CZOData Cyber-Seminar #4: 2014-03-27

CZOData Future Systems

Observations Data Model 2.0:

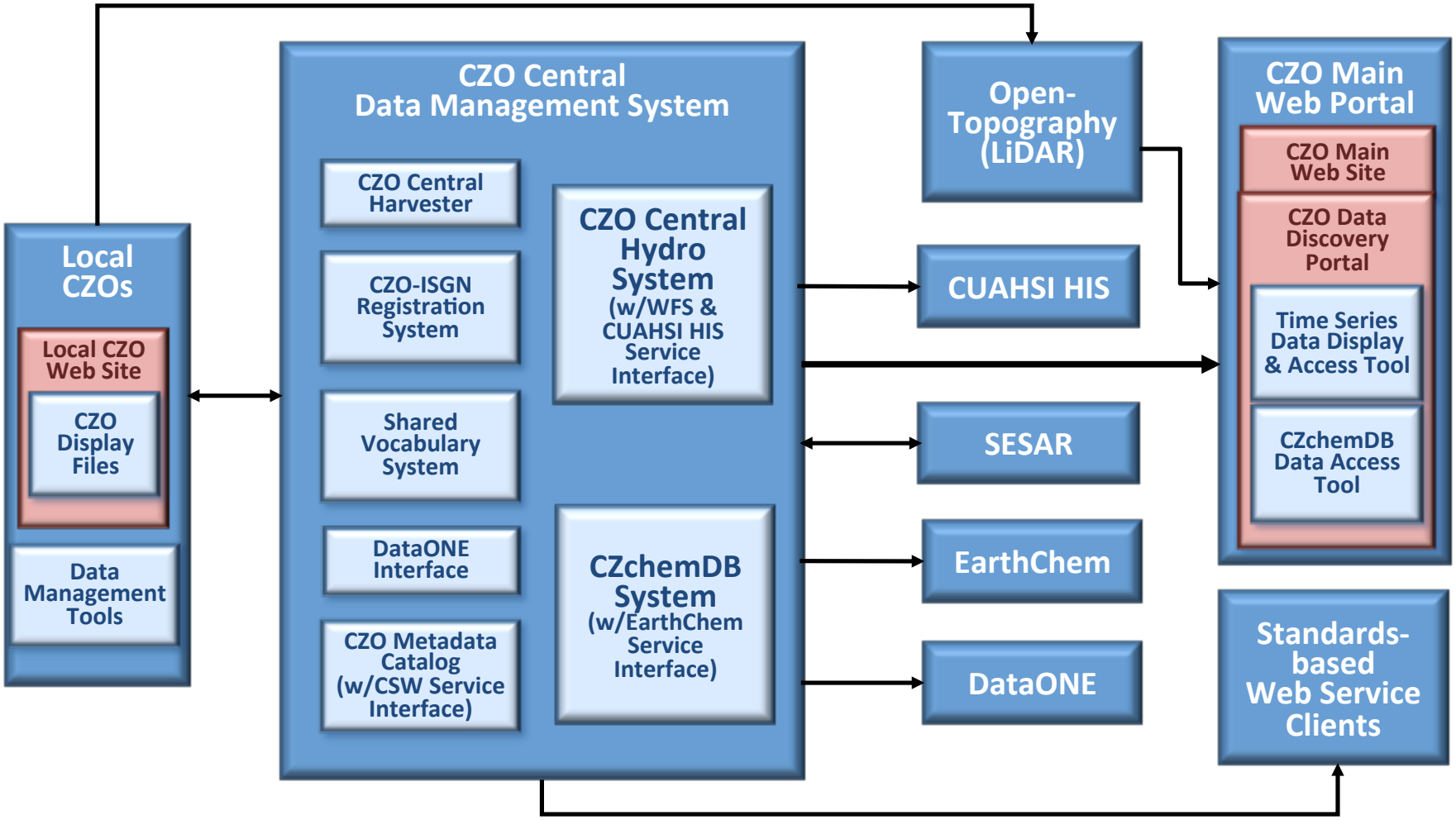
A community information model for interoperability among feature-based earth observations

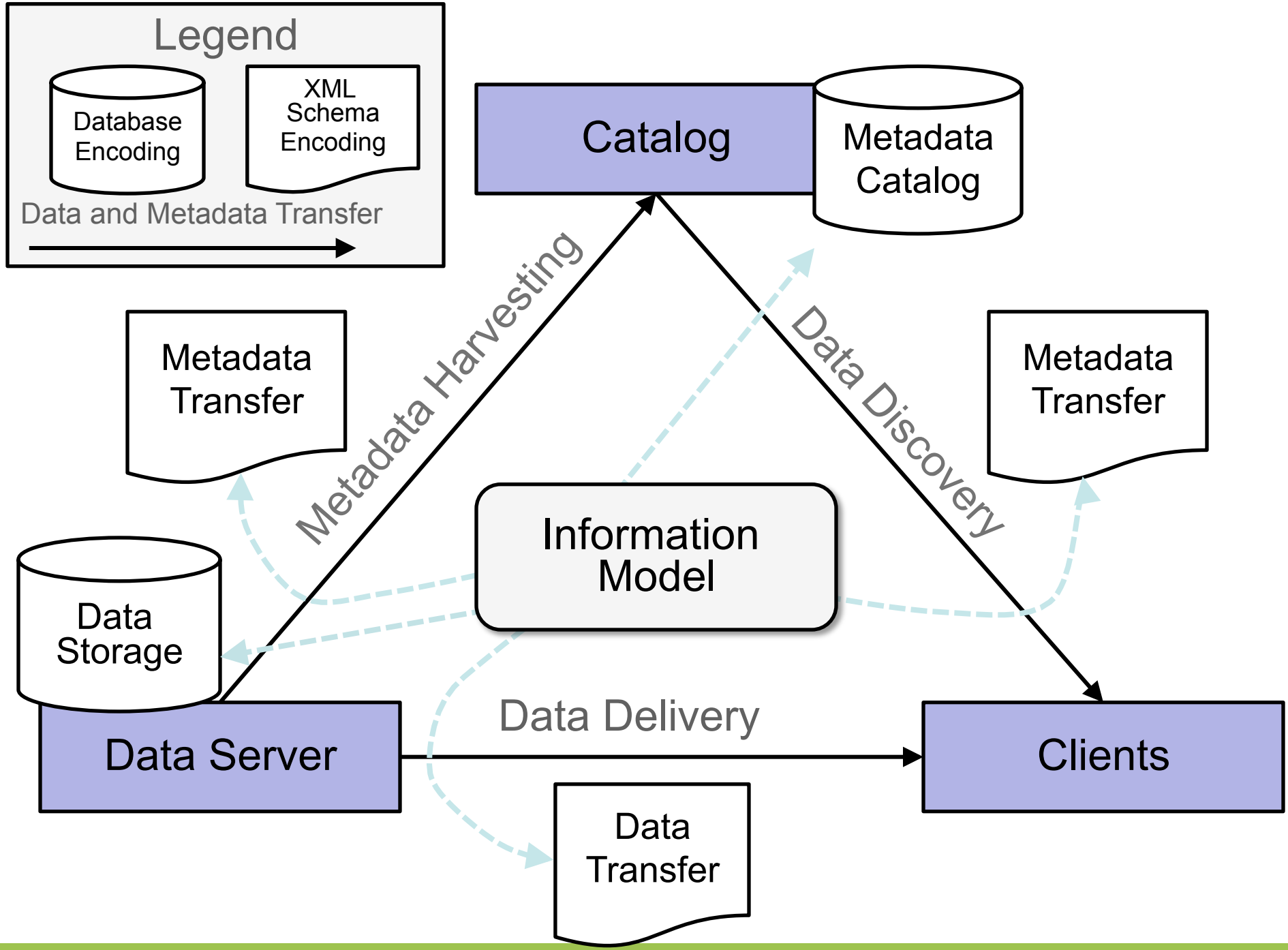
Anthony K. Aufdenkampe, Stroud Water Research Center
Jeff Horsburgh, USU. ODM2 Project PI.

Kerstin Lehnert, IEDA/Columbia
Emilio Mayorga, UW-APL
Ilya Zaslavsky, SDSC
David Valentine, SDSC
David Tarboton, USU
David Lubinski, UC-Boulder

STROUD WATER RESEARCH CENTER
UtahState University
SDSC SAN DIEGO SUPERCOMPUTER CENTER
UNIVERSITY of WASHINGTON
IEDA INTEGRATED EARTH DATA APPLICATIONS

CZOData II Architecture




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

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			

ODM – Linked Table Example

DataValues Table

ValueID	DataValue	ValueAccuracy	LocalDateTime	UTCOffset	SiteID	VariableID	MethodID	DerivedFromID
1	4.18		5/1/2006 00:00:00.000	-7	1	1	1	
97	748		5/1/2006 00:00:00.000	-7	1	2	2	1
193	722	22.89831642	5/1/2006 00:00:00.000	-7	1	3	3	100
2	4.18		5/1/2006 00:15:00.000	-7	1	1	1	
98	748		5/1/2006 00:15:00.000	-7	1	2	2	2
3	4.17		5/1/2006 00:30:00.000	-7	1	1	1	
99	742		5/1/2006 00:30:00.000	-7	1	2	2	3
4	4.17		5/1/2006 00:45:00.000	-7	1	1	1	
100	742		5/1/2006 00:45:00.000	-7	1	2	2	4
5	4.17		5/1/2006 01:00:00.000	-7	1	1	1	
101	742		5/1/2006 01:00:00.000	-7	1	2	2	5
6	4.17		5/1/2006 01:15:00.000	-7	1	1	1	
102	742		5/1/2006 01:15:00.000	-7	1	2	2	6

DerivedFrom Table

DerivedFromID	ValueID
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13

Variables Table

VariableID	VariableCode	VariableName	VariableUnitsID	ValueType	IsRegular	TimeSupport	TimeUnitsID	DataType	GeneralCategory
1	NWIS:00065	Gage height	1	Field Observation	TRUE	15	5	Continuous	Hydrologic
2	NWIS:00060	Discharge	2	Derived Value	TRUE	15	5	Continuous	Hydrologic
3	NWIS:00060	Discharge, daily average	2	Derived Value	TRUE	24	6	Average	Hydrologic
4	NWIS:00300	Dissolved oxygen concentration	3	Field Observation	FALSE	0	5	Sporadic	Water Quality

Units Table

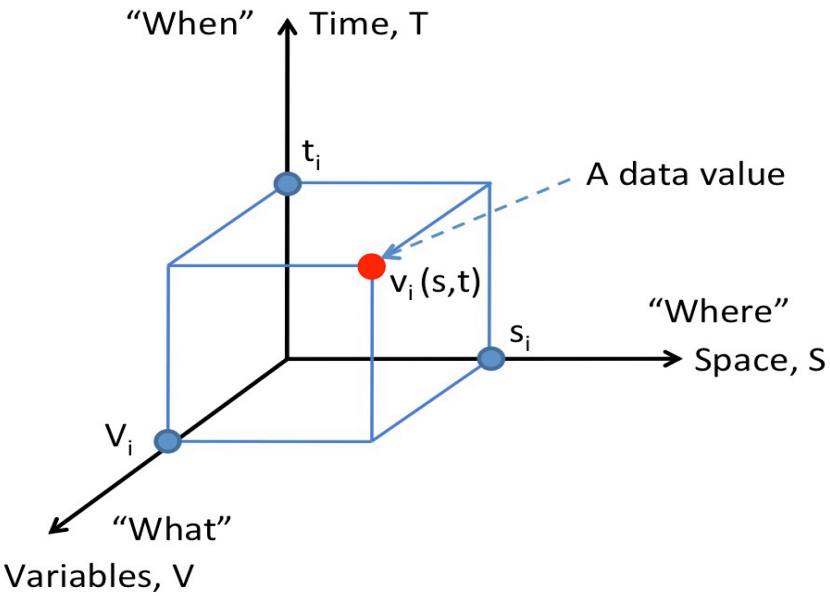
UnitsID	UnitsName	UnitsType	UnitsAbbreviation
1	Feet	Length	ft
2	Cubic feet per second	Flow	ft^3/s
3	Milligrams per liter	Concentration	mg/L
4	Meters	Length	m
5	Minutes	Time	min
6	Hours	Time	hr

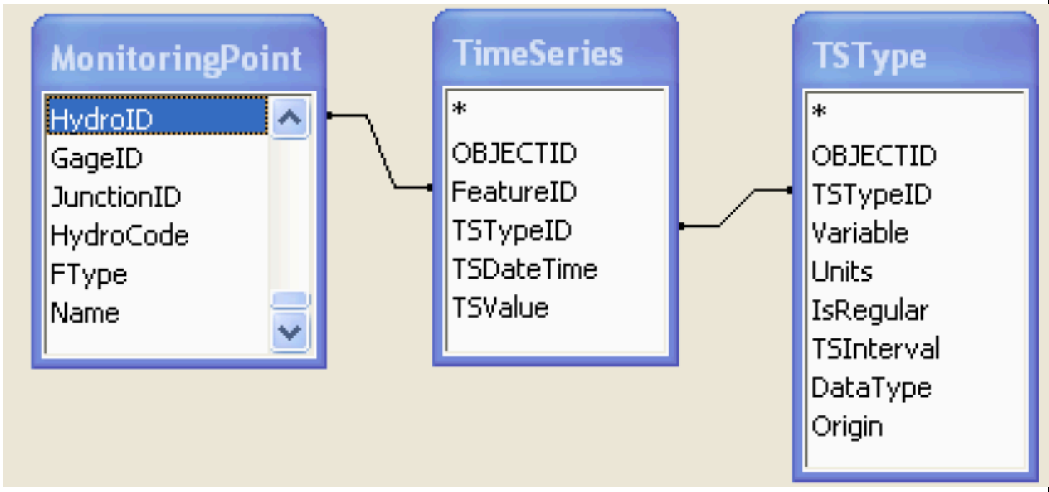
Methods Table

MethodID	MethodDescription
1	Gage height measured with continuous data logger
2	Discharge derived from water stage using site specific rating curve
3	Daily average discharge derived from 15 minute continuous discharge values
4	Dissolved oxygen measured with a Hydrolab multiprobe field instrument
7	Precipitation from a tipping bucket gage. 0 values not logged.
8	Daily precipitation from a tipping bucket gage. 0 values not logged.

