taken after precipitation events. The monthly samples are manual grab samples collected in 5-L polyethylene "space saver" bottles from a few centimeters below the surface and without disturbance to the stream bed. The event samples were collected in response to precipitation of 20 mm or more using an ISCO automated sampler which collected 1-L samples in polyethylene bottles at hourly intervals through an intake approximately 20 cm above the bed. Each of approximately four events per year are represented by approximately 10 samples selected from the hourly series to characterize the rise, peak, and falling limb of the hydrograph. Additional events are represented by the three samples nearest peak flow.

Data

 [Christina River Basin - Stream Suspended Sediment 1993](#)

(.csv) [Data Level 1](#), [Metadata](#)

CZO Field Areas


[Boulton Run](#)

[Christina River Basin](#)

[Forest Endmember: Spring Brook](#)


CZO Data Challenges

1. Community Engagement
2. Web: Get small “long tail” data out of the dark!
3. Metadata: Integrate data from all CZ disciplines

CZOData solutions:

- Strong integration with IEDA and SESAR
- Observations Data Model 2.0 (ODM2)
 - Separate NSF Geoinformatics project,
PI: J. Horsburgh

System for Earth Sample Registration

- Facilitate discovery of samples
- Ensure unique identification with International GeoSample Number (IGSN)
- Preserve sample metadata



Search An IGSN

Look Up

HOME

ABOUT THE IGSN

SERVICES

SAMPLES

NEWS

HELP

ABOUT US

LOG IN TO MYSESAR



get your igsn

Register your samples with SESAR to obtain IGSNs for unique sample identification.

search the catalog

Search the SESAR catalog to find registered samples and their current location.

curate your samples

Learn how SESAR can help you manage your samples and collections.

interoperability

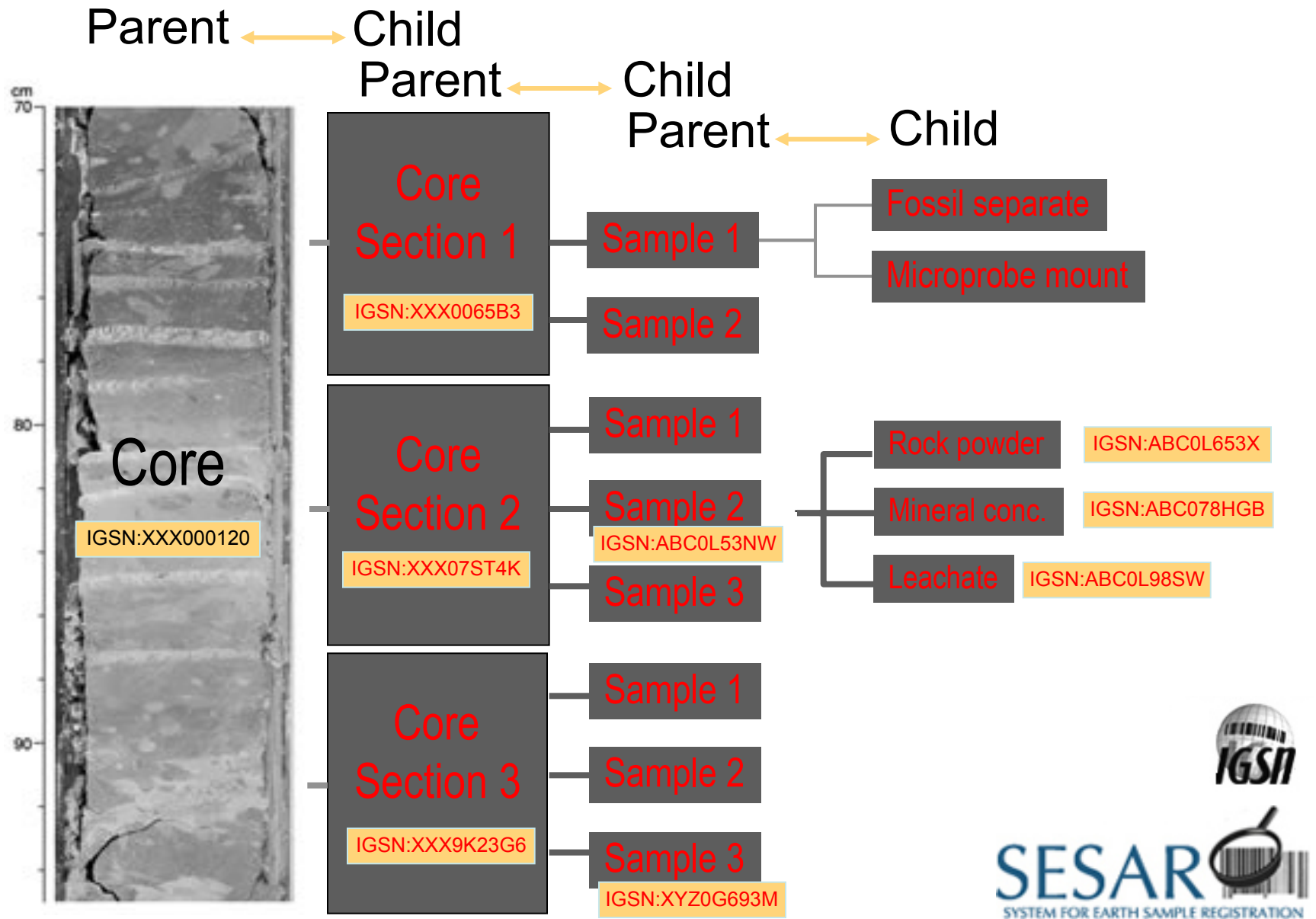
Access IGSN metadata profiles and register samples via web services.

