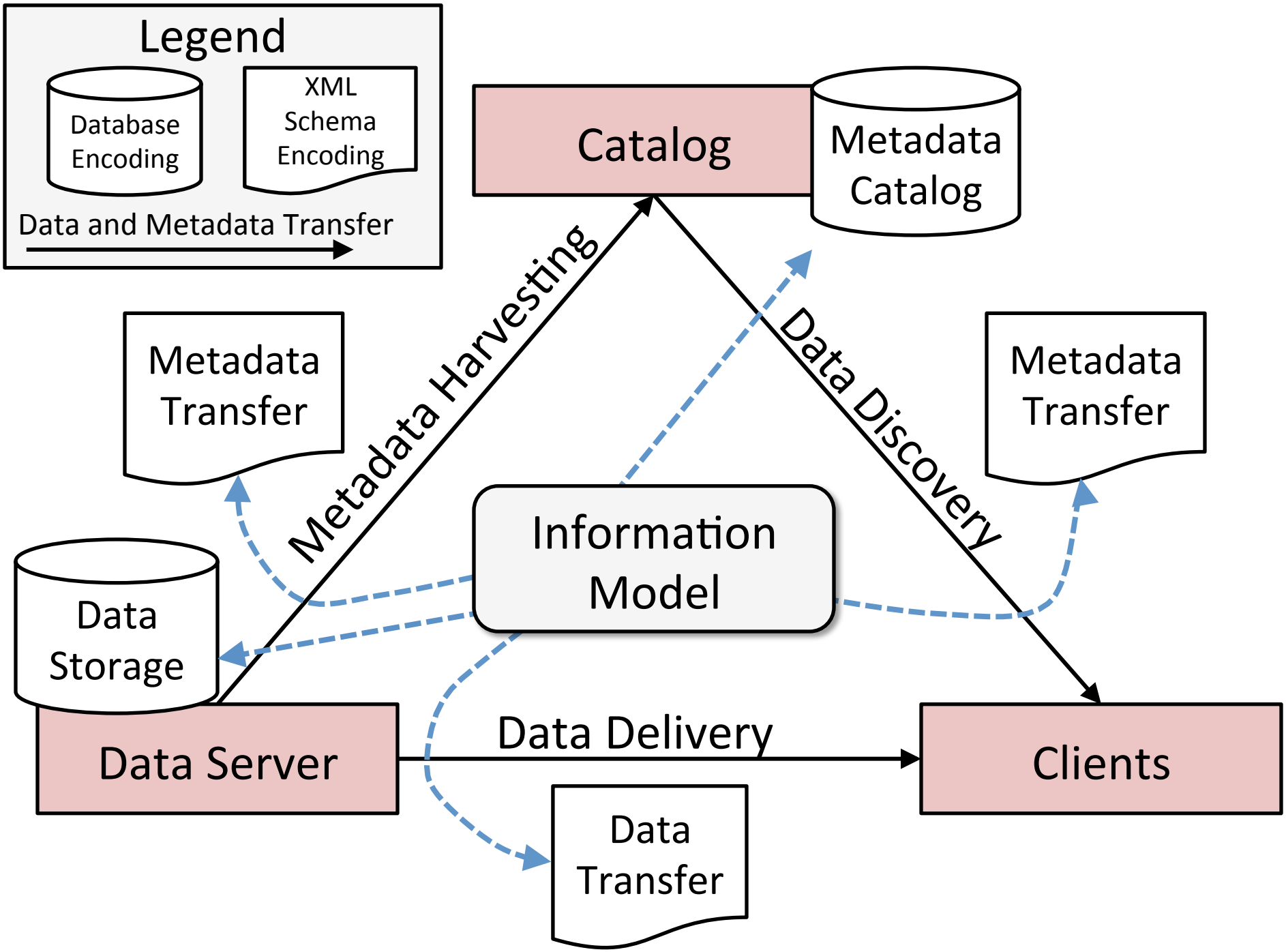
ODM2: Developing a Community Information Model and Supporting Software to Extend Interoperability of Sensor and Sample Based Earth Observations

Jeffery S. Horsburgh

Anthony Aufdenkampe, Kerstin Lehnert, Emilio Mayorga, David Tarboton, Ilya Zaslavsky, David Valentine

Support:
EAR 1224638

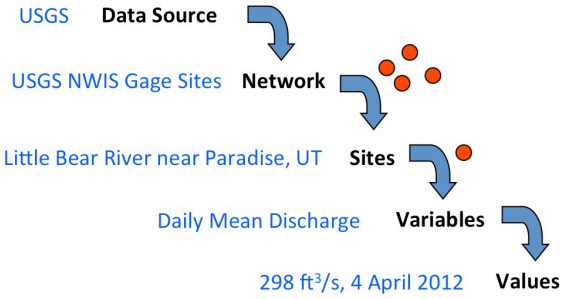

Legend
Database Encoding
XML Schema Encoding
Data and Metadata Transfer
Catalog
Metadata Catalog
Metadata Harvesting
Data Discovery
Metadata Transfer
Metadata Transfer
Information Model
Data Storage
Data Server
Data Delivery
Clients
Data Transfer

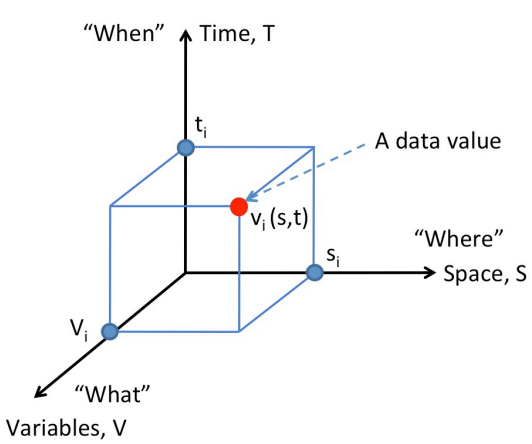
What is an "Information Model?"

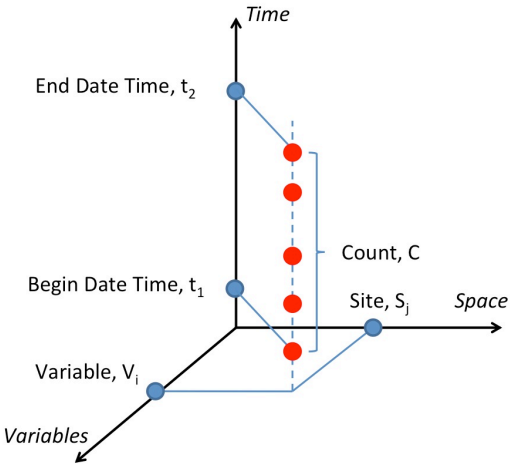
- Representation of concepts, relationships, constraints, rules, and operations that specify the semantics of data for a domain of discourse
  - Defines the domain's entity types and their attributes, relationships, and allowed operations on the entities
  - In a relational database implementation, entities become tables and their attributes become table columns
- Sharable, stable, and organized structure of information requirements for a domain context
  - Without constraining how that description is mapped to an actual implementation in software
  - There may be many physical implementations – e.g., relational database, XML schema, etc.

- Critically important to the effectiveness and interoperability of the cyberinfrastructure