Figure 3. Excerpts from tables illustrating the population of ODM with streamflow gauge height (stage) and discharge data.

A Data Model

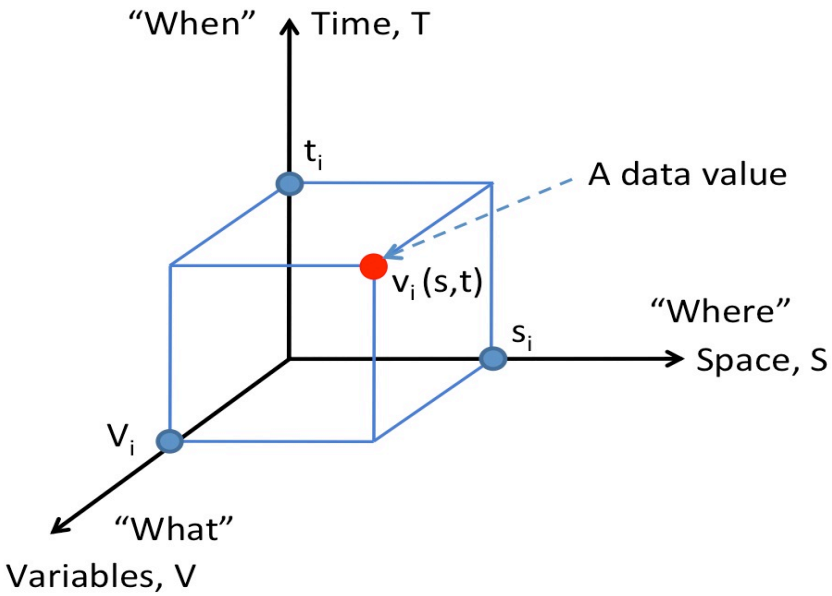


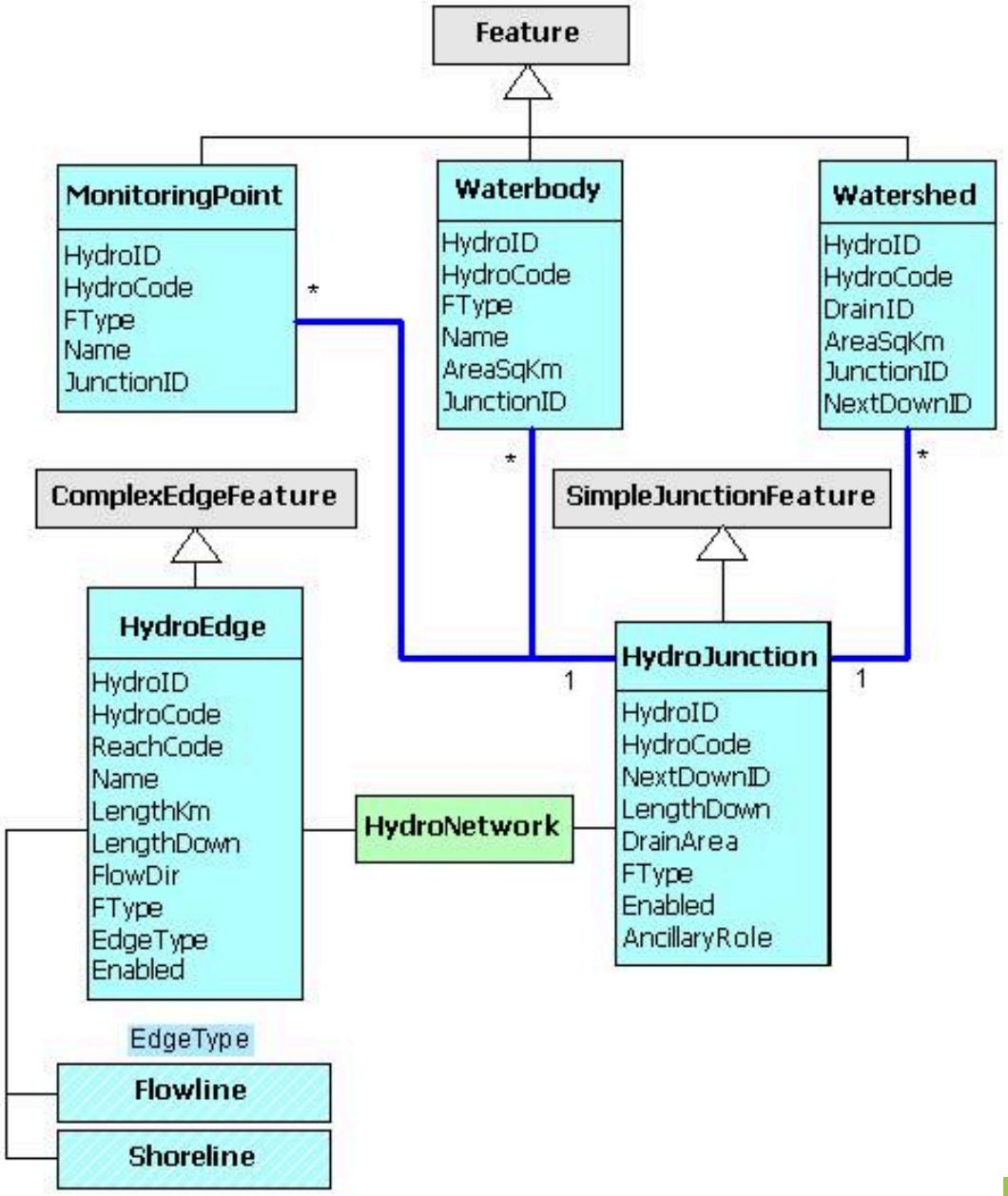


ODM History:

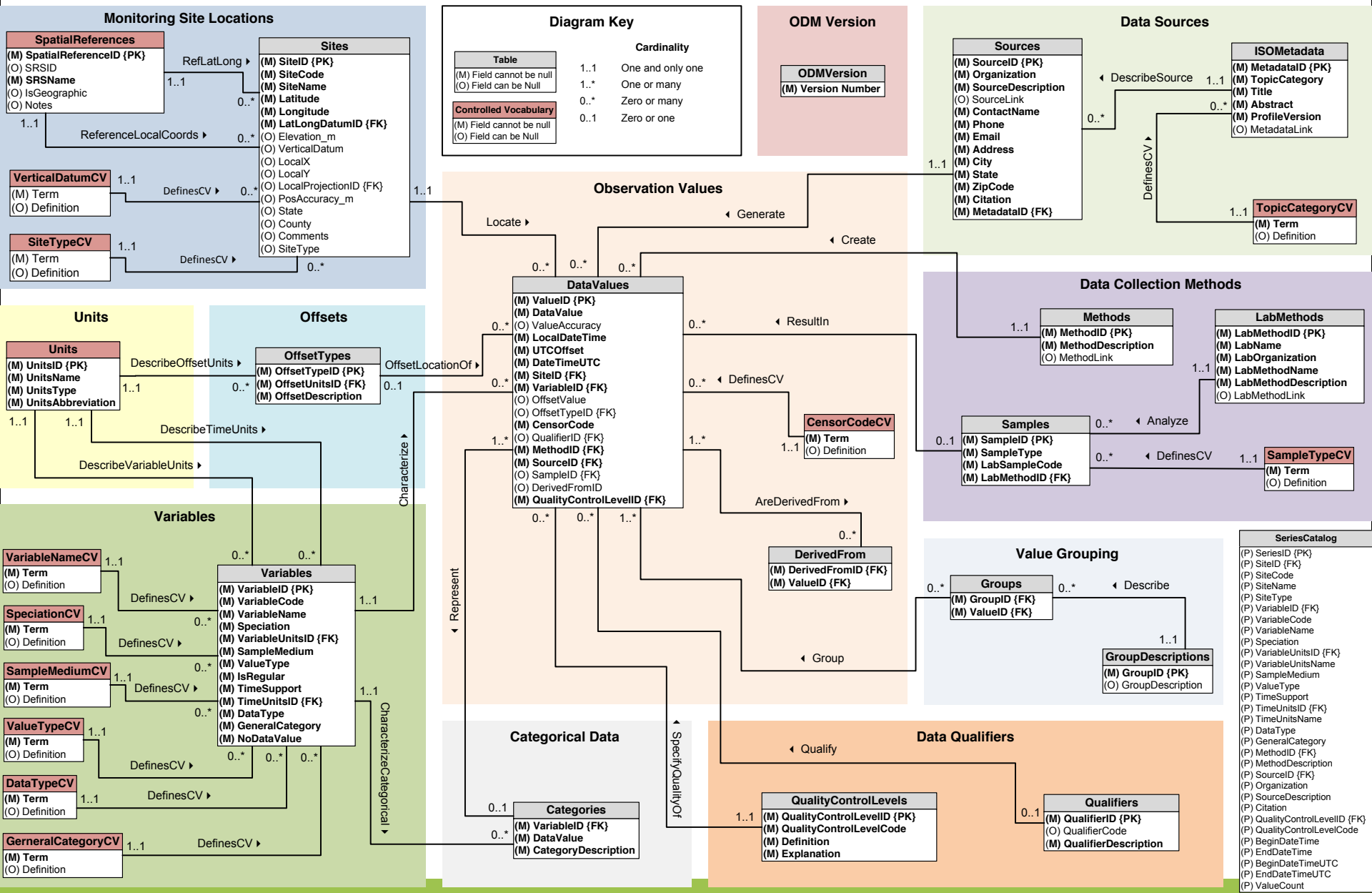
Maidment 2005 white paper suggested ArcHydro Time Series Data Model