new user?

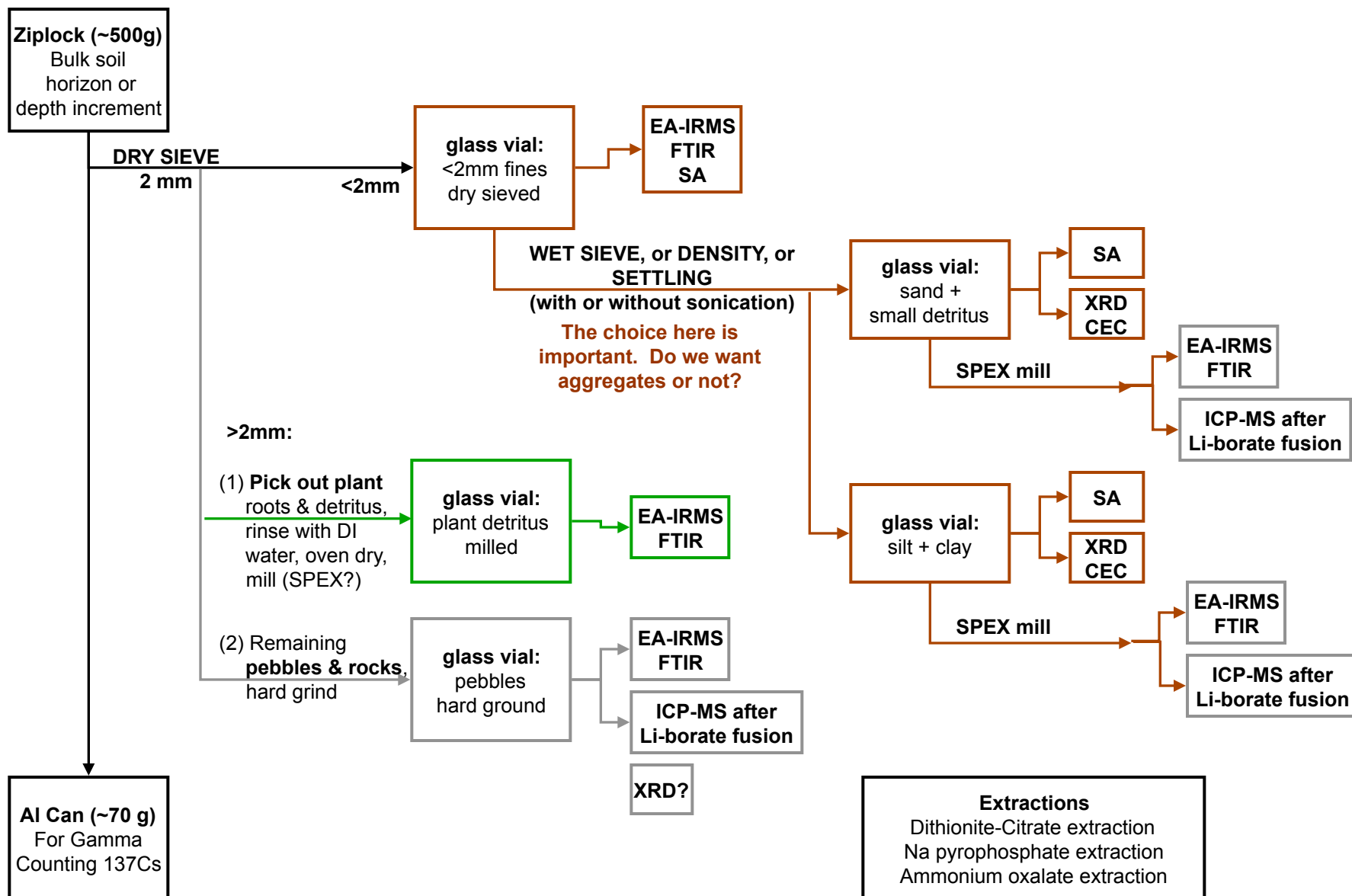
Get a MySESAR account to register your samples.