Example Information Model

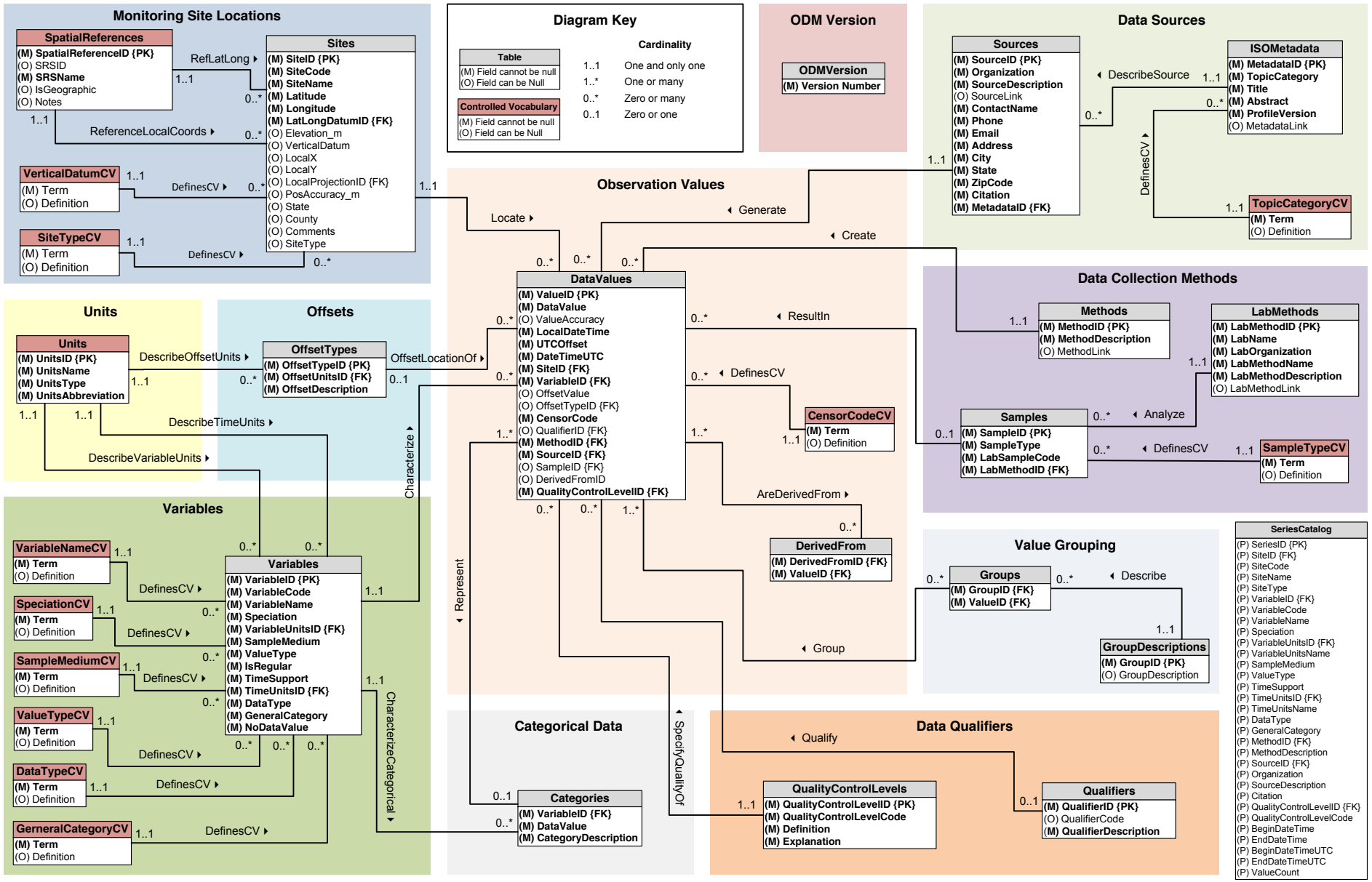






- An organization operates a network of monitoring sites
- At each monitoring site a number of variables are measured
- For each variable there is a time series of data values
- Each data series is made up of individual, time-indexed values that are each characterized by location (where the observation was collected), time (when the observation was collected), and variable (what the observation represents)

ODM 1.1.1


Monitoring Site Locations
SpatialReferences
(M) SpatialReferenceID {PK}
(O) SRSID
(M) SRSName
(O) IsGeographic
(O) Notes
RefLatLong
1..1
0..*
Sites
(M) SiteID {PK}
(M) SiteCode
(M) SiteName
(M) Latitude
(M) Longitude
(M) LatLongDatumID {FK}
(O) Elevation_m
(O) VerticalDatum
(O) LocalX
(O) LocalY
(O) LocalProjectionID {FK}
(O) PosAccuracy_m
(O) State
(O) County
(O) Comments
(O) SiteType
ReferenceLocalCoords
VerticalDatumCV
(M) Term
(O) Definition
DefinesCV
SiteTypeCV
(M) Term
(O) Definition
Diagram Key
Table
(M) Field cannot be null
(O) Field can be Null
Controlled Vocabulary
(M) Field cannot be null
(O) Field can be Null
Cardinality
1..1 One and only one
1..* One or many
0..* Zero or many
0..1 Zero or one
ODM Version
ODMVersion
(M) Version Number
Data Sources
Sources
(M) SourceID {PK}
(M) Organization
(M) SourceDescription
(O) SourceLink
(M) ContactName
(M) Phone
(M) Email
(M) Address
(M) City
(M) State
(M) ZipCode
(M) Citation
(M) MetadataID {FK}
DescribeSource
ISOMetadata
(M) MetadataID {PK}
(M) TopicCategory
(M) Title
(M) Abstract
(M) ProfileVersion
(O) MetadataLink
DefinesCV
TopicCategoryCV
(M) Term
(O) Definition
Observation Values
Locate
Generate
Create
DataValues
(M) ValueID {PK}
(M) DataValue
(O) ValueAccuracy
(M) LocalDateTime
(M) UTCOffset
(M) DateTimeUTC
(M) SiteID {FK}
(M) VariableID {FK}
(O) OffsetValue
(O) OffsetTypeID {FK}
(M) CensorCode
(O) QualifierID {FK}
(M) MethodID {FK}
(M) SourceID {FK}
(O) SampleID {FK}
(O) DerivedFromID
(M) QualityControlLevelID {FK}
ResultIn
DefinesCV
CensorCodeCV
(M) Term
(O) Definition
AreDerivedFrom
DerivedFrom
(M) DerivedFromID {FK}
(M) ValueID {FK}
Represent
Group
Data Collection Methods
Methods
(M) MethodID {PK}
(M) MethodDescription
(O) MethodLink
LabMethods
(M) LabMethodID {PK}
(M) LabName
(M) LabOrganization
(M) LabMethodName
(M) LabMethodDescription
(O) LabMethodLink
Samples
(M) SampleID {PK}
(M) SampleType
(M) LabSampleCode
(M) LabMethodID {FK}
Analyze
DefinesCV
SampleTypeCV
(M) Term
(O) Definition
Value Grouping
Groups
(M) GroupID {FK}
(M) ValueID {FK}
Describe
GroupDescriptions
(M) GroupID {PK}
(O) GroupDescription
Units
Units
(M) UnitsID {PK}
(M) UnitsName
(M) UnitsType
(M) UnitsAbbreviation
DescribeOffsetUnits
DescribeTimeUnits
DescribeVariableUnits
Offsets
OffsetTypes
(M) OffsetTypeID {PK}
(M) OffsetUnitsID {FK}
(M) OffsetDescription
OffsetLocationOf
Characterize
Variables
VariableNameCV
(M) Term
(O) Definition
SpeciationCV
(M) Term
(O) Definition
SampleMediumCV
(M) Term
(O) Definition
ValueTypeCV
(M) Term
(O) Definition
DataTypeCV
(M) Term
(O) Definition
GerneralCategoryCV
(M) Term
(O) Definition
DefinesCV
Variables
(M) VariableID {PK}
(M) VariableCode
(M) VariableName
(M) Speciation
(M) VariableUnitsID {FK}
(M) SampleMedium
(M) ValueType
(M) IsRegular
(M) TimeSupport
(M) TimeUnitsID {FK}
(M) DataType
(M) GeneralCategory
(M) NoDataValue
CharacterizeCategorical
Categorical Data
Categories
(M) VariableID {FK}
(M) DataValue
(M) CategoryDescription
SpecifyQualityOf
Data Qualifiers
Qualify
QualityControlLevels
(M) QualityControlLevelID {PK}
(M) QualityControlLevelCode
(M) Definition
(M) Explanation
Qualifiers
(M) QualifierID {PK}
(O) QualifierCode
(M) QualifierDescription
SeriesCatalog
(P) SeriesID {PK}
(P) SiteID {FK}
(P) SiteCode
(P) SiteName
(P) SiteType
(P) VariableID {FK}
(P) VariableCode
(P) VariableName
(P) Speciation
(P) VariableUnitsID {FK}
(P) VariableUnitsName
(P) SampleMedium
(P) ValueType
(P) TimeSupport
(P) TimeUnitsID {FK}
(P) TimeUnitsName
(P) DataType
(P) GeneralCategory
(P) MethodID {FK}
(P) MethodDescription
(P) SourceID {FK}
(P) Organization
(P) SourceDescription
(P) Citation
(P) QualityControlLevelID {FK}
(P) QualityControlLevelCode
(P) BeginDateTime
(P) EndDateTime
(P) BeginDateTimeUTC
(P) EndDateTimeUTC
(P) ValueCount