An Information Model





Observations Data Model 1.1.1



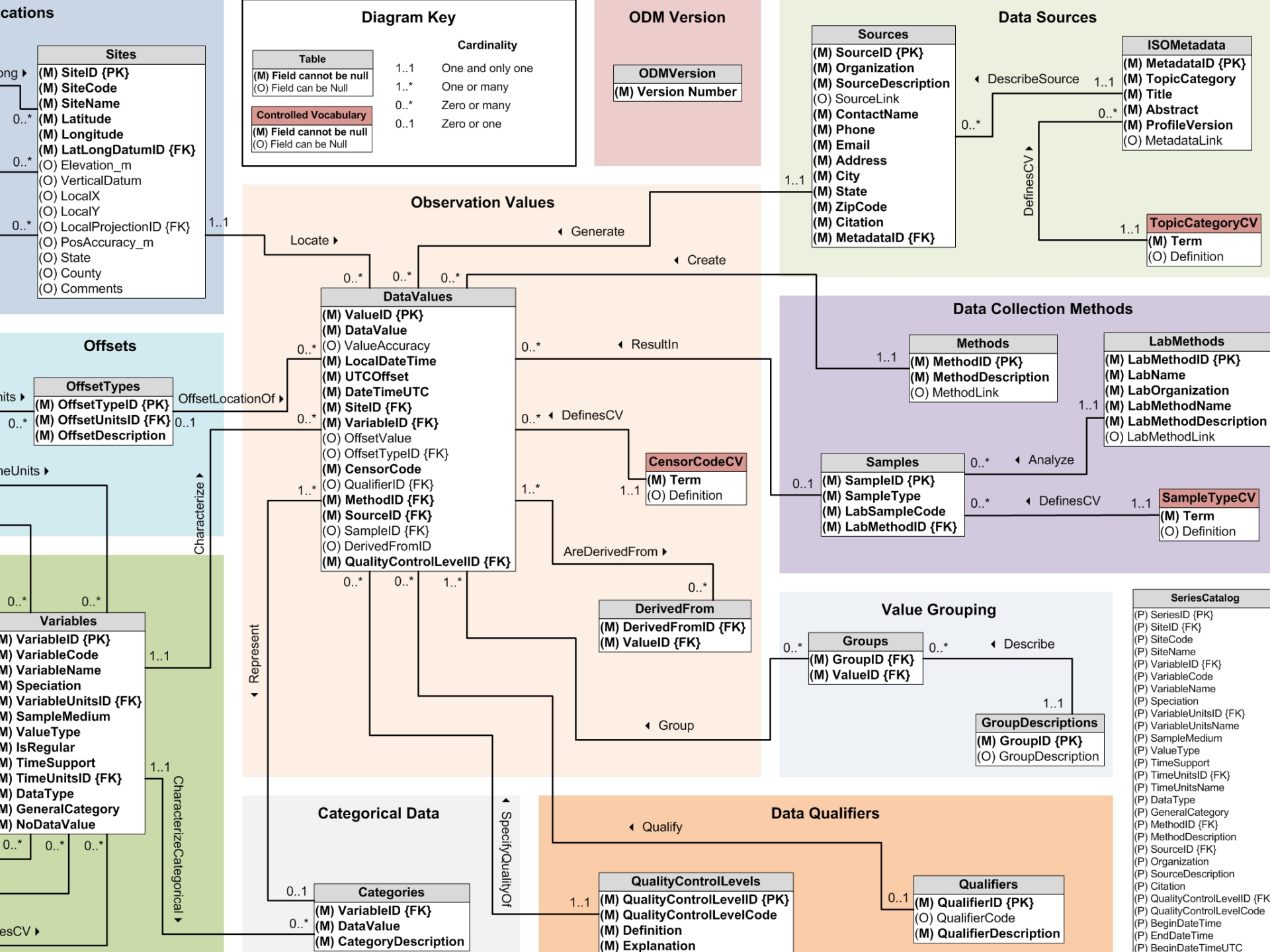
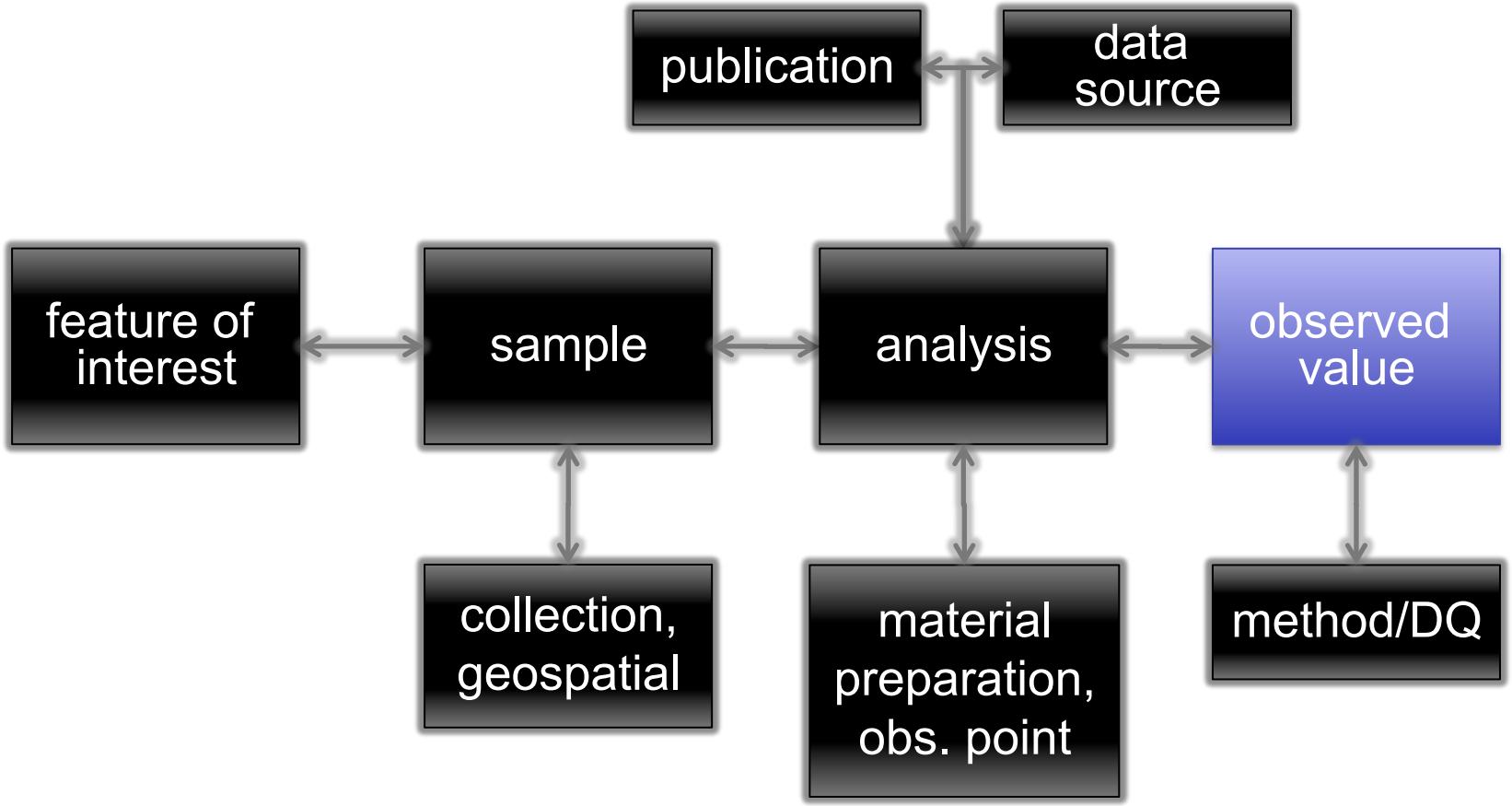


Diagram Key
Table
(M) Field cannot be null
(O) Field can be Null
Controlled Vocabulary
(M) Field cannot be null
(O) Field can be Null
Cardinality
1..1 One and only one
1..* One or many
0..* Zero or many
0..1 Zero or one
ODM Version
ODMVersion
(M) Version Number
Data Sources
Sources
(M) SourceID {PK}
(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
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
Locate
Observation Values
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}
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
Analyze
Samples
(M) SampleID {PK}
(M) SampleType
(M) LabSampleCode
(M) LabMethodID {FK}
DefinesCV
SampleTypeCV
(M) Term
(O) Definition
Offsets
OffsetTypes
(M) OffsetTypeID {PK}
(M) OffsetUnitsID {FK}
(M) OffsetDescription
OffsetLocationOf
Characterize
Represent
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
Value Grouping
Groups
(M) GroupID {FK}
(M) ValueID {FK}
Describe
GroupDescriptions
(M) GroupID {PK}
(O) GroupDescription
Group
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) 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

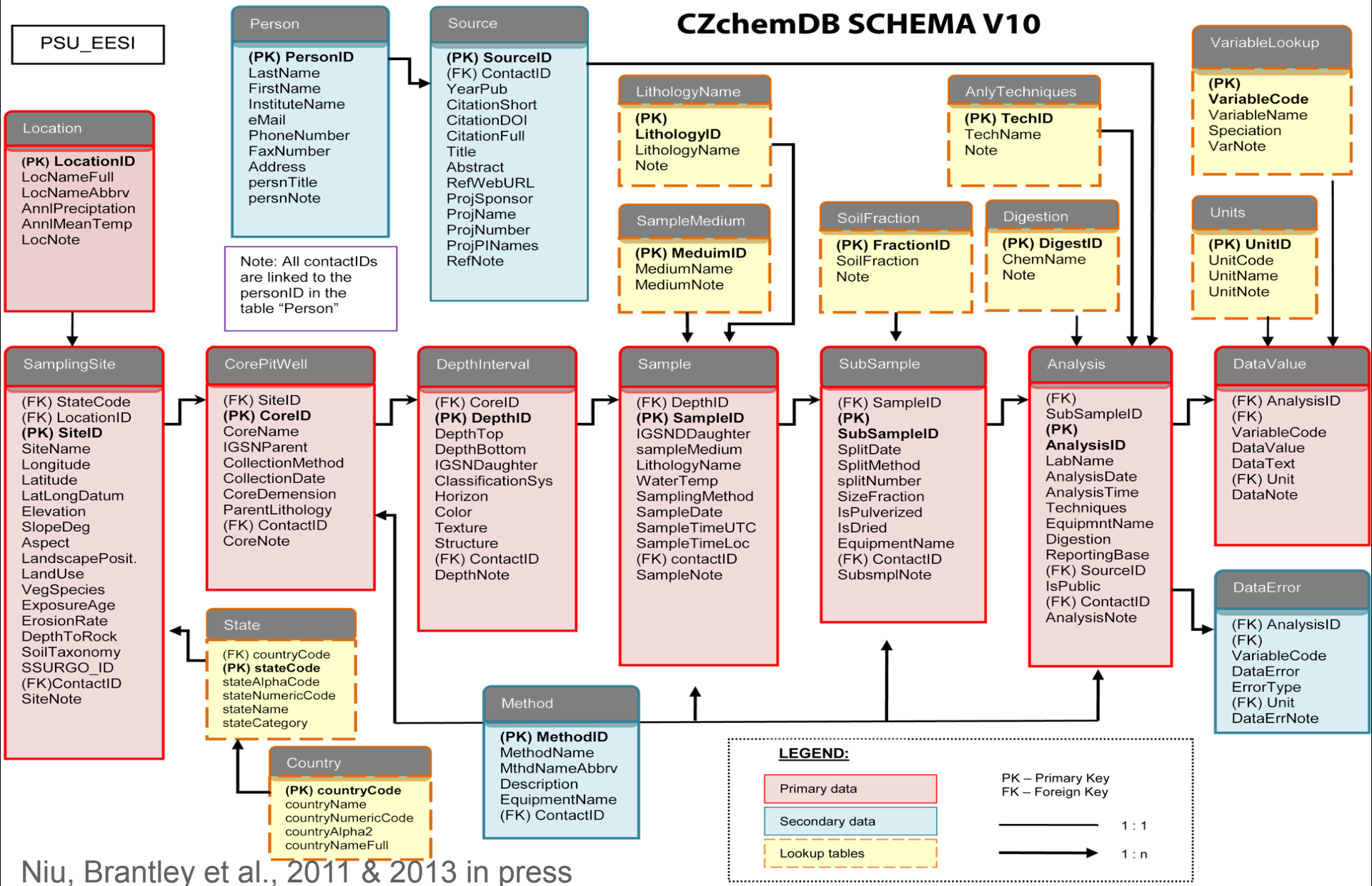
General GeoChemical Data Model



EarthChem

IEDA
INTEGRATED EARTH DATA APPLICATIONS

CZchemDB: the CZO Chemistry Database



Niu, Brantley et al., 2011 & 2013 in press

CZO Disciplines

- Biogeochemistry
- Biology/Ecology
- Biology/Molecular
- Climatology/ Meteorology
- Data Management/ CyberInfrastructure
- Engineering/Method Development
- Geochemistry/ Mineralogy
- Geology/Chronology
- Geomorphology
- Geophysics
- GIS/Remote Sensing
- Hydrology
- Modeling/ Computational Science
- Outreach/ Education Research
- Soil Science/Pedology
- Water Chemistry