www.geosamples.org

Sample Management for Geochemistry



Sample Fractions for Soil Geochemistry



Observation Data Model 2.0

- NSF funded project: PI. Jeff Horsburgh
 - “Developing a Community Information Model and Supporting Software to Extend Interoperability of Sensor and Sample Based Earth Observations”
 - To achieve interoperability between IEDA, EarthCHEM, CUAHSI HIS, and other data systems
 - Better support for samples and unique identifiers (IGSN/SESAR)
 - Extensibility to table attributes
 - Better annotation and provenance
 - Enable integrated web service based publication of a broader class of CZO data

Data Management Concepts

- Flat files → relational databases
 - Linking tables
 - Reducing redundancy
 - Why a “data model” matters
- Meta-data “standards” via
 - Controlled Vocabulary (or Shared Vocab)
 - Choice of fields in data model
- Web Services
 - Clients-Server exchanges via XML
- Data discovery

Flat files → relational databases

Field,
Attribute



Record →

DateTime	Station	Latitude	Longitude	Sensor	StreamStage	Units
1/23/12 12:00	WCC-PumpHouse	39.86066	-75.783855	Solinist	23	cm
1/23/12 12:15	WCC-PumpHouse	39.86066	-75.783855	Solinist	24	cm
1/23/12 12:30	WCC-PumpHouse	39.86066	-75.783855	Solinist	25	cm
1/23/12 12:45	WCC-PumpHouse	39.86066	-75.783855	Solinist	26	cm
1/23/12 13:00	WCC-PumpHouse	39.86066	-75.783855	Solinist	27	cm
1/23/12 13:15	WCC-PumpHouse	39.86066	-75.783855	Solinist	28	cm
1/23/12 13:30	WCC-PumpHouse	39.86066	-75.783855	Solinist	29	cm
1/23/12 13:45	WCC-PumpHouse	39.86066	-75.783855	Solinist	30	cm
1/23/12 14:00	WCC-PumpHouse	39.86066	-75.783855	Solinist	31	cm