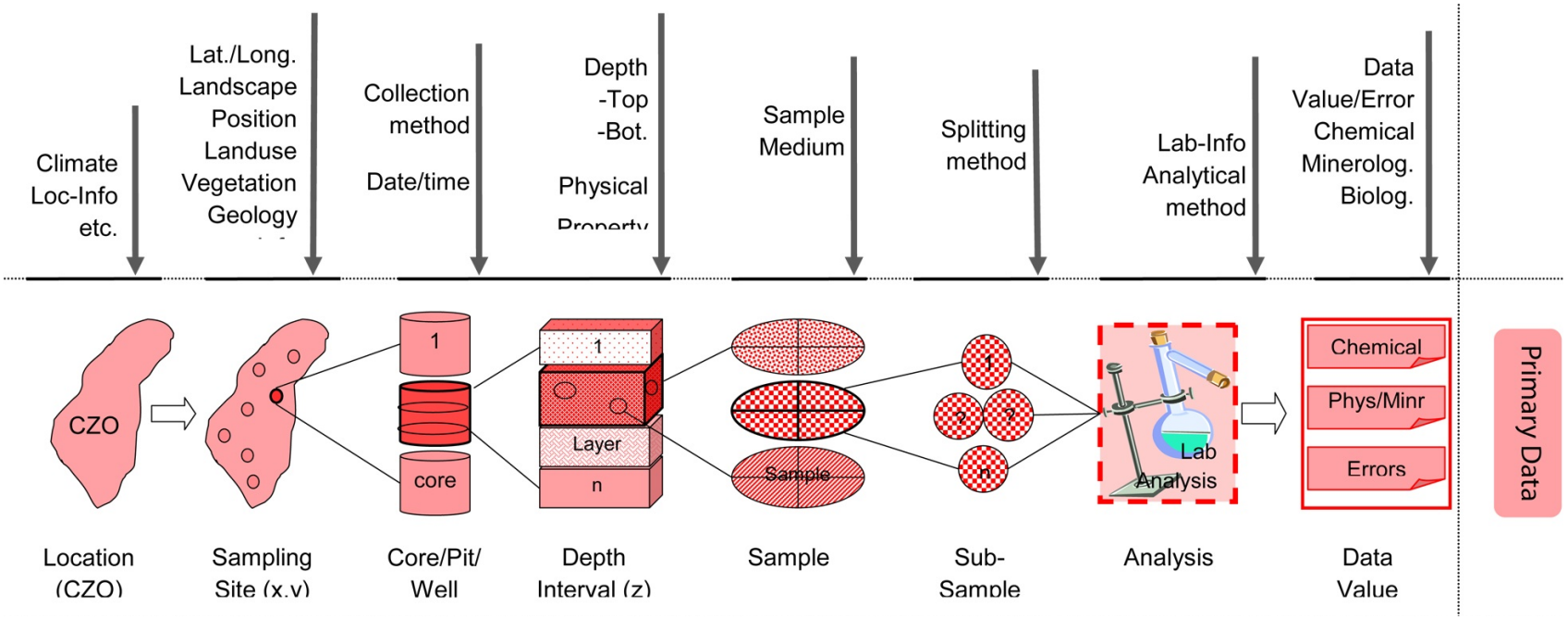
WaterML 1.1


```
<timeSeriesResponse>
  <queryInfo>
    <creationTime>2012-04-13T11:45:35.3363483-06:00</creationTime>
    <criteria MethodCalled="GetValues"></criteria>
  </queryInfo>
  <timeSeries>
    <sourceInfo xsi:type="SiteInfoType">
      <siteName>Little Bear River at Mendon Road near Mendon, Utah</siteName>
      <siteCode network="LittleBearRiver" siteID="1">USU-LBR-Mendon</siteCode>
      <geoLocation>
        <geogLocation xsi:type="LatLonPointType" srs="EPSG:4269">
          <latitude>41.718473</latitude>
          <longitude>-111.946402</longitude>
        </geogLocation>
        <localSiteXY projectionInformation="NAD83 / UTM zone 12N"></localSiteXY>
      </geoLocation>
      <elevation_m>1345</elevation_m>
      <verticalDatum>NGVD29</verticalDatum>
      <siteProperty name="County">Cache</siteProperty>
      <siteProperty name="State">Utah</siteProperty>
      <siteProperty name="Site Comments">
        Located below county road bridge at Mendon Road crossing
      </siteProperty>
      <siteProperty name="PosAccuracy_m">10</siteProperty>
    </sourceInfo>
    <variable>
      <variableCode vocabulary="LittleBearRiver" default="true" variableID="44">USU44</variableCode>
      <variableName>Discharge</variableName>
      <valueType>Derived Value</valueType>
      <dataType>Average</dataType>
      <generalCategory>Hydrology</generalCategory>
      <sampleMedium>Surface Water</sampleMedium>
      <unit>
        <unitName>cubic feet per second</unitName>
        <unitType>Flow</unitType>
        <unitAbbreviation>cfs</unitAbbreviation>
        <unitCode>35</unitCode>
      </unit>
      <noDataValue>-9999</noDataValue>
      <timeScale isRegular="true">
        <unit>
          <unitName>minute</unitName>
          <unitType>Time</unitType>
          <unitAbbreviation>min</unitAbbreviation>
          <unitCode>102</unitCode>
        </unit>
        <timeSupport>30</timeSupport>
      </timeScale>
      <speciation>Not Applicable</speciation>
    </variable>
    <values>
      <value censorCode="nc" dateTime="2008-01-01T00:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T07:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.7815402112921</value>
      <value censorCode="nc" dateTime="2008-01-01T00:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T07:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.7775823779046</value>
      <value censorCode="nc" dateTime="2008-01-01T01:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T08:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.767691480856</value>
      <value censorCode="nc" dateTime="2008-01-01T01:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T08:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.7083692101962</value>
      <value censorCode="nc" dateTime="2008-01-01T02:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T09:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.6313420514041</value>
      <value censorCode="nc" dateTime="2008-01-01T02:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T09:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.6214746973181</value>
      <value censorCode="nc" dateTime="2008-01-01T03:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T10:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.5169693587662</value>
      <value censorCode="nc" dateTime="2008-01-01T03:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T10:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.6668869468178</value>
      <value censorCode="nc" dateTime="2008-01-01T04:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T11:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.7538407227539</value>
      <value censorCode="nc" dateTime="2008-01-01T04:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T11:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.9241658910827</value>
      <value censorCode="nc" dateTime="2008-01-01T05:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T12:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">26.8924473946182</value>
      <value censorCode="nc" dateTime="2008-01-01T05:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T12:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">27.5420329911293</value>
      <value censorCode="nc" dateTime="2008-01-01T06:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T13:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">28.1680112854555</value>
      <value censorCode="nc" dateTime="2008-01-01T06:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T13:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">28.7858029935971</value>
      <value censorCode="nc" dateTime="2008-01-01T07:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T14:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">29.7358364392533</value>
      <value censorCode="nc" dateTime="2008-01-01T07:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T14:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">30.3620096087977</value>
      <value censorCode="nc" dateTime="2008-01-01T08:00:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T15:00:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">31.2155850870149</value>
      <value censorCode="nc" dateTime="2008-01-01T08:30:00" timeOffset="-07:00" dateTimeUTC="2008-01-01T15:30:00" methodCode="29" sourceCode="2" qualityControlLevelCode="2">31.8465165668535</value>
```

Some Limitations of ODM 1.1.1

- ODM supports only point-based observations
- ODM doesn't support sample-based data well
- Not all of the structure of ODM is required for each type of data
- Versioning and provenance of data can be difficult in ODM
- The central DataValues table doesn't support every functional use case (e.g., metadata catalog)