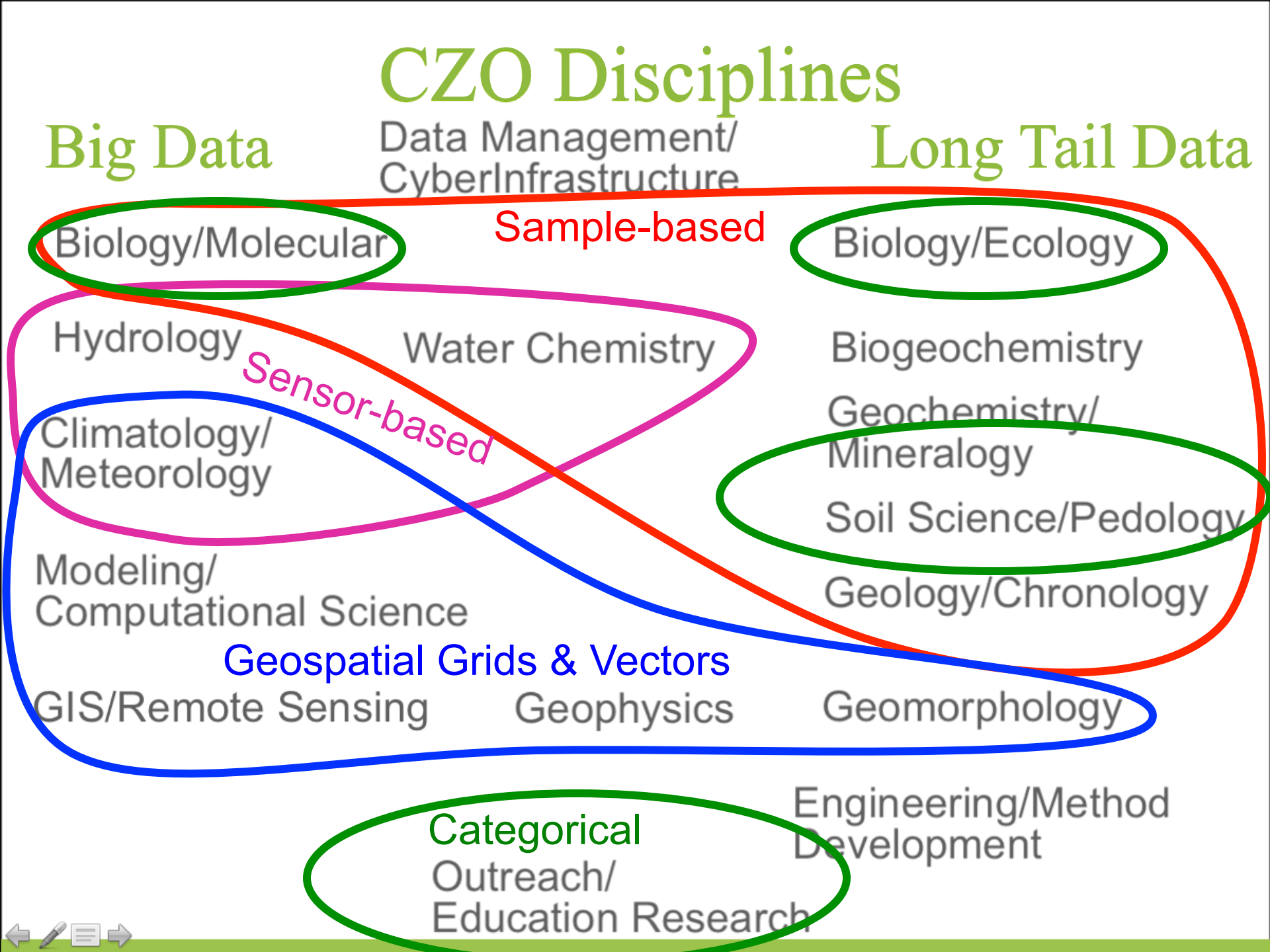
CZO Disciplines

Big Data

- Biogeochemistry
- Biology/Ecology
- Biology/Molecular
- Climatology/ Meteorology
- Data Management/ CyberInfrastructure
- Engineering/Method Development
- Geochemistry/ Mineralogy
- Geology/Chronology

Long Tail Data

- Geomorphology
- Geophysics
- GIS/Remote Sensing
- Hydrology
- Modeling/ Computational Science
- Outreach/ Education Research
- Soil Science/Pedology
- Water Chemistry


CZO Disciplines
Big Data
Data Management/
CyberInfrastructure
Long Tail Data
Sample-based
Biology/Molecular
Biology/Ecology
Hydrology
Water Chemistry
Biogeochemistry
Sensor-based
Geochemistry/
Mineralogy
Climatology/
Meteorology
Soil Science/Pedology
Modeling/
Computational Science
Geology/Chronology
Geospatial Grids & Vectors
GIS/Remote Sensing
Geophysics
Geomorphology
Engineering/Method
Development
Categorical
Outreach/
Education Research

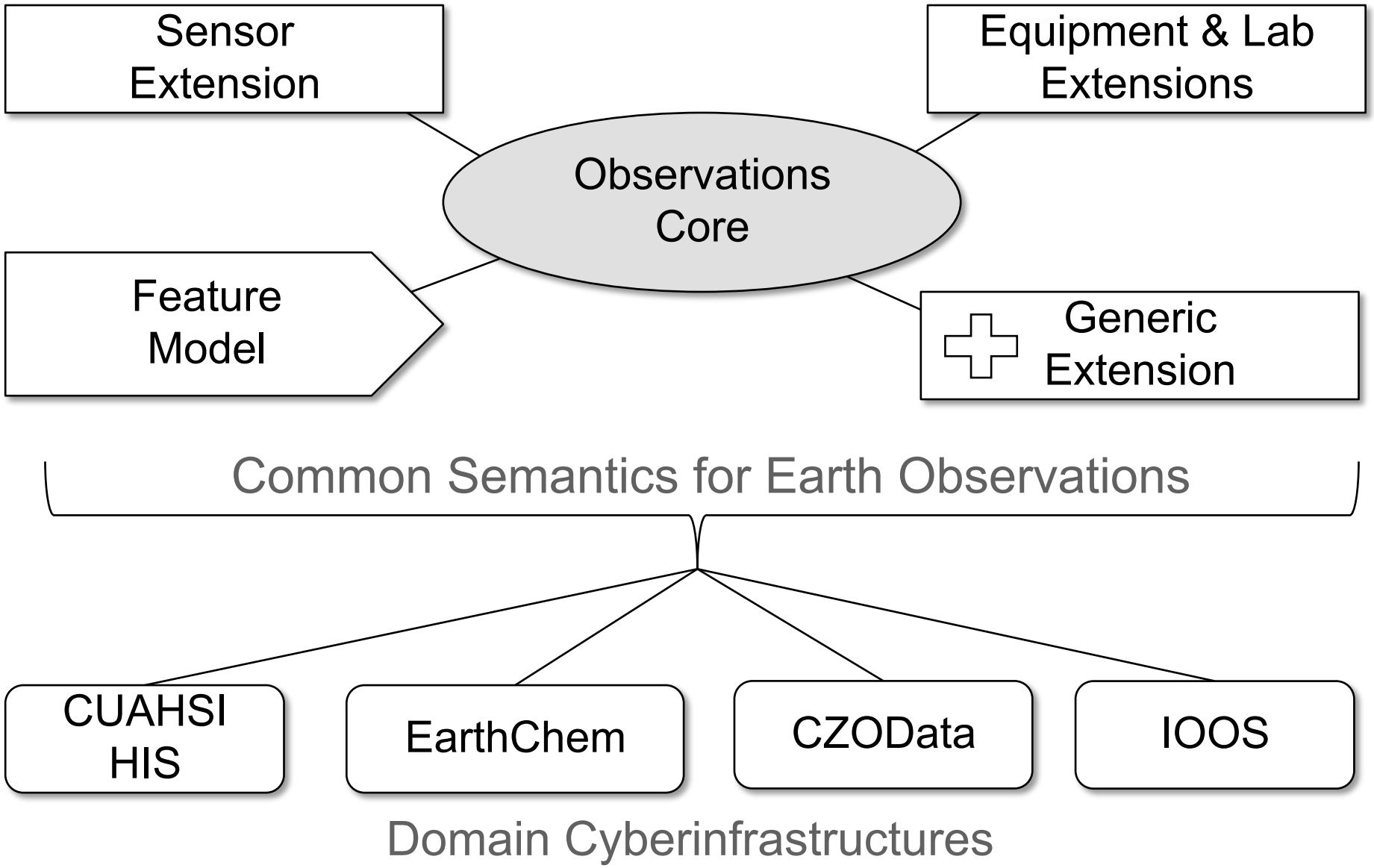
Dublin Core Metadata Element Set

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

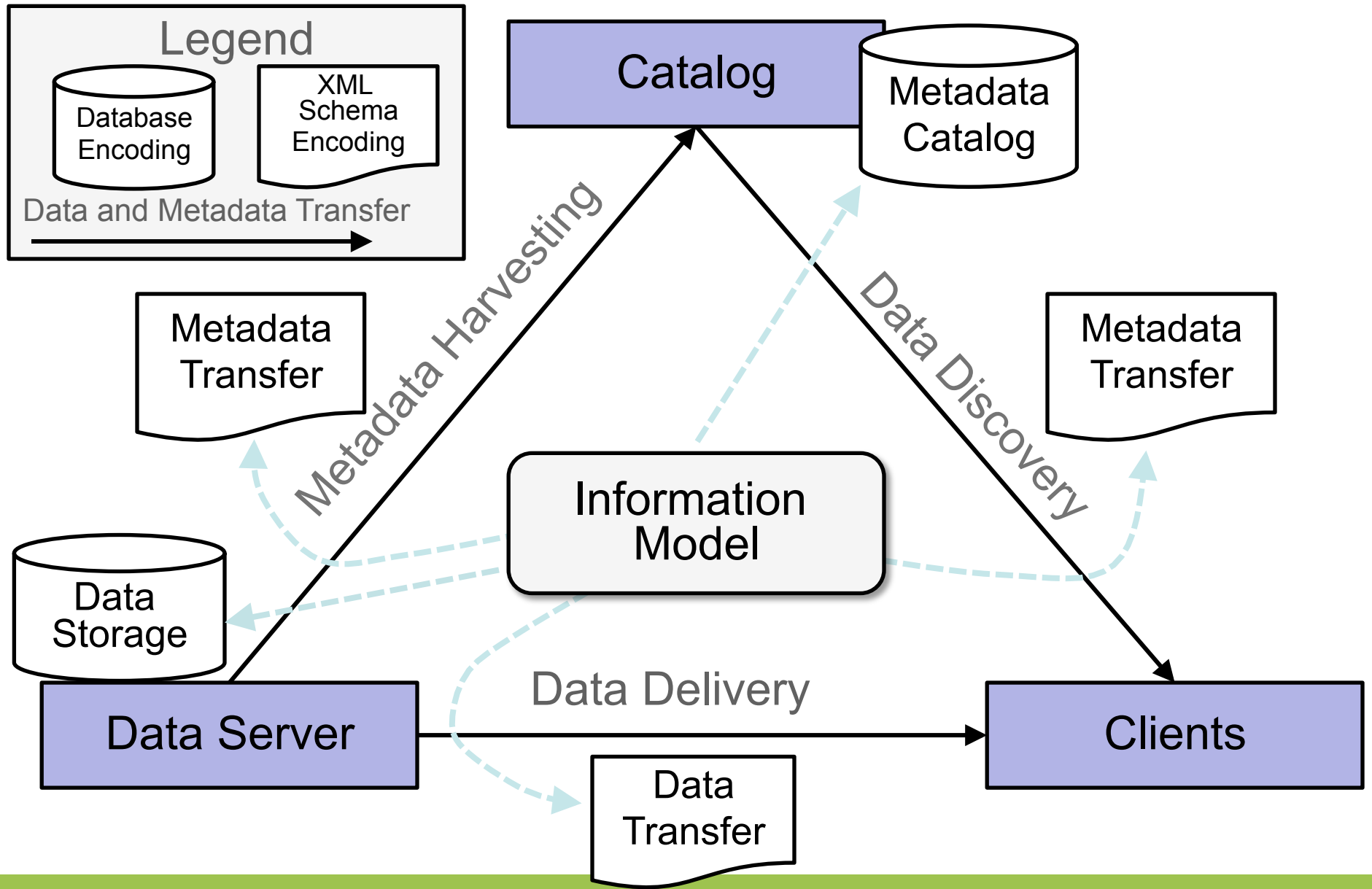
Darwin Core adds:

- Taxon, Occurrence, Geological Context, [Measurement]