Flat files → relational databases

Foreign
Key



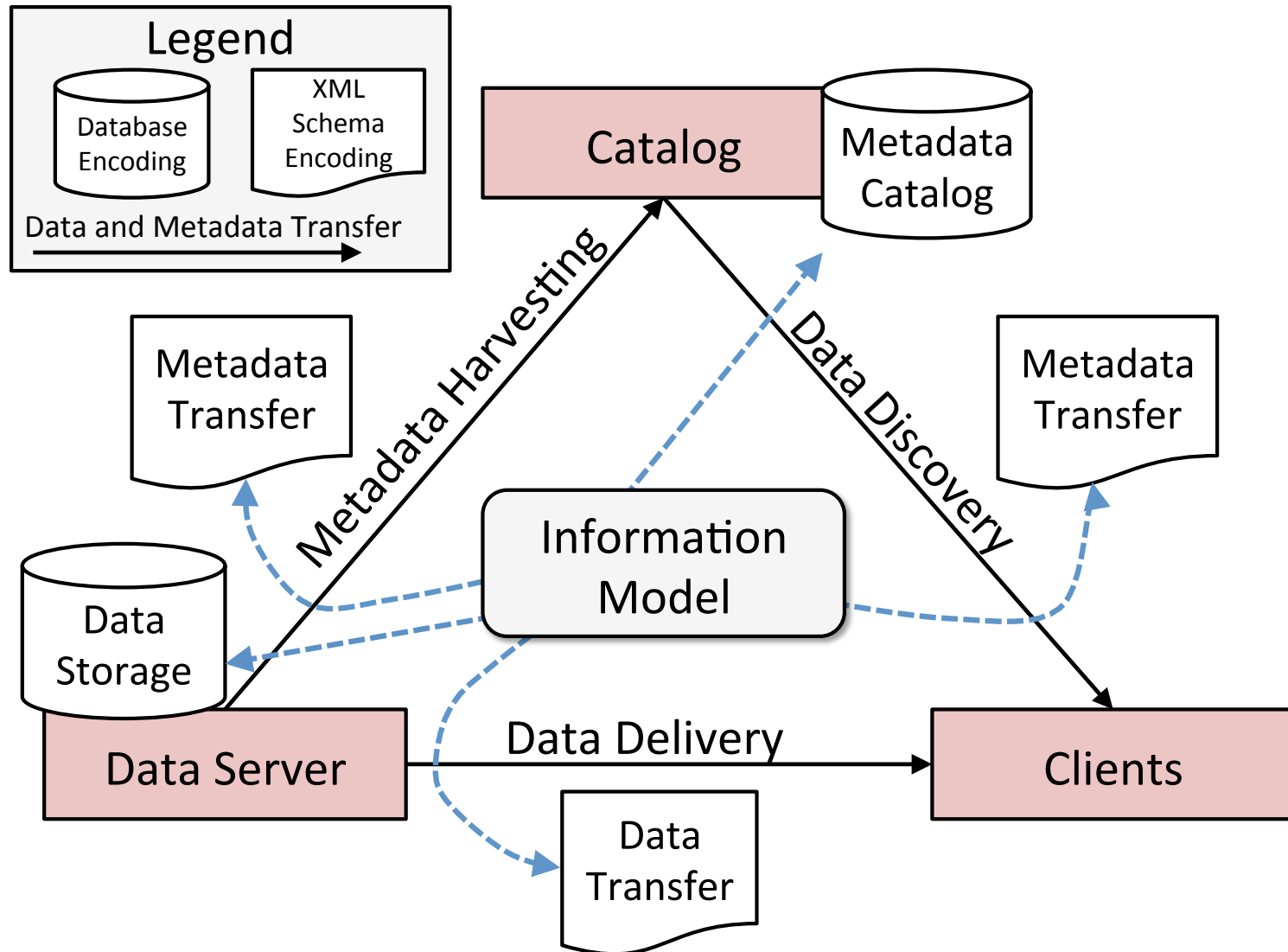
DateTime	StationID (FK)	Sensor	StreamStage	Units (CV)
1/23/12 12:00	1	Solinist	23	cm
1/23/12 12:15	1	Solinist	24	cm
1/23/12 12:30	1	Solinist	25	cm
1/23/12 12:45	1	Solinist	26	cm
1/23/12 13:00	1	Solinist	27	cm
1/23/12 13:15	1	Solinist	28	cm
1/23/12 13:30	1	Solinist	29	cm
1/23/12 13:45	1	Solinist	30	cm
1/23/12 14:00	1	Solinist	31	cm