Another Information Model Example

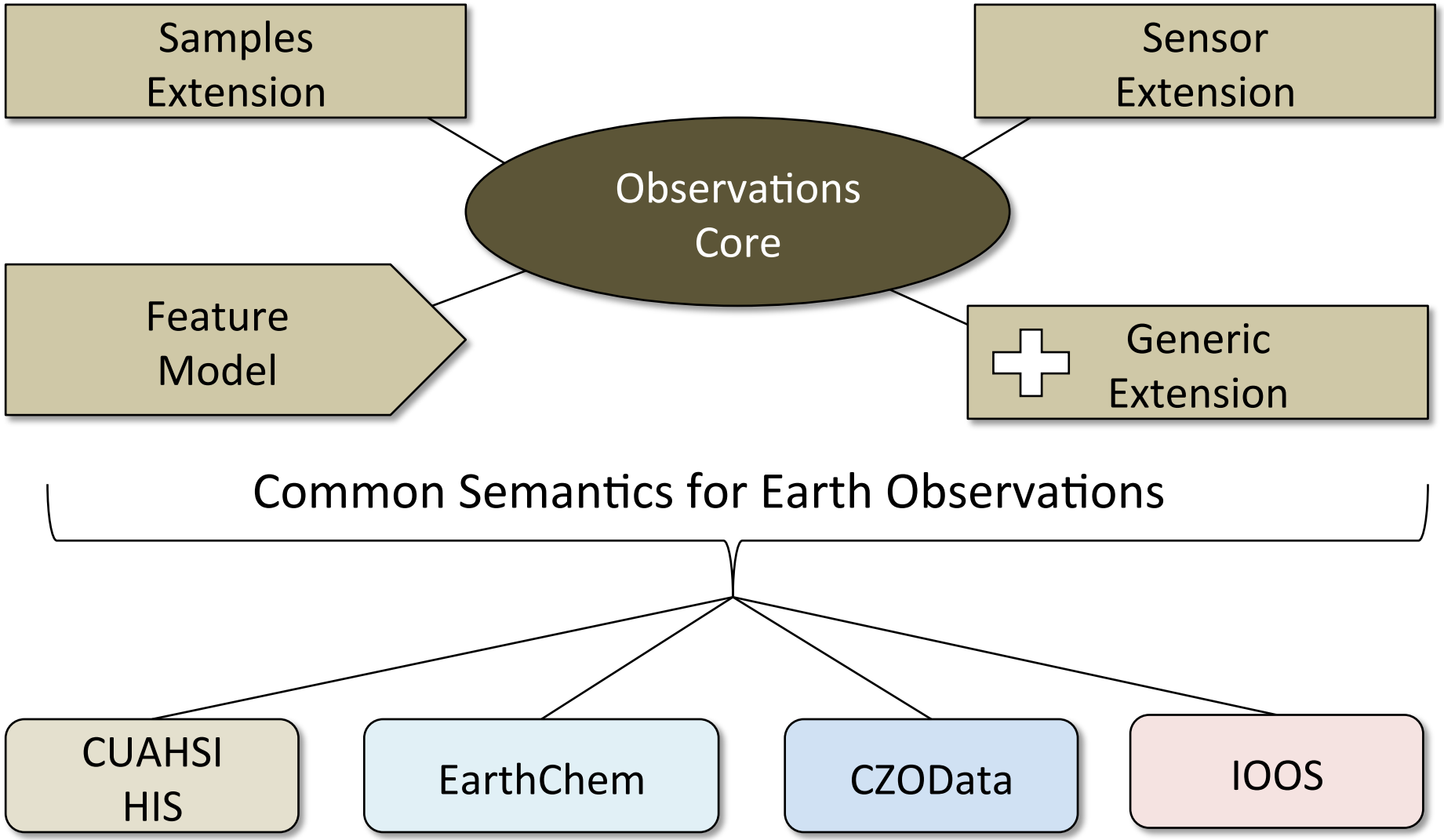


Xianzeng Niu, Jennifer Z. Williams, Doug Miller, Kerstin Lehnert, Brian Bills, Susan L. Brantley, (2013), An Ontology Driven Relational Geochemical Database for the Earth's Critical Zone: CZchemDB, Submitted to *Journal of Environmental Informatics*.

Overarching Goals for ODM2 Project

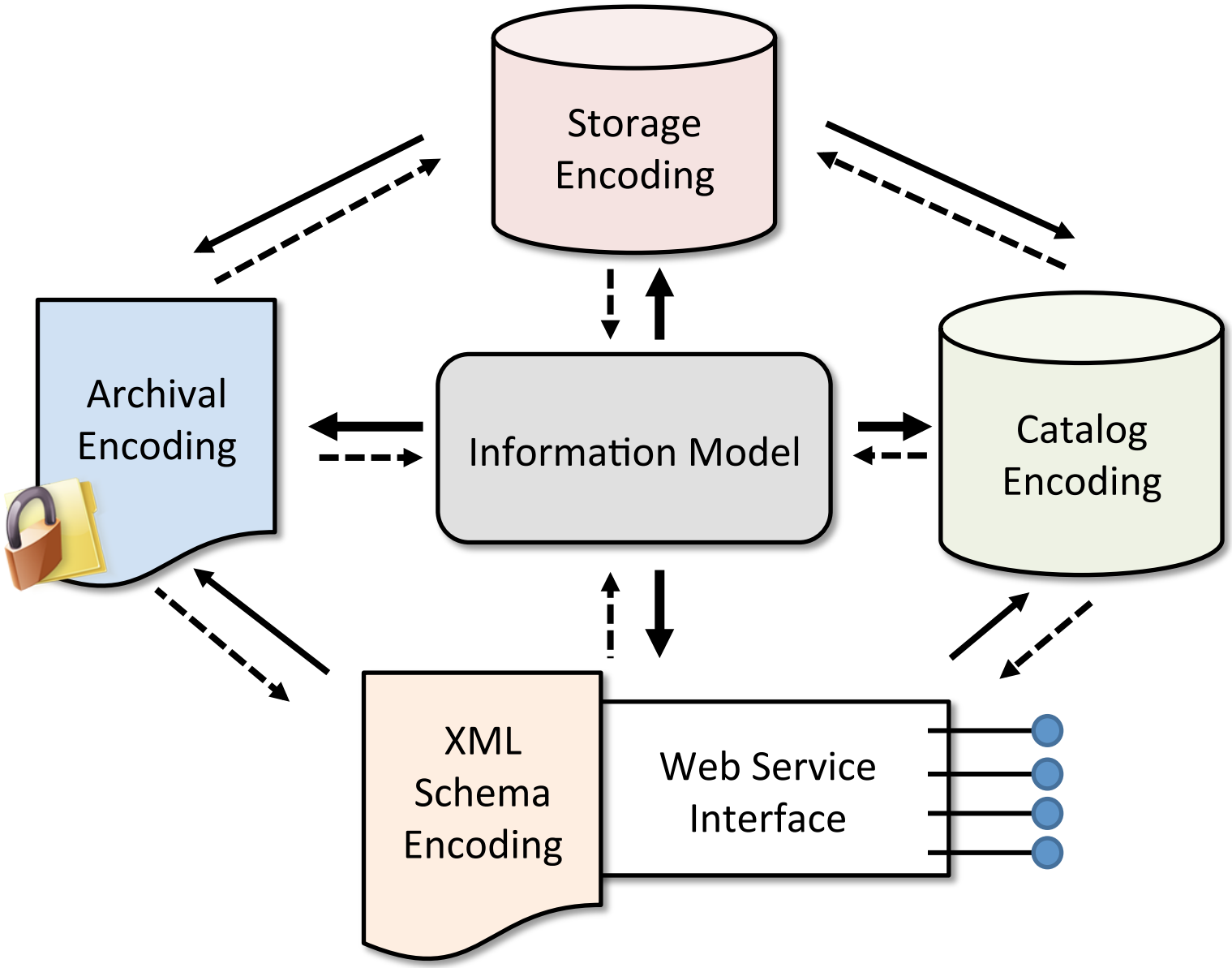
- Create an ***information model*** that is ***integrative*** and ***extensible***
  - Accommodating a wide range of observational data
  - Aimed at achieving interoperability across multiple disciplines and systems that support publication of earth observations
- Allow a diverse range of geoscience observations to be consistently shared, discovered, accessed, and interpreted