ODM2: Common to Most Data Types



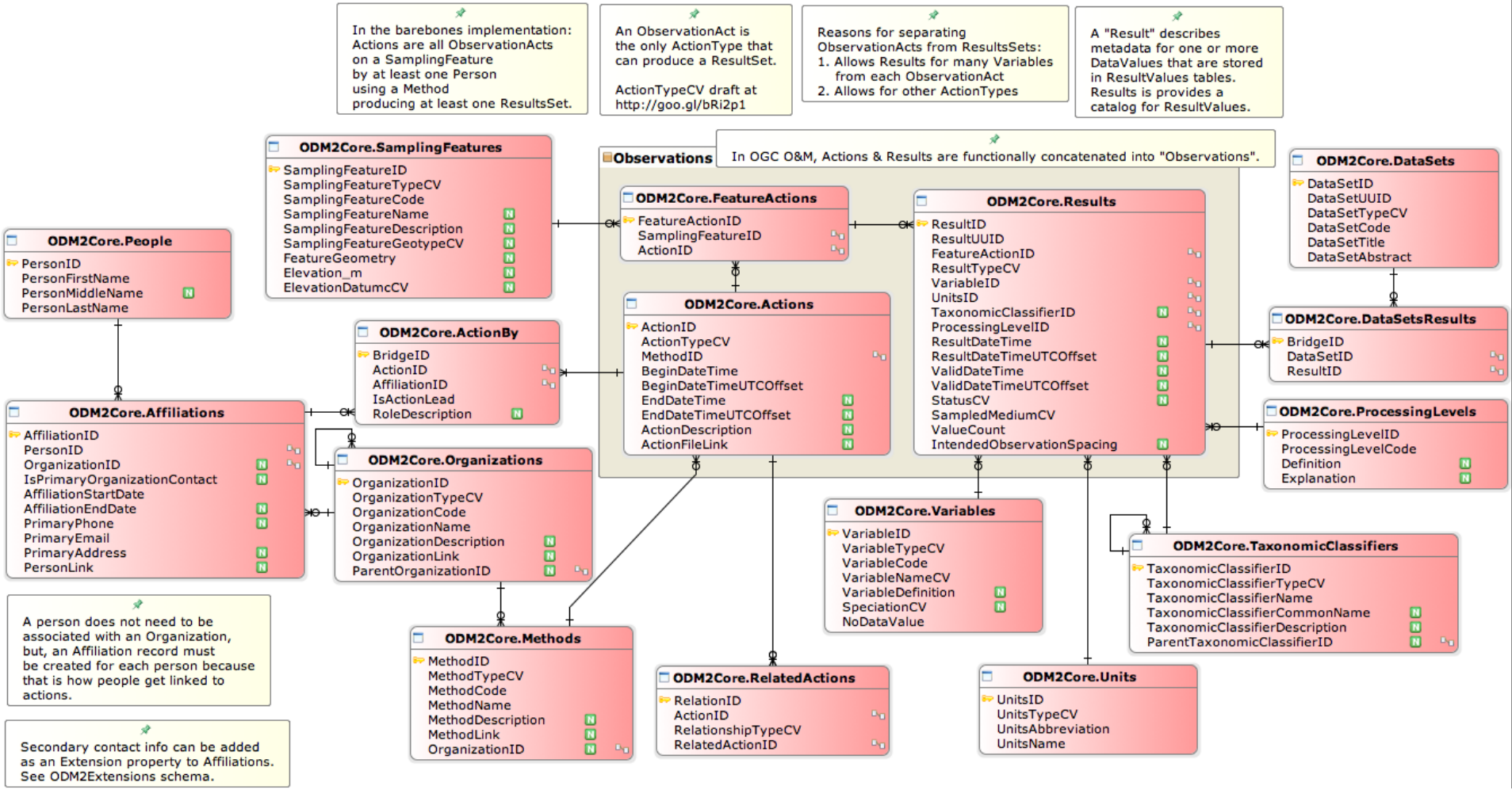
ODM2: Common to All Components



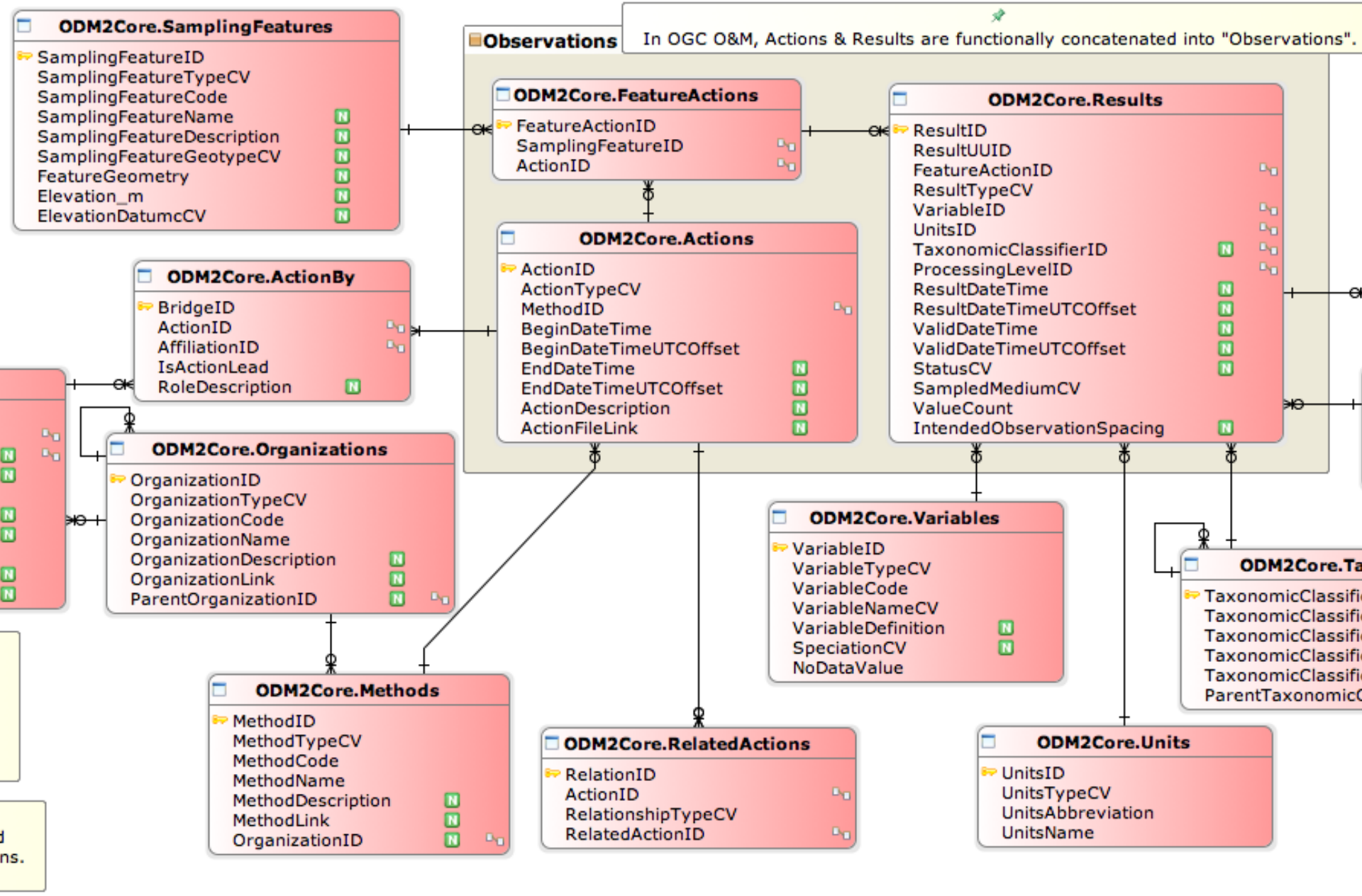
ODM2: Additional Goals

- Driven by Community & Use Cases:
  - 3 workshops + ~12 data models + much feedback
  - use cases: CZOData, Little Bear River, PetDB, IOOS
- Balance between general vs. understandable
- External unique identifiers, vocabularies & taxonomies
- Rich Specimen, Site & other Sampling Features
- Granular Methods, Data Quality & Equipment
- Dataset publishing & archiving via:
  - Result “packages”, Versions, Citations, Provenance
- Strong Annotations & general extensibility

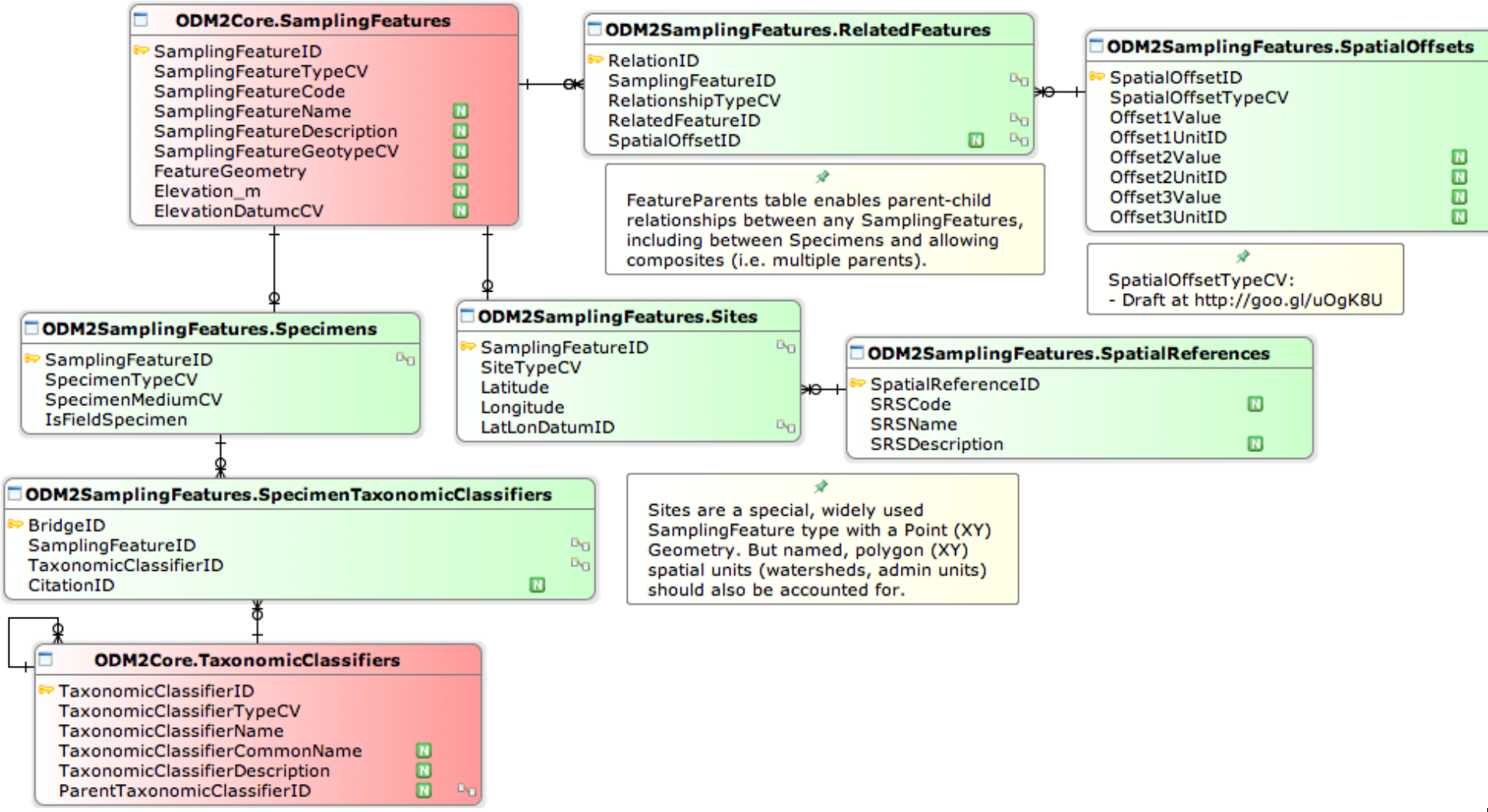
ODM2Core


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 ResultSet.
ActionTypeCV draft at http://goo.gl/bRi2p1
Reasons for separating ObservationActs from ResultsSets:
1. Allows Results for many Variables from each ObservationAct
2. Allows for other ActionTypes
A "Result" describes metadata for one or more DataValues that are stored in ResultValues tables. Results is provides a catalog for ResultValues.
Observations
In OGC O&M, Actions & Results are functionally concatenated into "Observations".
ODM2Core.SamplingFeatures
SamplingFeatureID
SamplingFeatureTypeCV
SamplingFeatureCode
SamplingFeatureName
SamplingFeatureDescription
SamplingFeatureGeotypeCV
FeatureGeometry
Elevation_m
ElevationDatumcCV
ODM2Core.FeatureActions
FeatureActionID
SamplingFeatureID
ActionID
ODM2Core.Results
ResultID
ResultUUID
FeatureActionID
ResultTypeCV
VariableID
UnitsID
TaxonomicClassifierID
ProcessingLevelID
ResultDateTime
ResultDateTimeUTCOffset
ValidDateTime
ValidDateTimeUTCOffset
StatusCV
SampledMediumCV
ValueCount
IntendedObservationSpacing
ODM2Core.DataSets
DataSetID
DataSetUUID
DataSetTypeCV
DataSetCode
DataSetTitle
DataSetAbstract
ODM2Core.People
PersonID
PersonFirstName
PersonMiddleName
PersonLastName
ODM2Core.ActionBy
BridgeID
ActionID
AffiliationID
IsActionLead
RoleDescription
ODM2Core.Actions
ActionID
ActionTypeCV
MethodID
BeginDateTime
BeginDateTimeUTCOffset
EndDateTime
EndDateTimeUTCOffset
ActionDescription
ActionFileLink
ODM2Core.DataSetsResults
BridgeID
DataSetID
ResultID
ODM2Core.Affiliations
AffiliationID
PersonID
OrganizationID
IsPrimaryOrganizationContact
AffiliationStartDate
AffiliationEndDate
PrimaryPhone
PrimaryEmail
PrimaryAddress
PersonLink
ODM2Core.Organizations
OrganizationID
OrganizationTypeCV
OrganizationCode
OrganizationName
OrganizationDescription
OrganizationLink
ParentOrganizationID
ODM2Core.ProcessingLevels
ProcessingLevelID
ProcessingLevelCode
Definition
Explanation
ODM2Core.Variables
VariableID
VariableTypeCV
VariableCode
VariableNameCV
VariableDefinition
SpeciationCV
NoDataValue
ODM2Core.TaxonomicClassifiers
TaxonomicClassifierID
TaxonomicClassifierTypeCV
TaxonomicClassifierName
TaxonomicClassifierCommonName
TaxonomicClassifierDescription
ParentTaxonomicClassifierID
A person does not need to be associated with an Organization, but, an Affiliation record must be created for each person because that is how people get linked to actions.
ODM2Core.Methods
MethodID
MethodTypeCV
MethodCode
MethodName
MethodDescription
MethodLink
OrganizationID
ODM2Core.RelatedActions
RelationID
ActionID
RelationshipTypeCV
RelatedActionID
ODM2Core.Units
UnitsID
UnitsTypeCV
UnitsAbbreviation
UnitsName
Secondary contact info can be added as an Extension property to Affiliations. See ODM2Extensions schema.