Primary
Key

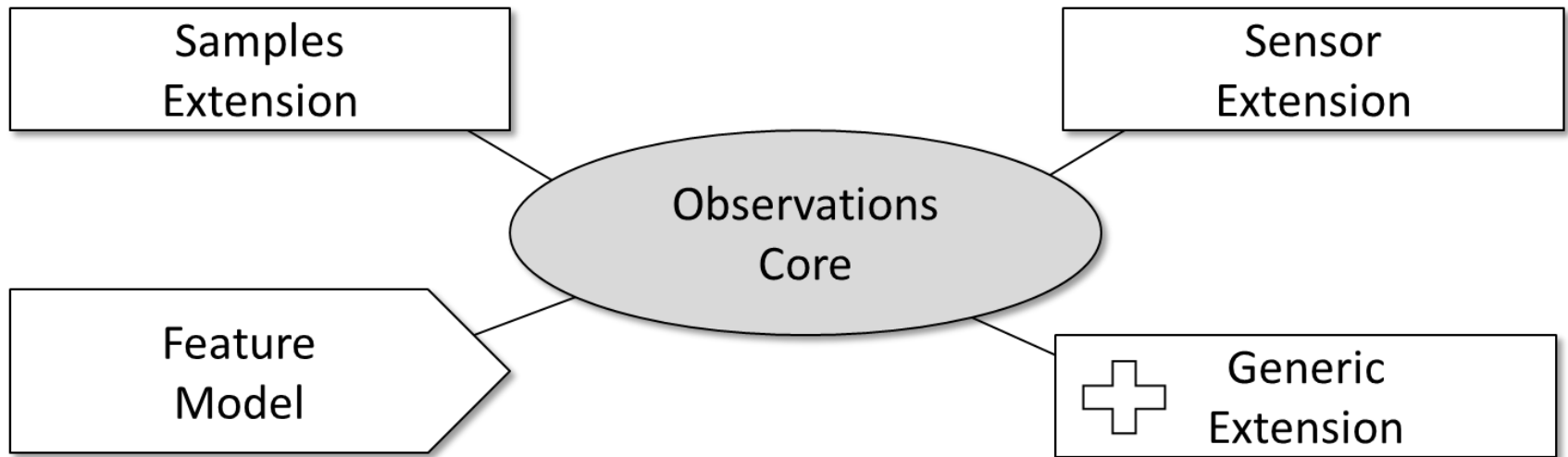


StationID (PK)	StationName	Latitude	Longitude
1	WCC-PumpHouse	39.86066	-75.783855
2	WCCLab	39.8594	-75.78381
3			
4			