Observations Information Model



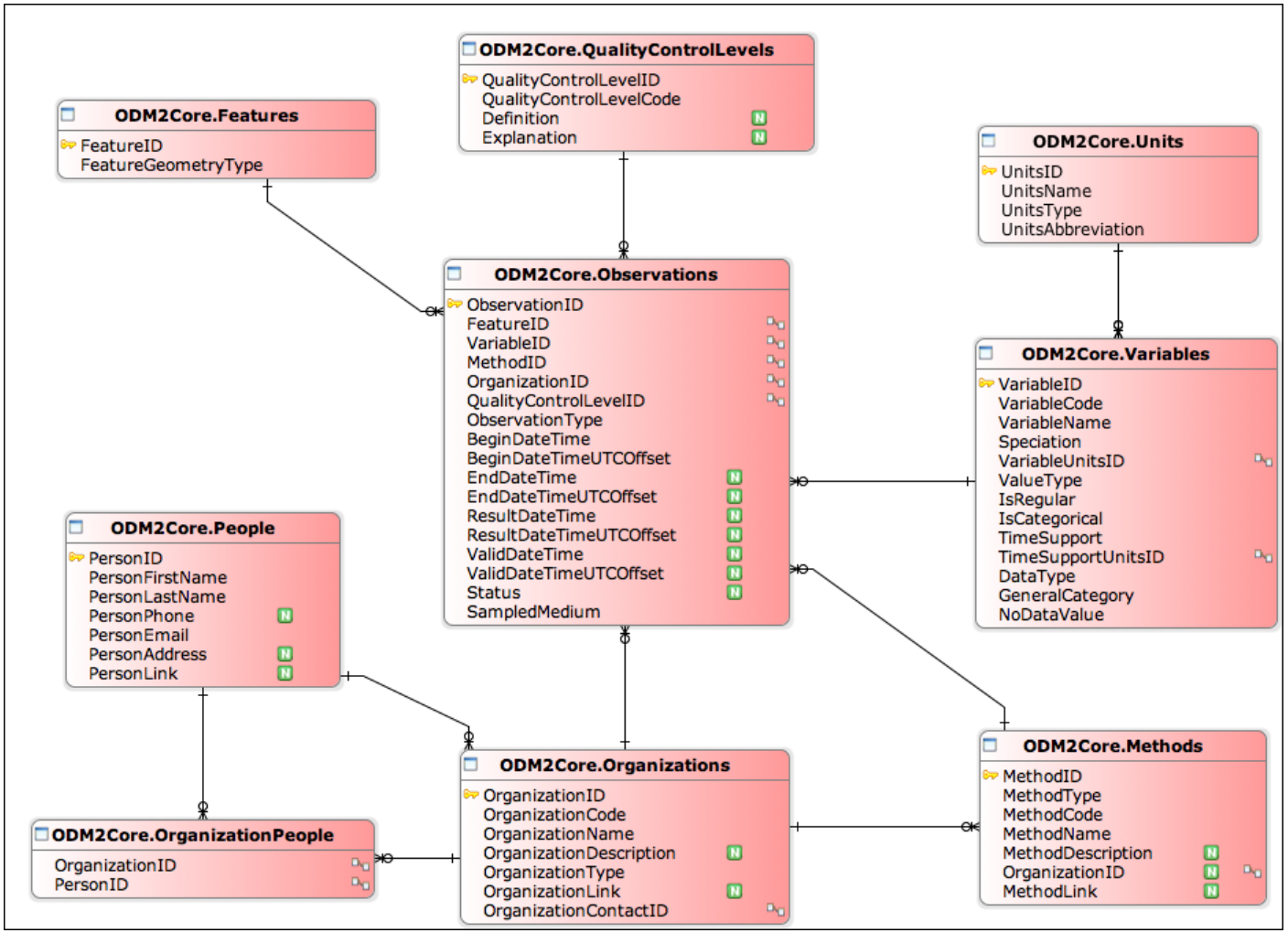
Domain Cyberinfrastructures

ODM2 Functional Use Cases



ODM2 - Core

- **Observations**
- Features
- People
- Organizations
- Methods
- Variables
- QCLevels
- Units



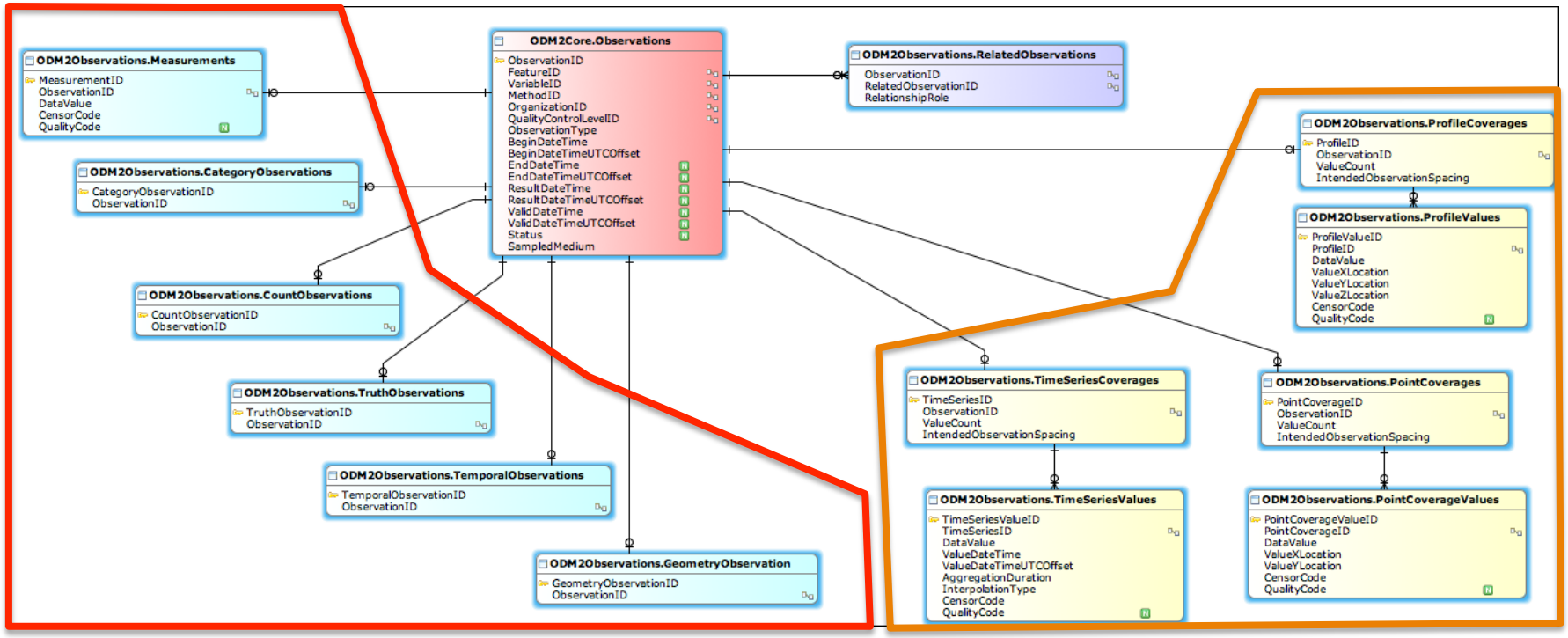
ODM2 Observations

Open Geospatial Consortium Observations & Measurements – 2 Types of observations

1. Observations whose result is constant
2. Observations whose result varies with space or time

- Separate the concept of an “observation” and its “result”

ODM2 Observations



Observations whose result is constant

- Measurements
- Category observations
- Count observations
- Truth observations
- Temporal observations
- Geometry observations