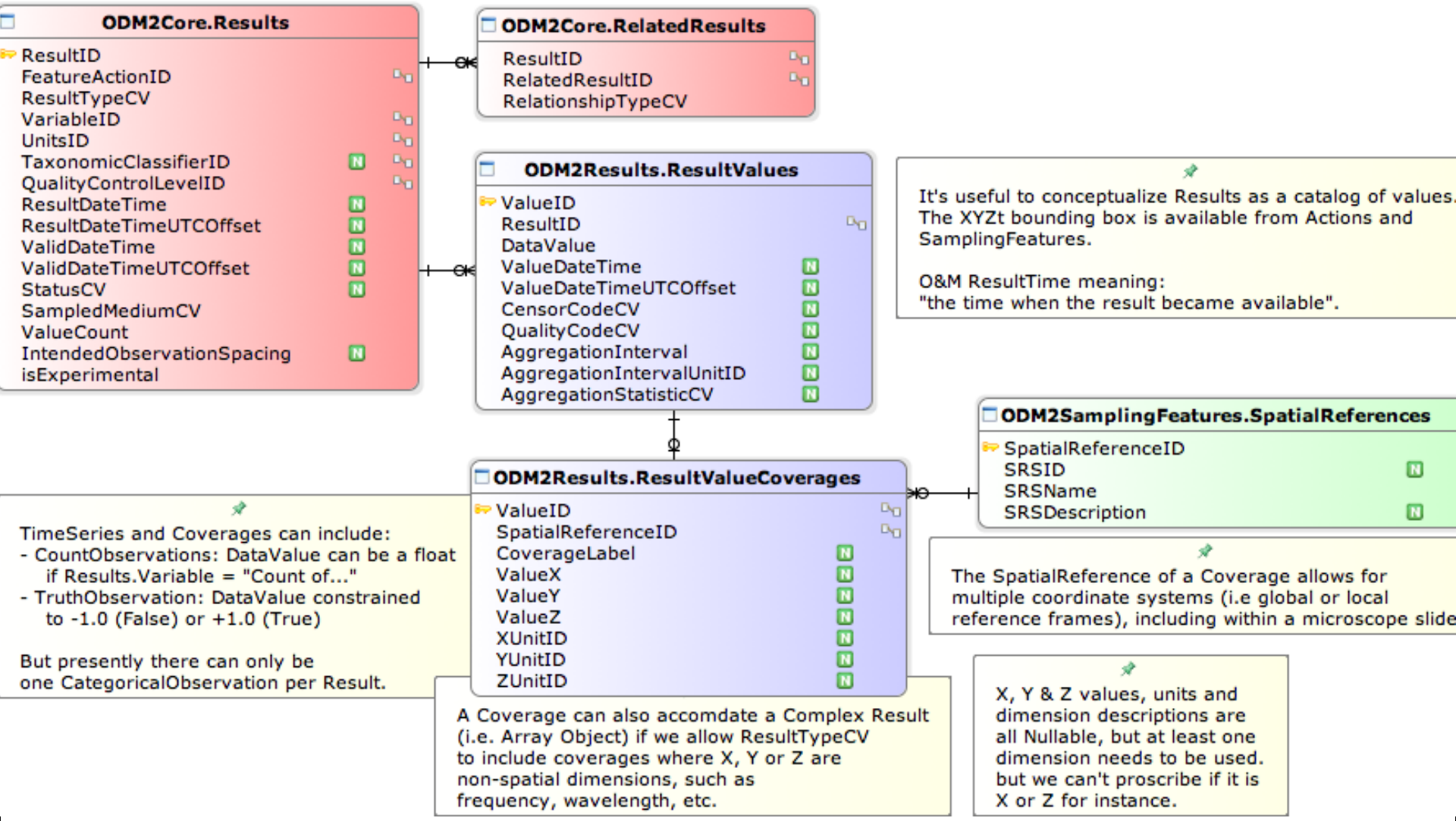
ODM2Core


ODM2Core.SamplingFeatures
SamplingFeatureID
SamplingFeatureTypeCV
SamplingFeatureCode
SamplingFeatureName
SamplingFeatureDescription
SamplingFeatureGeotypeCV
FeatureGeometry
Elevation_m
ElevationDatumCV
Observations
In OGC O&M, Actions & Results are functionally concatenated into "Observations".
ODM2Core.FeatureActions
FeatureActionID
SamplingFeatureID
ActionID
ODM2Core.Results
ResultID
ResultUUID
FeatureActionID
ResultTypeCV
VariableID
UnitsID
TaxonomicClassifierID
ProcessingLevelID
ResultDateTime
ResultDateTimeUTCOffset
ValidDateTime
ValidDateTimeUTCOffset
StatusCV
SampledMediumCV
ValueCount
IntendedObservationSpacing
ODM2Core.Actions
ActionID
ActionTypeCV
MethodID
BeginDateTime
BeginDateTimeUTCOffset
EndDateTime
EndDateTimeUTCOffset
ActionDescription
ActionFileLink
ODM2Core.ActionBy
BridgeID
ActionID
AffiliationID
IsActionLead
RoleDescription
ODM2Core.Organizations
OrganizationID
OrganizationTypeCV
OrganizationCode
OrganizationName
OrganizationDescription
OrganizationLink
ParentOrganizationID
ODM2Core.Variables
VariableID
VariableTypeCV
VariableCode
VariableNameCV
VariableDefinition
SpeciationCV
NoDataValue
ODM2Core.Methods
MethodID
MethodTypeCV
MethodCode
MethodName
MethodDescription
MethodLink
OrganizationID
ODM2Core.RelatedActions
RelationID
ActionID
RelationshipTypeCV
RelatedActionID
ODM2Core.Units
UnitsID
UnitsTypeCV
UnitsAbbreviation
UnitsName