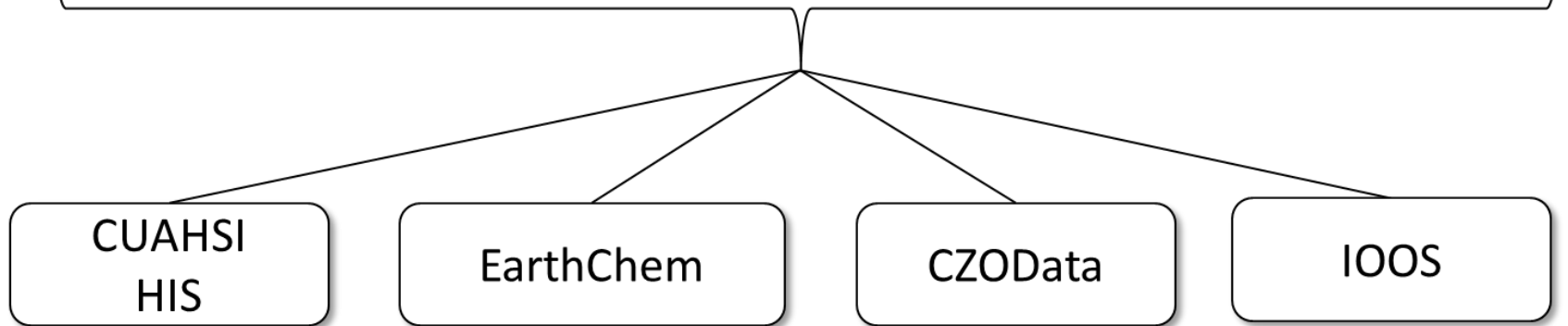
ODM2: Common to All Components



ODM2: Common to Most Data Types



Common Semantics for Earth Observations



Domain Cyberinfrastructures

ODM2: Core Schema

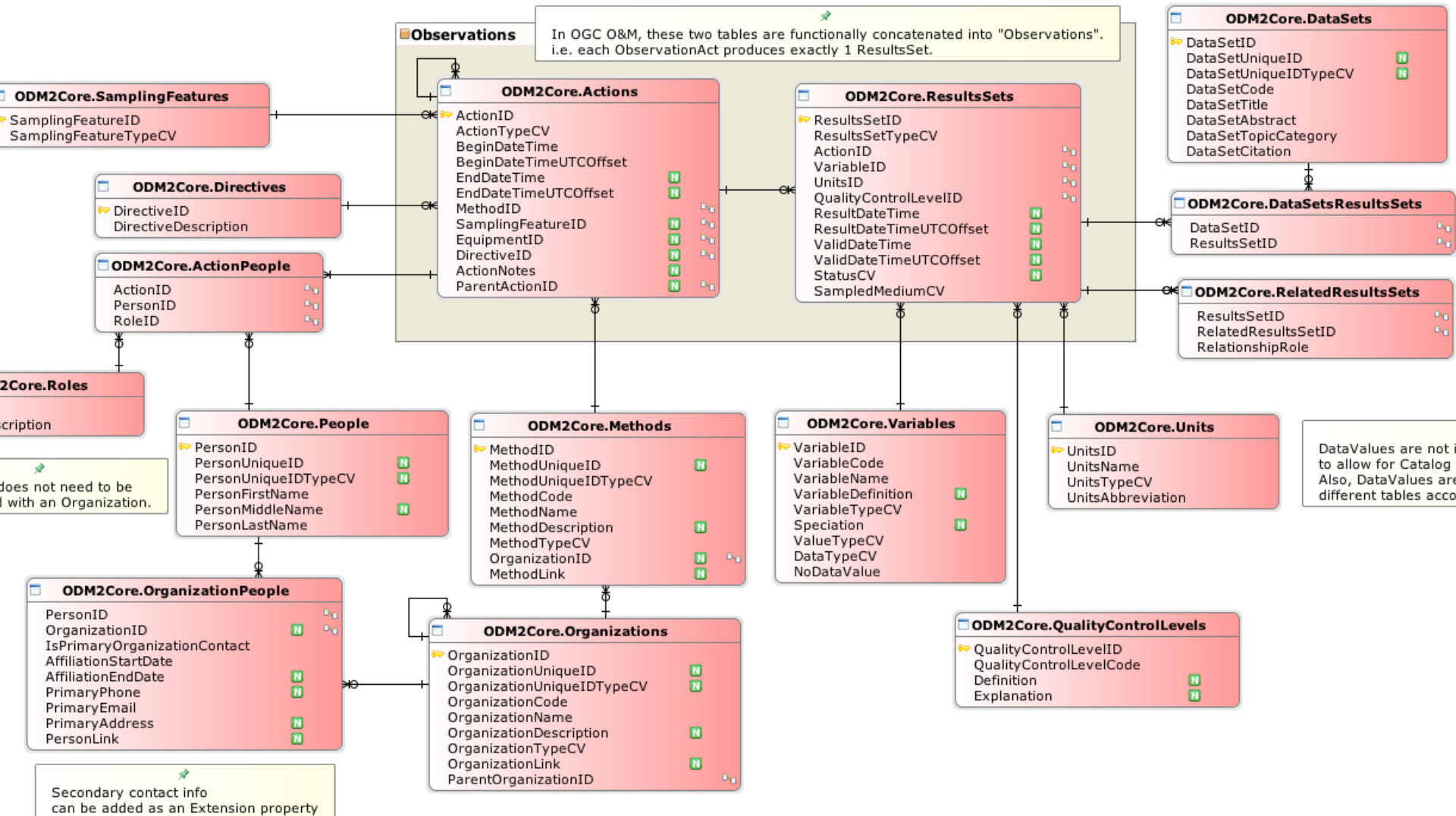
In the barebones implementation: Actions are all ObservationActs on a SamplingFeature by at least one Person using a Method producing at least one ResultsSet.

An ObservationAct is the only ActionType that can produce a ResultsSet.

A ResultsSet describes metadata for one or more DataValues that are stored in Results tables specific to the ResultsSetType

Reasons for separating ObservationActs from ResultsSets:

1. Allows Results for many Variables from each ObservationAct
2. Allows for other ActionTypes



CZO Data Challenges

1. Community Engagement
2. Web: Get small “long tail” data out of the dark!
3. Metadata: Integrate data from all CZ disciplines
4. Workflows: Enhance data management practice

CZOData solutions:

- CZO Display File version 2
- ODM2 extensions for data management
developing with BiG CZ SSI Toolbox
 - Sensor/Instrument tracking
 - Sample tracking

CZO Data Challenges

1. Community Engagement
2. Web: Get small “long tail” data out of the dark!
3. Metadata: Integrate data from all CZ disciplines
4. Workflows: Enhance data management practice
5. Web-based data discovery & visualization client