Observations whose result varies in space and/or time

- Time series coverage
- Point coverage
- Profile coverage

ODM2 Sensor Extension

Equipment

- Model
- Serial number
- Owner
- Vendor
- Manufacturer
- Service history

Calibrations

- Method
- Standard

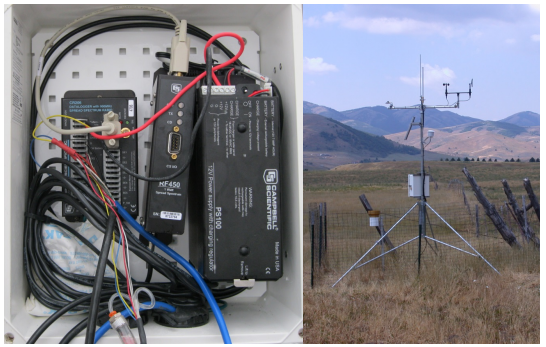
Site visits

- Location
- Date
- People
- Conditions

Field activities

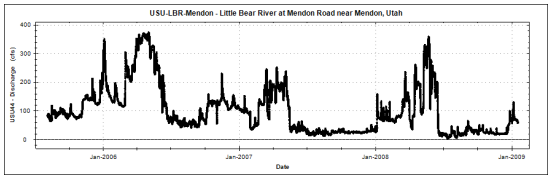
- Activity type
- Description
- Date

Deployments

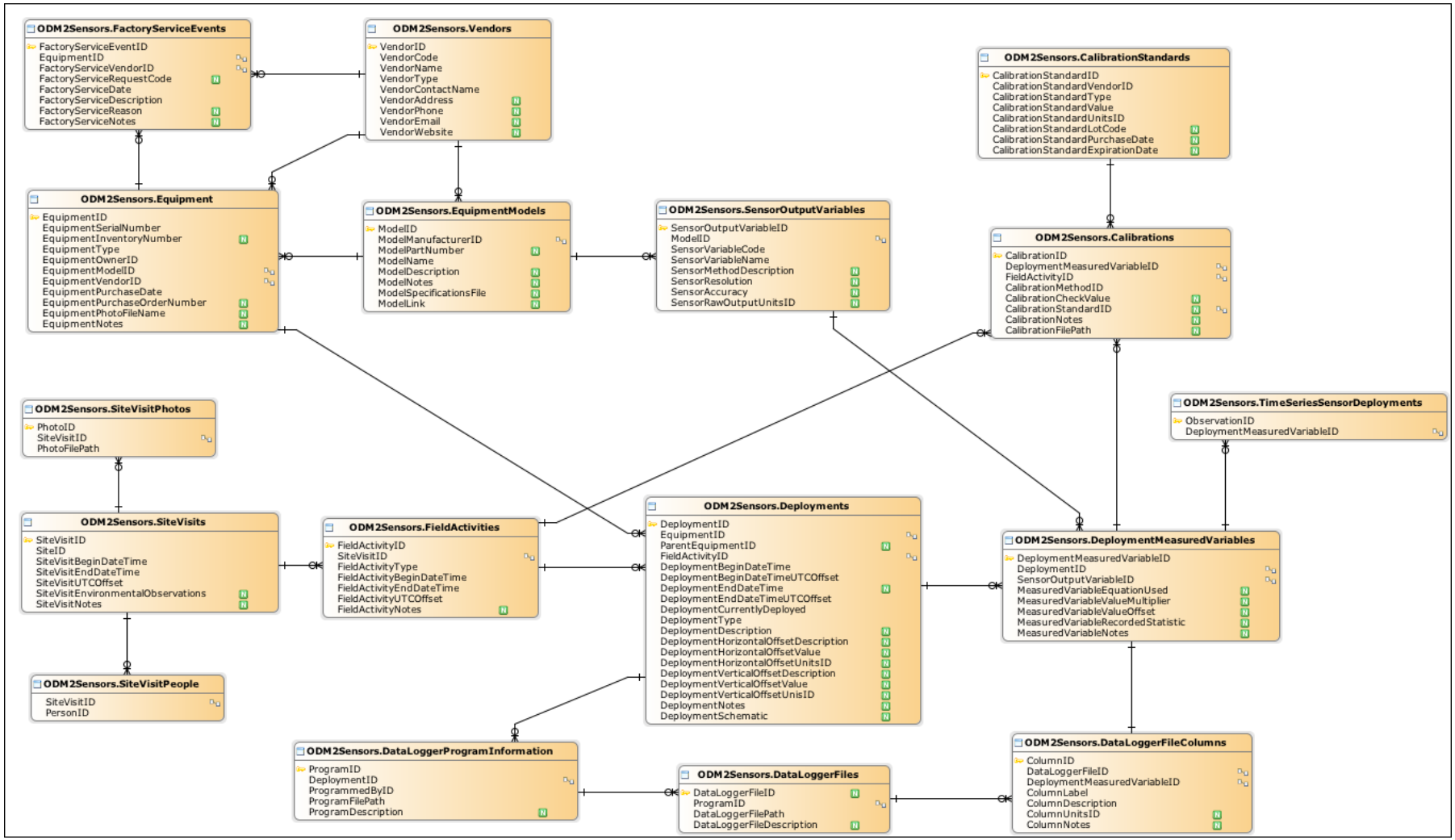


- Deployment type
- Description
- Dates
- Offsets

Time Series Observations



ODM2 Sensor Extension



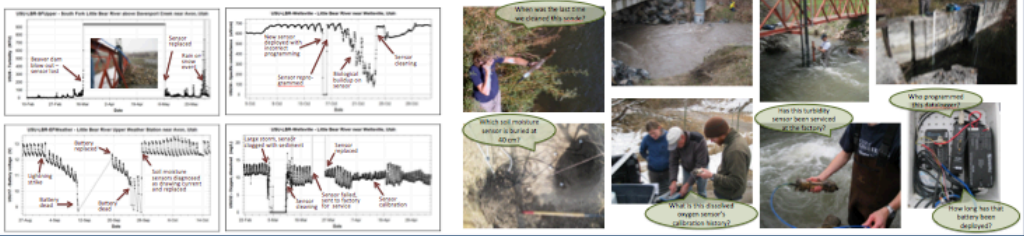
Managing Sensor Infrastructure: A Sensor Extension for ODM2

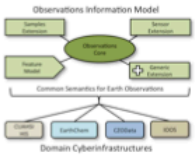
Amber Spackman Jones, Jeffery S. Horsburgh, Juan Caraballo, Maurier Ramirez
Utah Water Research Laboratory, Utah State University

This project is funded by National Science Foundation grants EPS-1208732 and EAR-1224638.

Background: The Observations Data Model (ODM) is a relational data model for storage and management of environmental observations data designed to capture consistent descriptions for unambiguous interpretation among users. ODM 1.1 was intended for publication of point-based hydrologic observations using the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS). Because of its focus on data publication, ODM 1.1 does not provide all of the functionality needed to fully describe the data collection and management process for sensor and sample-based observations. We are working to develop a new version of ODM (ODM2) that better supports both management and publication of sensor and sample-based observations.

Motivation: Research sites conducting long term monitoring using *in situ* sensors need the functionality to track equipment, deployments, calibrations, and other events related to site maintenance and to link this information to the observational data that they are collecting.



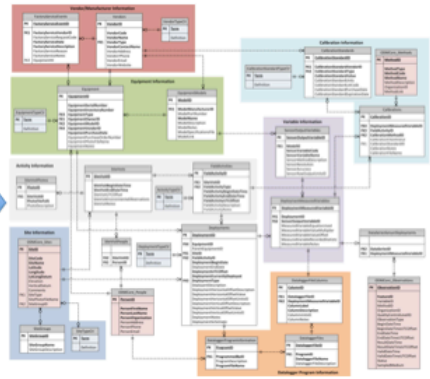


ODM2 Core: ODM2 consists of a core set of entities common to all types of observational data to describe location, observed variable, and time.

ODM2 Extensions: Additional metadata can be provided via extensions to the core to support particular types of observations.

Extensions Under Development:

- **Sensors:** describe monitoring equipment, deployments, and *in situ* time series.
- **Samples**: capture sample hierarchy and analytical methods.
- **Features:** detail the physical feature(s) from which an observation is made.
- **Annotations:** provide qualifiers or comments to variables or observations.
- **Provenance**: describe data versioning and processing.



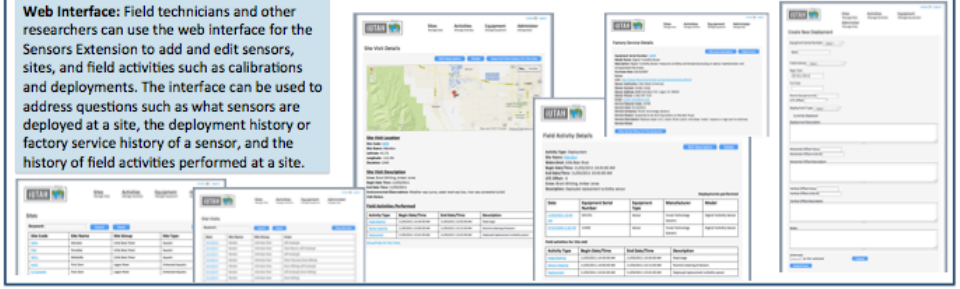
Linkage to ODM2 Core: The connection between the Sensors Extension and the ODM2 Core provides linkages between observations, equipment, and events associated with the observations. The linking tables are Methods, Observations, People, and Sites.

Sensors Extension. The Sensors Extension permits information about individual sensors, deployments, and the associated time series to be recorded and managed.

Table Descriptions

- **Equipment** contains attributes of individual sensors and other pieces of field equipment while **EquipmentModels** describes attributes of equipment of the same type.
- **Deployments** provides information about the sensor deployment in space and time.
- **SiteVisits/FieldActivities/Calibrations** record activities of field crews and sensor calibrations.
- **Vendors** provides information on equipment manufacturers and suppliers and **Factory Service Events** records sensor servicing.
- **SensorOutputVariables** information on a specific variable being measured by a sensor. Linked to the specific deployment through **DeploymentMeasuredVariables**.
- **DataloggerProgram** tables allow for tracking of programs, files, and each variable measured.
- **Controlled Vocabularies:** used for types of equipment, deployments, sites, vendors, activities, and calibration standards.

Web Interface: Field technicians and other researchers can use the web interface for the Sensors Extension to add and edit sensors, sites, and field activities such as calibrations and deployments. The interface can be used to address questions such as what sensors are deployed at a site, the deployment history or factory service history of a sensor, and the history of field activities performed at a site.