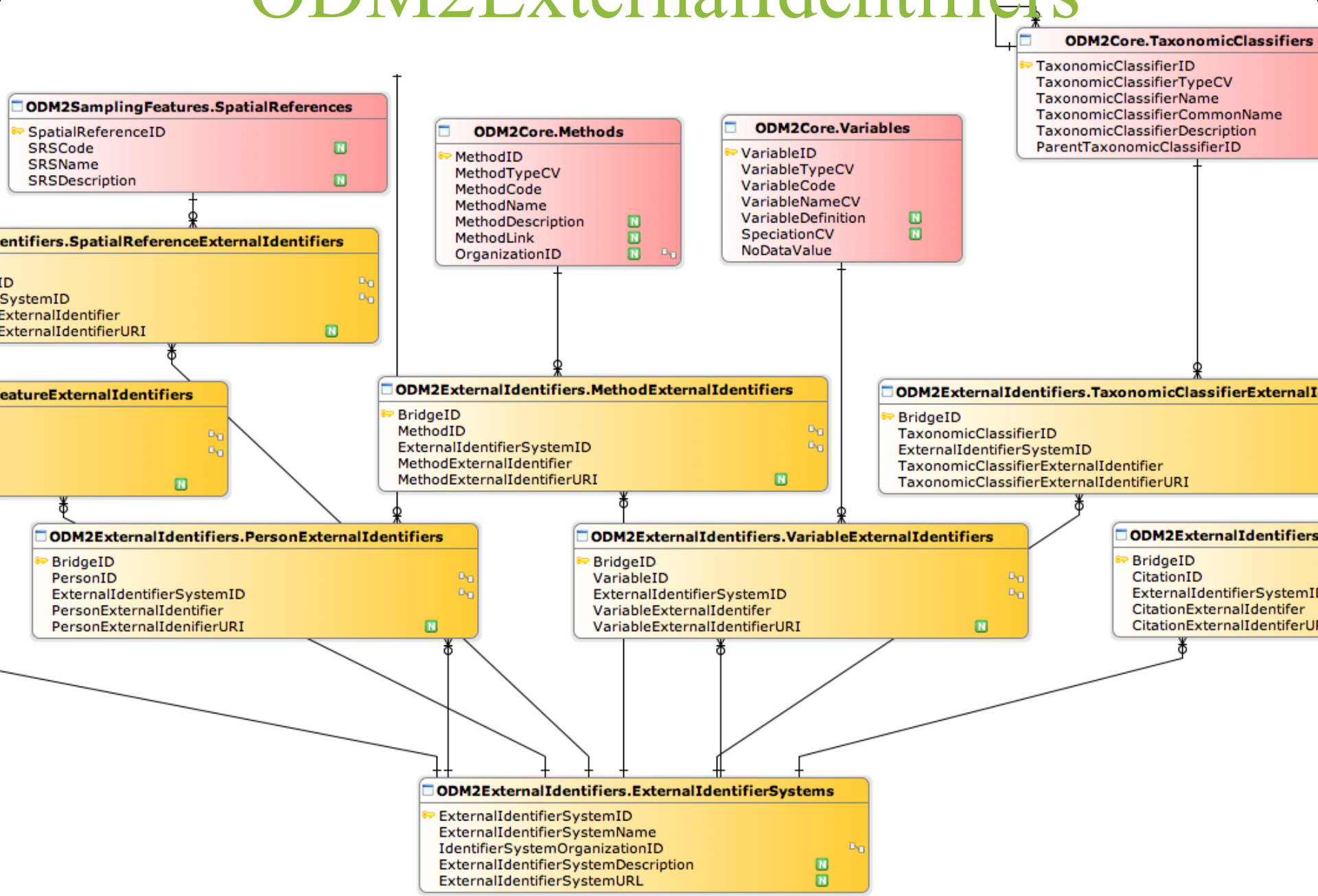
ODM2SamplingFeatures



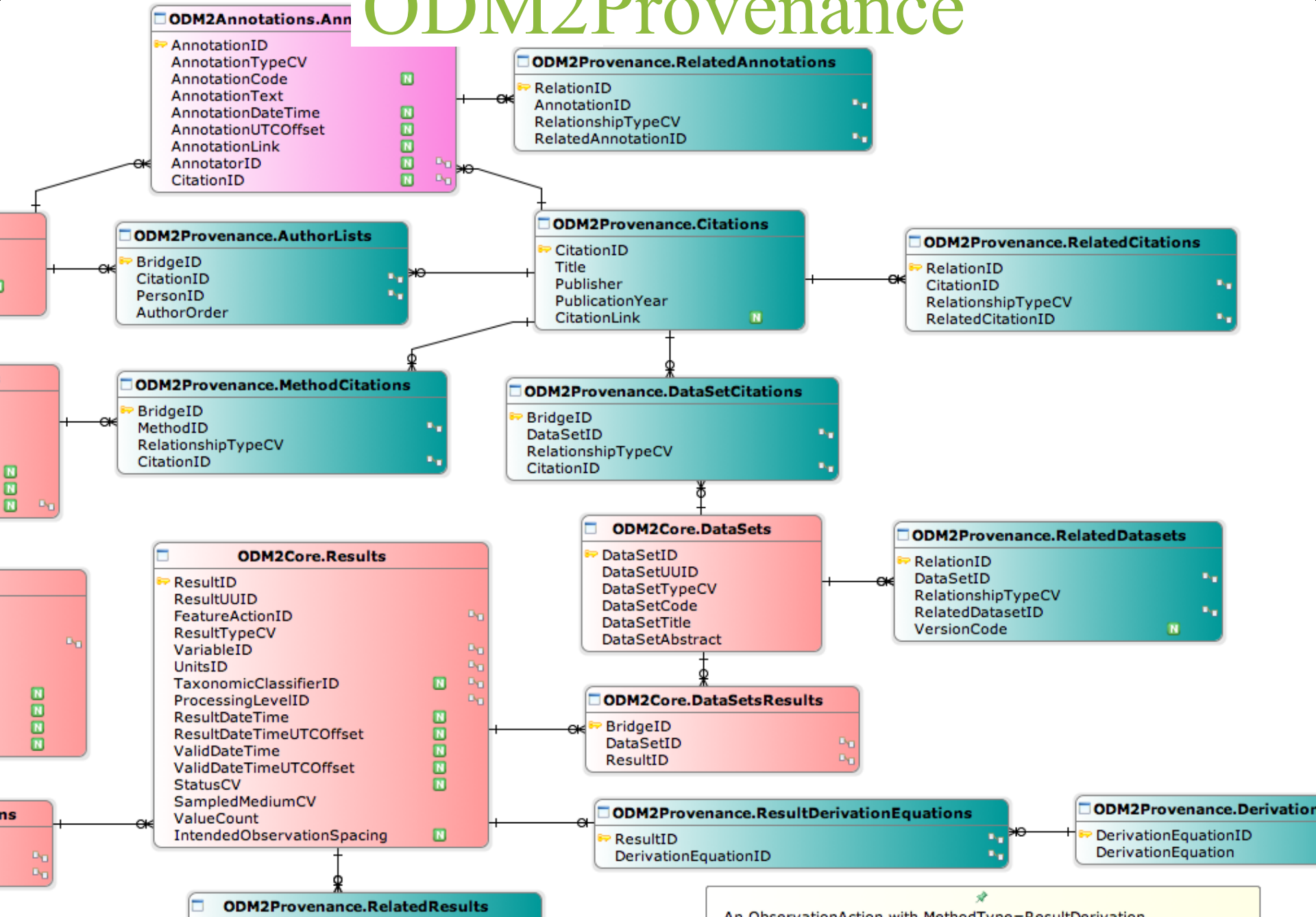
ODM2Results



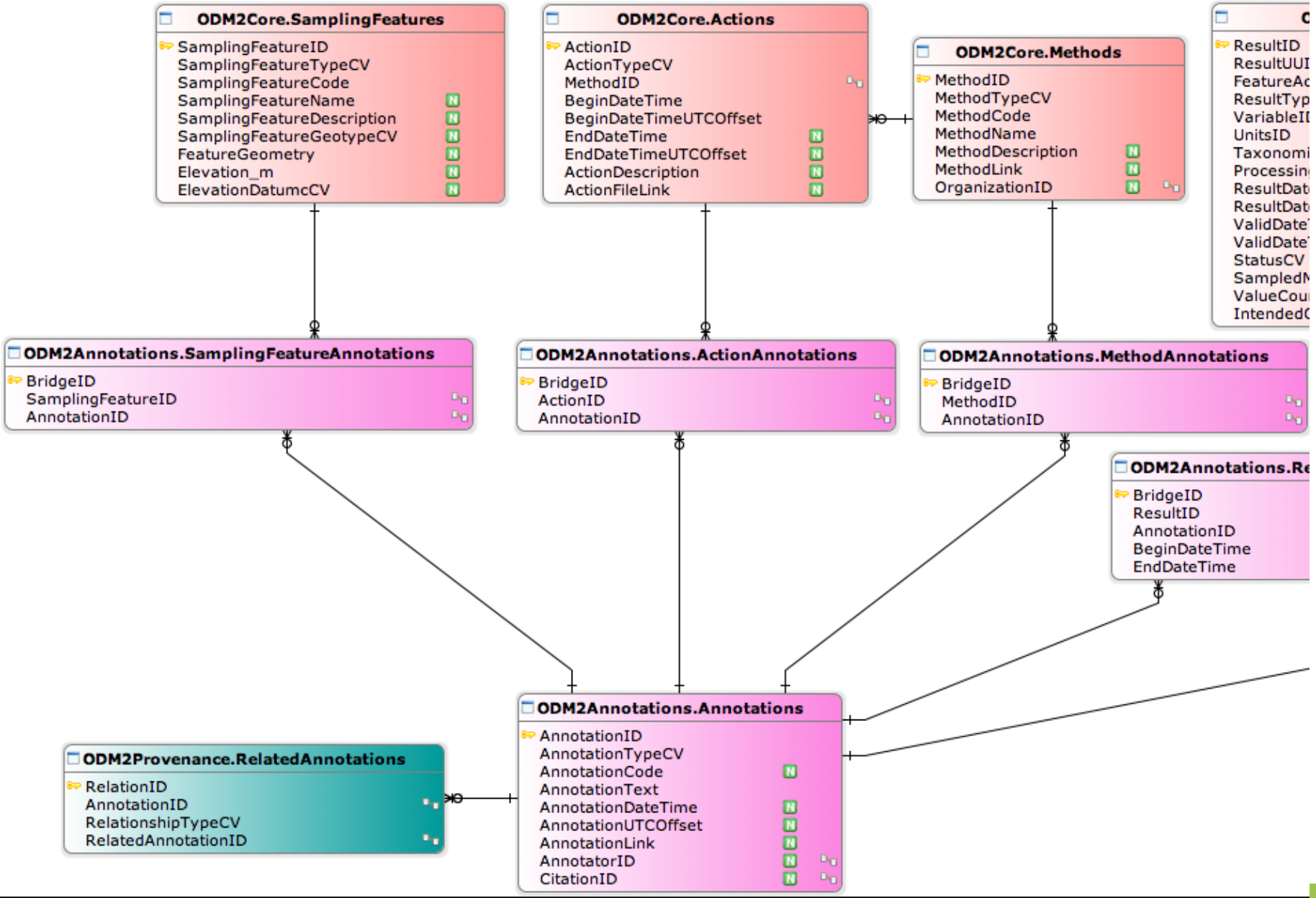
ODM2ExternalIdentifiers



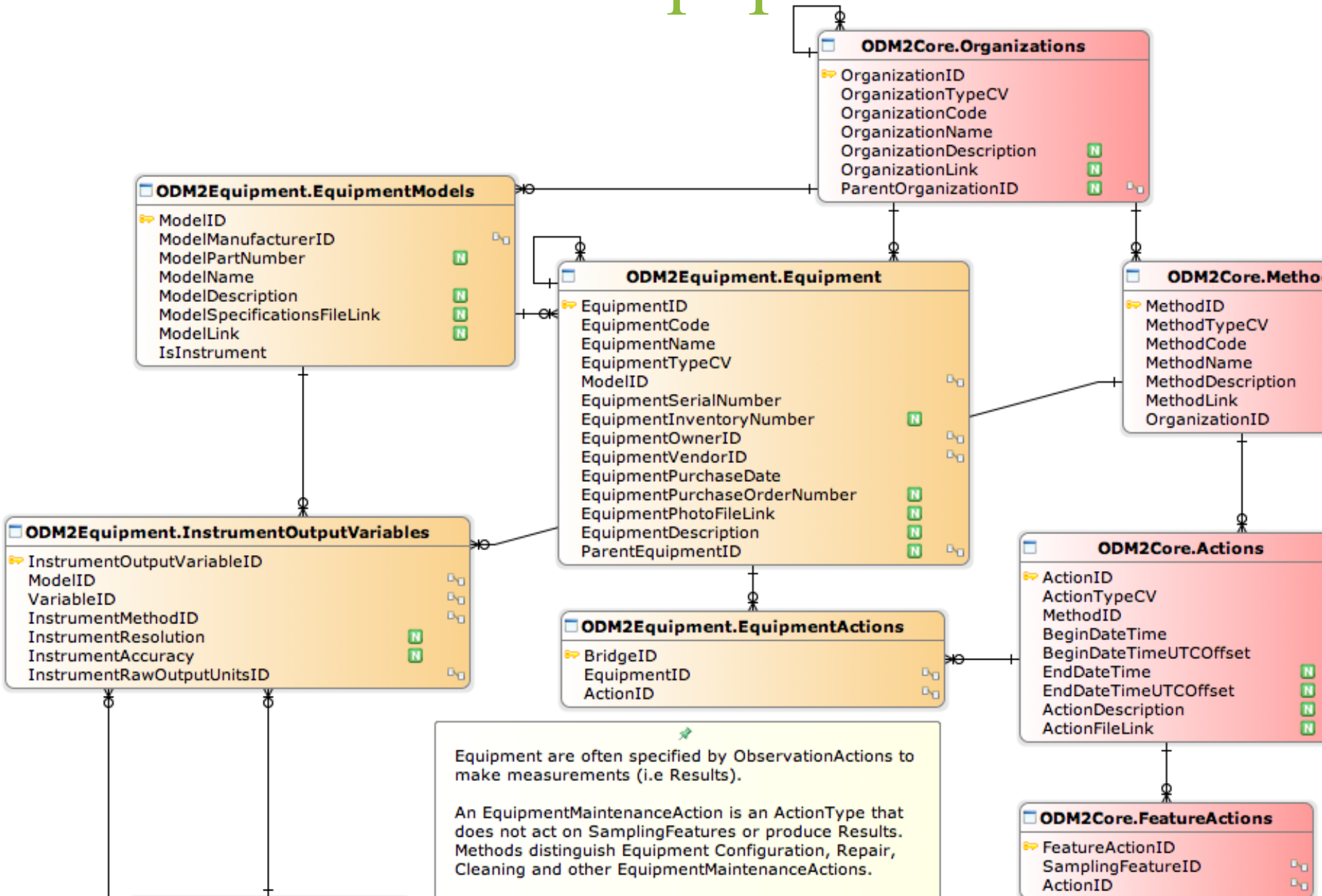
ODM2Provenance



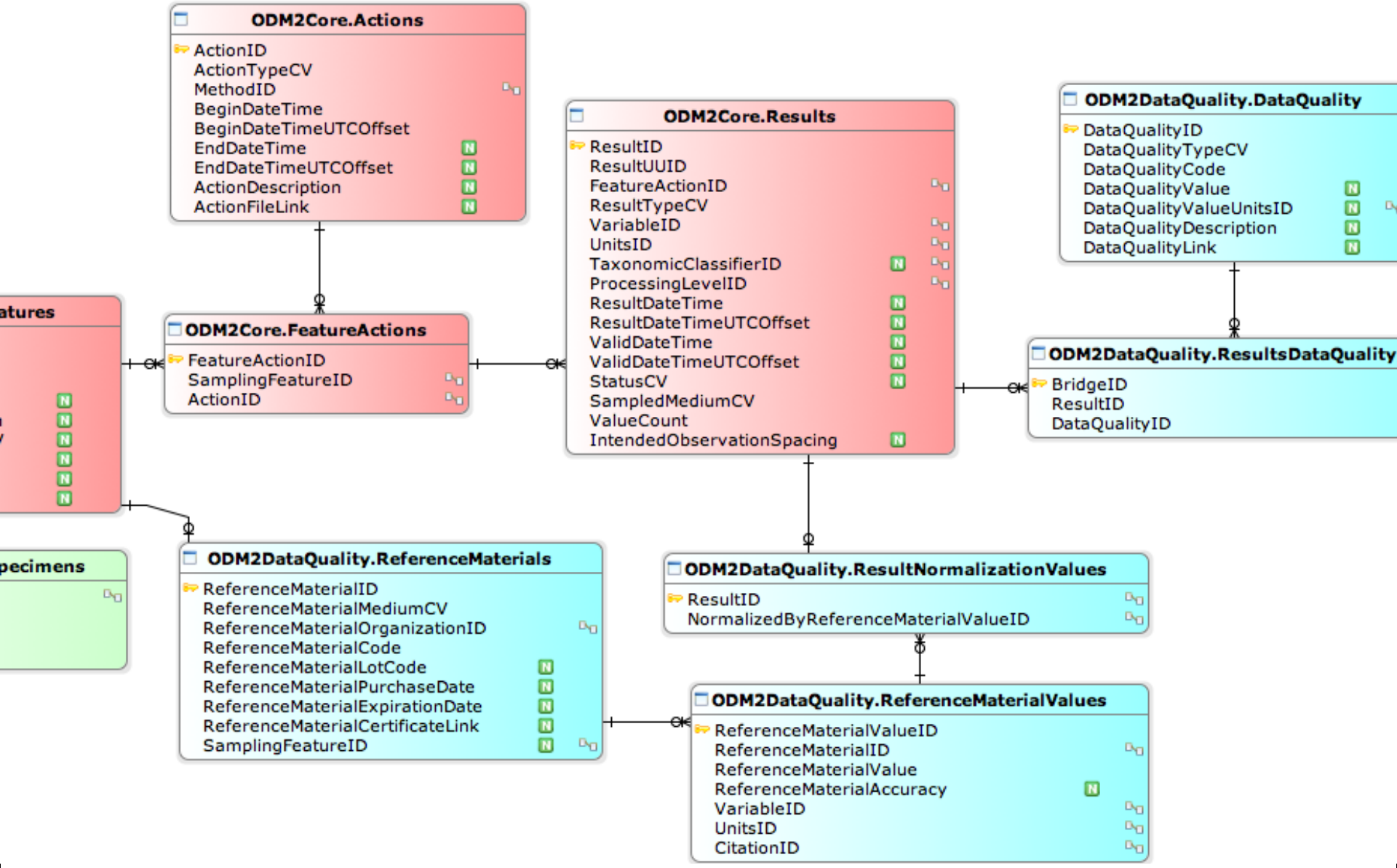
ODM2Annotations



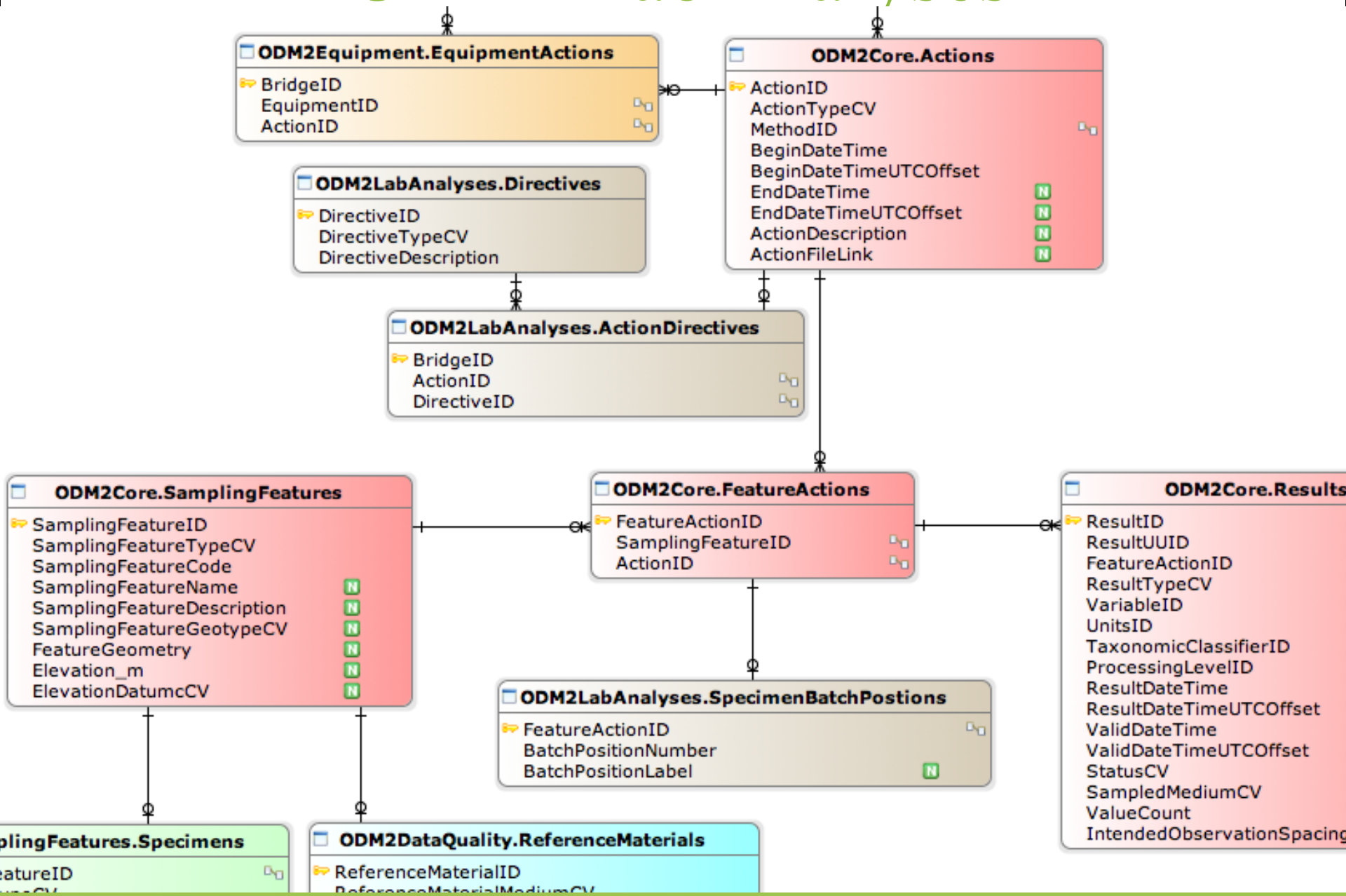
ODM2Equipment



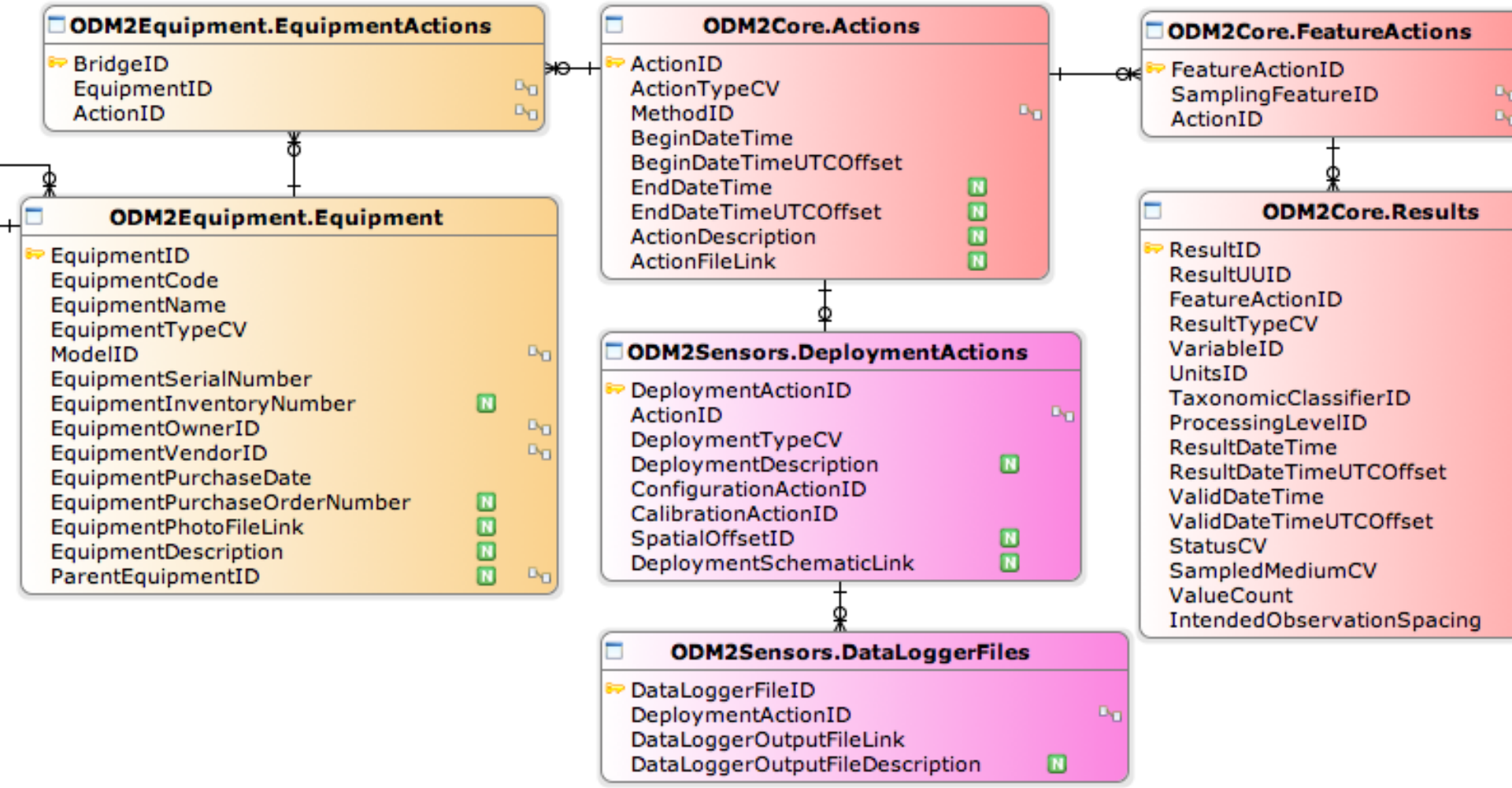
ODM2DataQuality



ODM2LabAnalyses



ODM2Sensors



Little Bear River Sensor Network

- 7 water quality and streamflow monitoring sites
  - Temperature
  - Dissolved Oxygen
  - pH
  - Specific Conductance
  - Turbidity
  - Water level/discharge
- 4 weather stations
  - Air Temperature
  - Relative Humidity
  - Solar radiation
  - Precipitation
  - Barometric Pressure
  - Wind speed and direction
  - Soil moisture and temperature at 5 depths
- Spread spectrum radio telemetry network

Support: CBET 0610075

Horsburgh, J.S., Spackman Jones, A., Stevens, D.K., Tarboton, D.G., Mesner, N.O., (2010), A sensor network for high frequency estimation of water quality consitutent fluxes using surrogates, *Environmental Modelling & Software*, 25, 1031-1044, doi:10.1016/j.envsoft.2009.10.012.