CZOData solutions:

- Adapting NANOOS Visualization System (NVS) for CZOs
- See <http://nvs.nanoos.org/Explorer>

CHRISTINA RIVER BASIN
NVS

v2.5

Assets

Chart

List

Help

Map

Map Layers

Filters

Assets

Regions

Settings

Legend

Assets

- NOS Reedy Point
- NOS Ship John Shoal
- USGS 01482800

Land Station

- DEOS DAGF
- DEOS DBKB
- DEOS DBUK1
- DEOS DCHI
- DEOS DCLY
- DEOS DDMV
- DEOS DFHM
- DEOS DGLW
- DEOS DGRN
- DEOS DHOC
- DEOS DLWG
- DEOS DMTG
- DEOS DPRC
- DEOS DSCR
- DEOS DSMY
- DEOS DTLY
- DEOS DTOW
- DEOS DWCC
- USCRN Avondale

River Gage

- SWRC Half-way Run
- SWRC Mine Hill Run
- SWRC Morris Run
- USGS 01463500
- USGS 01467200
- USGS 01474500

Brandywine Creek

Observations

Details

Lat 40.0166, Lon: -75.8050

Terrain

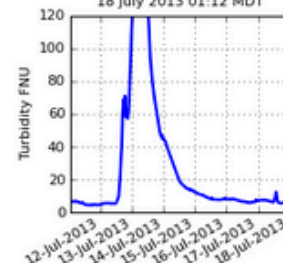
USGS 01481500, Brandywine Creek at Wilmington

Website

Location: Brandywine Creek, New Castle Co., Delaware Lat: 39.7694 Lon: -75.5733

Provider: USGS Data Source: CUAHSI-HIS

Data Updated: 17 Jul 2013 23:15 MDT

USGS 01481500 - Turbidity - 7 Days
18 July 2013 01:12 MDT

24 Hours 7 Days 30 Days



Link

Conductivity (-2ft):	0.3 mS/cm
Discharge (0ft):	4.4 k ft ³ /s
Oxygen Conc. (-2ft):	8 mg/L
Oxygen Pct. Sat. (-2ft):	100 %
pH (-2ft):	7.8
River Stage Height (0ft):	12.2 ft
Turbidity (-2ft):	6.3 FNU
Water Temp. (-2ft):	80.6 °F

White Clay Creek

NSF Scientific Software Integration

BiG CZ SSI project (Dec. 2013 to Nov. 2015):

The community-driven BiG CZ software system for integration and analysis of bio- and geoscience data in the critical zone

- **Community Engagement in Software Design**
through co-design, training & testing workshops.
- **BiG CZ Portal web application**
for high-performance map-based discovery, visualization, access & publication of data on critical zone structure & function
- **BiG CZ Toolbox**
to enable cyber-savvy CZ scientists & data managers to manage and publish the data they produce through a single scientist-focused toolkit
- **BiG CZ Central software stack**
to bridge data systems developed for multiple critical zone domains

BiG CZ Portal

- A web-based integration and visualization environment for joint analysis of cross-scale bio and geoscience processes in the critical zone, spanning experimental and observational designs
 - intuitive, high-performance map-based discovery, visualization, access and publication of heterogeneous datasets, including:
 - Points with sensor & sample observations;
 - 2D satellite and GIS imagery from many different agencies and sources;
 - 2D and 3D interactive visualization of datasets.
 - Data publication/registration forms, to easily “shed light on dark data”

BiG CZ Toolbox

- A single, easy-to-install and easy-to-teach cross-platform package with components to facilitate local data management workflows and analysis, and publication to data repositories.
 - A ready-to-use relational database for many data types
 - Web service interfaces for both accessing and sharing near real-time or historical data
 - Streaming sensor middleware and other software for managing, visualizing and QAing data from sensor networks
 - A publication coordination tool to maintain common identifiers, spatial-temporal registration and controlled vocabularies when publishing to repositories.

BiG CZ Central

- Cloud software stack to bridge data systems developed for multiple critical zone domains into a single metadata catalog
 - Central catalog for single search of multiple repositories, with focus on bridging bio & geo.
 - Unified web services API to access multiple repositories using a variety of standards/protocols

Thank You