Implementation: The iUTAH (innovative Urban Transitions and Aridregion Hydro-sustainability) network of aquatic and terrestrial sensors is being used as a test case for the Sensors Extension. The ODM2 Sensors Extension and associated tools will be useful for similar large scale and long term monitoring networks. See data.iutahepscor.org.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Contact Information
Amber Spackman Jones, (435)797-7147, amber.jones@usu.edu
Jeffery S. Horsburgh, (435)797-2946, jeff.horsburgh@usu.edu
Juan Caraballo, (435)797-0045, juan.caraballo17@gmail.com
Maurier Ramirez, (435)797-0045, mauriel.ramirez@gmail.com

ODM2 Samples Extension

Sample collection

- Sample type
- Location / sampled feature
- Collection method
- Date
- People
- Preservation
- Identifiers
- Sampled medium

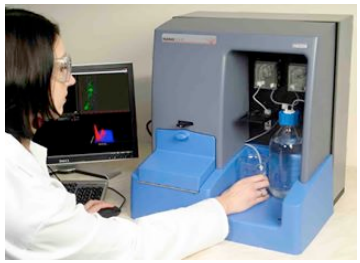
Sample splitting

- Sample hierarchy
- Parent/child relationships

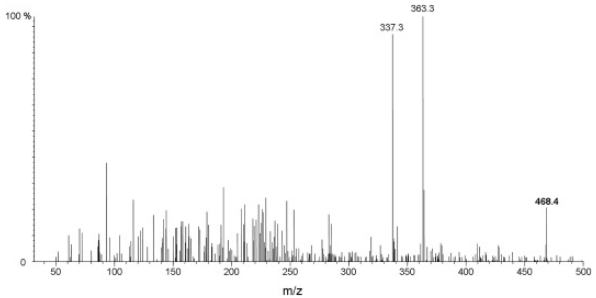
Sample processing

- Preparation method
- Processing method
- People

Sample analysis



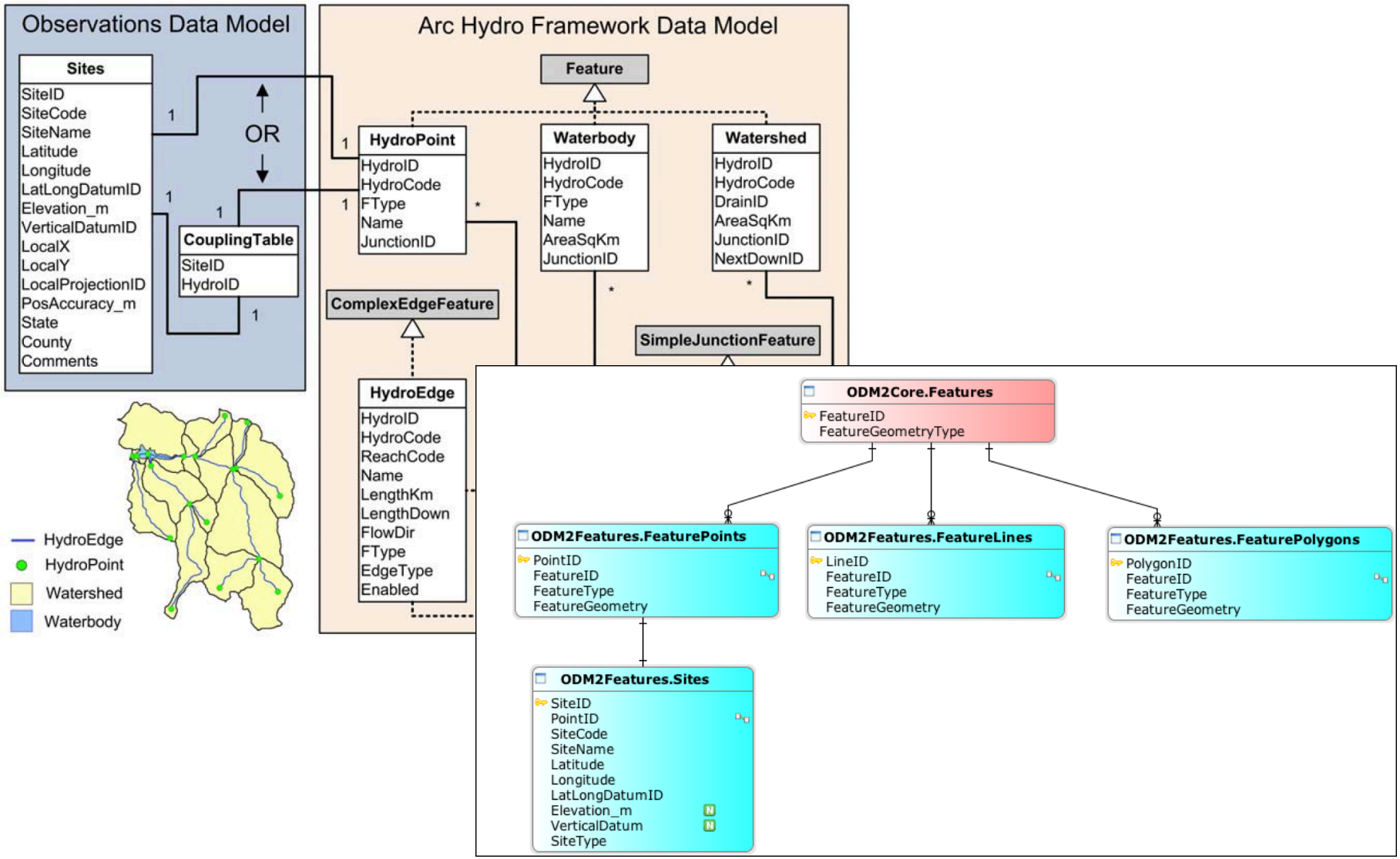
- Analytical method
- Instrument
- People
- Date



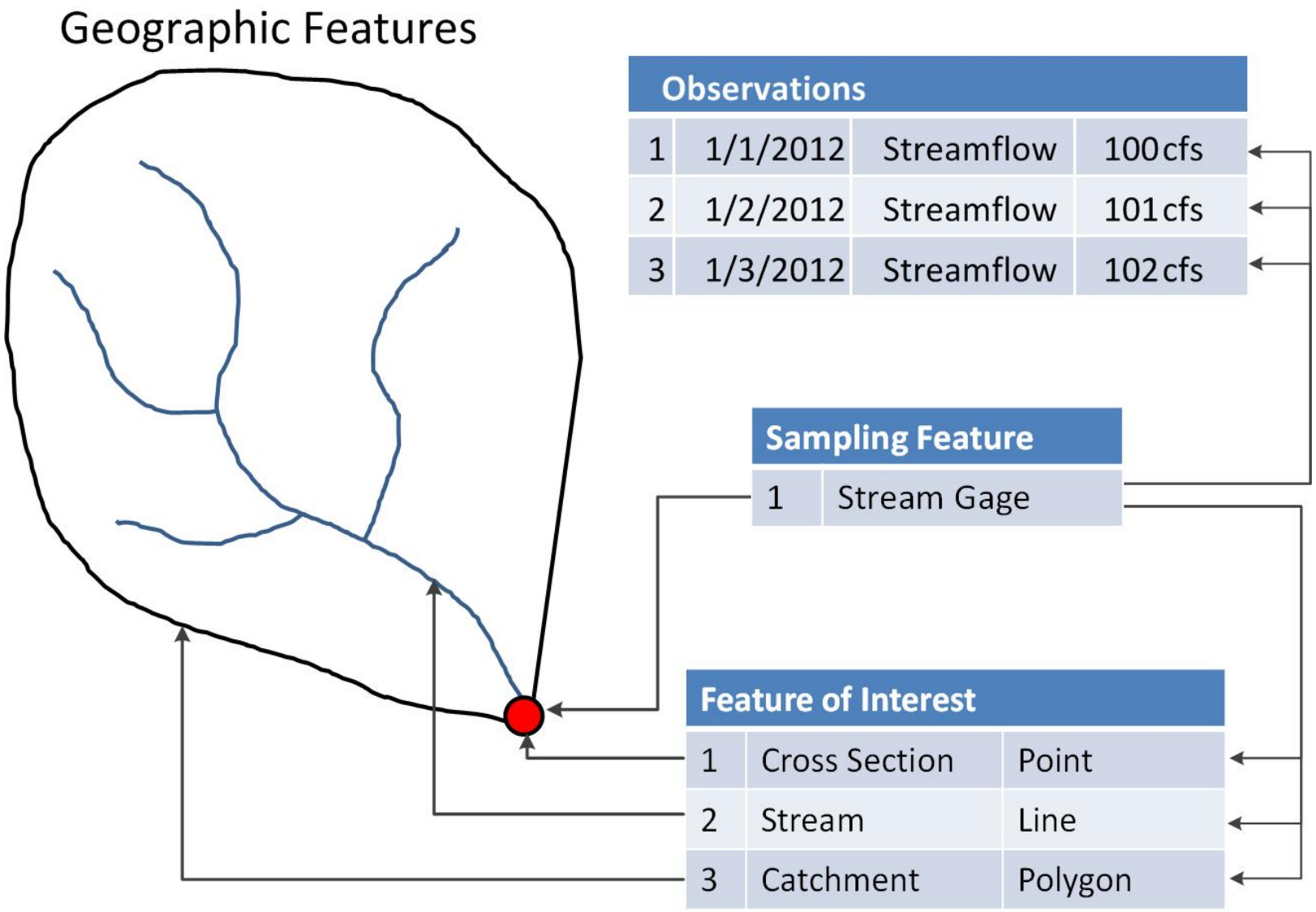
Sample analyses result in observations

ODM2 Samples Extension

ODM2 Features Extension



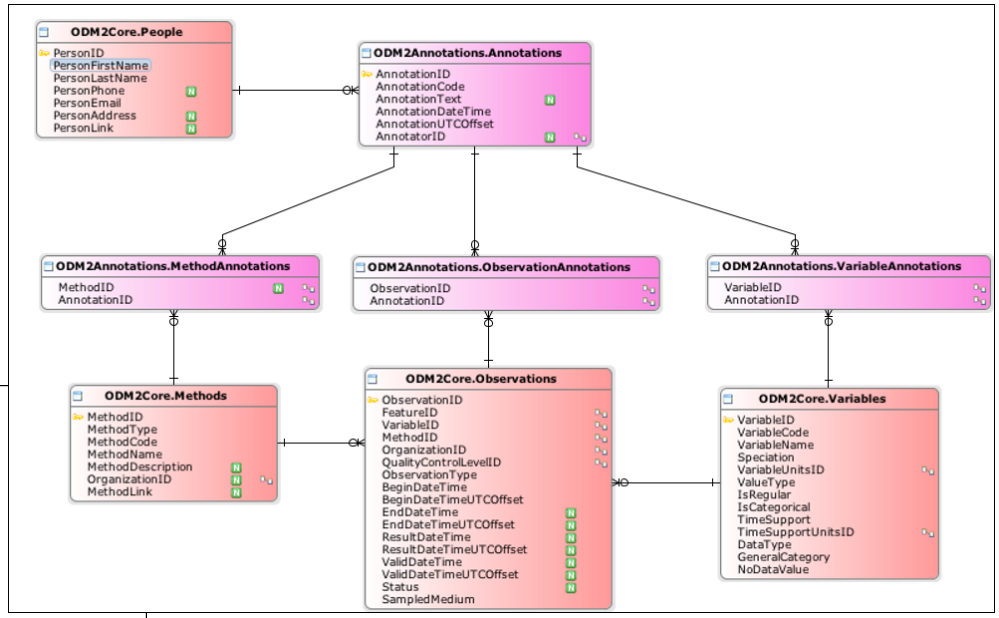
Linking Observations to the Geo-Environment



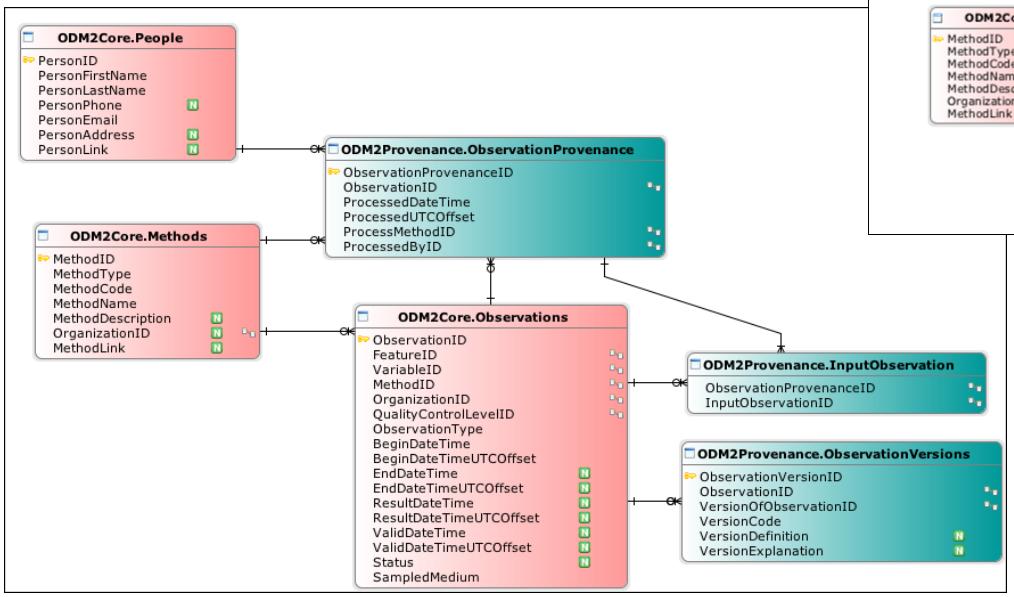
Provenance and Annotations Extensions

- Better support for storing provenance of observational data

Annotations

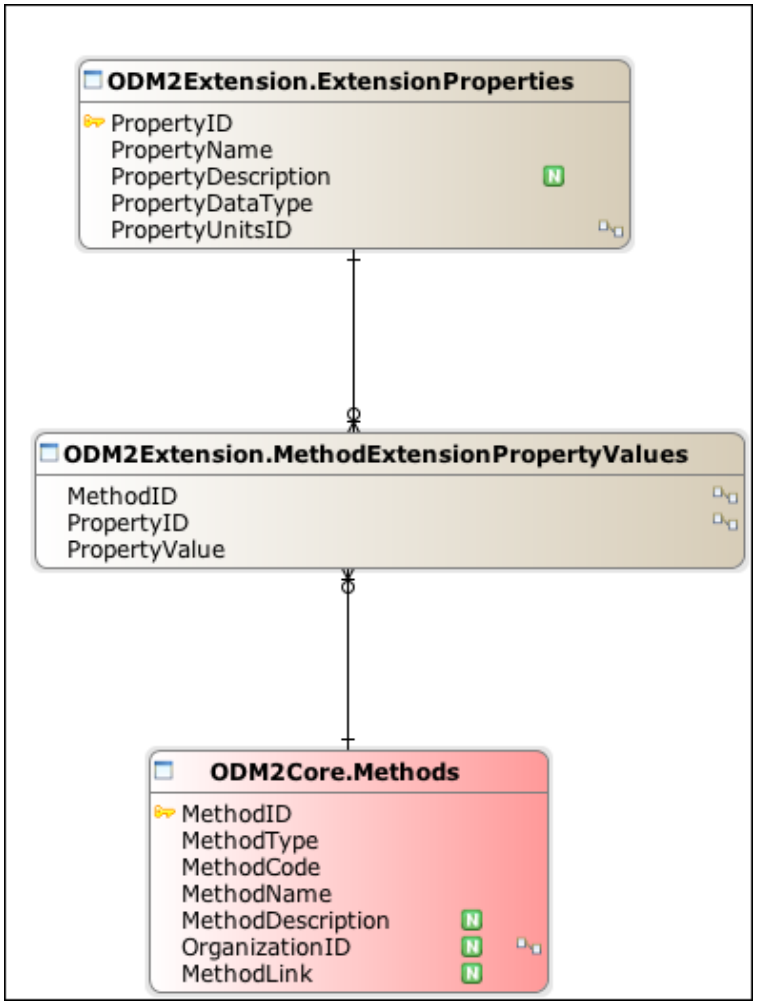


Versions and Provenance



General Extensibility

- Ability to add extension properties to any entity



Impact

- Better support for sample based data in CUAHSI HIS, HydroShare, and related tools
- Support for and integration of sensor and sample-based data in the Critical Zone Observatory Integrated Data Management System (CZOData)
- Potential adoption of ODM within EarthChem
- Providing context and guidance for future WaterML 2.0 development (water quality samples)

Questions?

Support:
EAR 1224638