Christina River Basin CZO: Sediment Fingerprint

1. Field collection of suspended materials, landscape materials, precipitation
2. Analyze for several chemical signatures
3. Use data in a statistical model to determine fraction of stream material from various sources

Karwan, Pizzuto, Aalto, Aufdenkampe et al.

Particulate Analyses:

- Fallout radionuclides (7Be, 137Cs, 210Pb)
- C, N, stable isotopes
- Mineral surface area
- Grain size
- Mineral content (elemental and composition)
- Microbial community characterization (16S, rRNA, PCR, ITS-DGGE, sequencing)

Streamwater Chemistry:

- TSS
- DOC
- Anions + Cations
- Others?

Sediment & Soil Processing:

Characterization of organic & inorganic bulk properties

Field Data Collection

Notebook:
Site ID, Date, Time, Site Notes
Sample list & descriptions

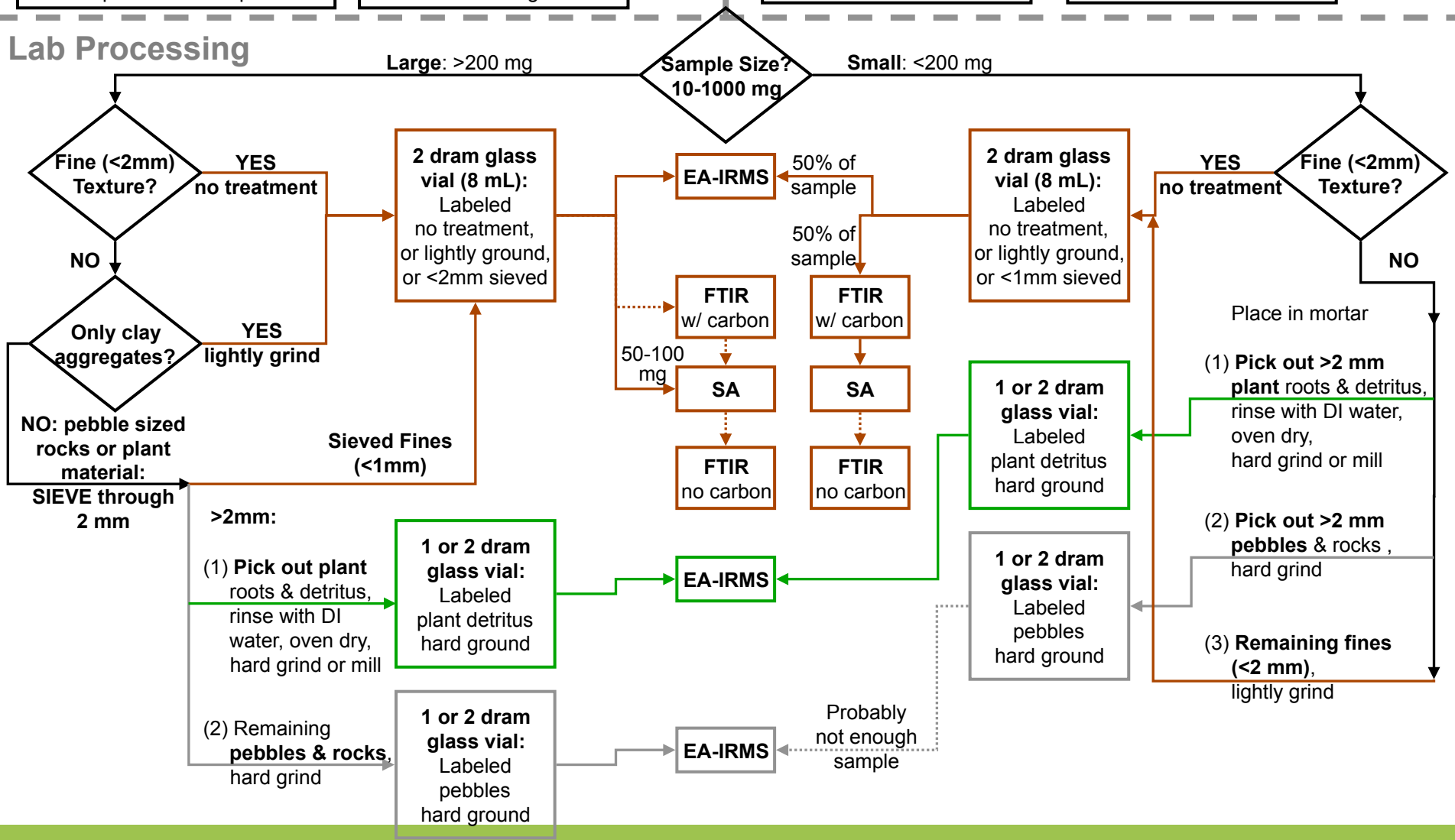
GPS:
Waypoint: Lat., Lon., Elev.
Track Log

Field Sample Collection

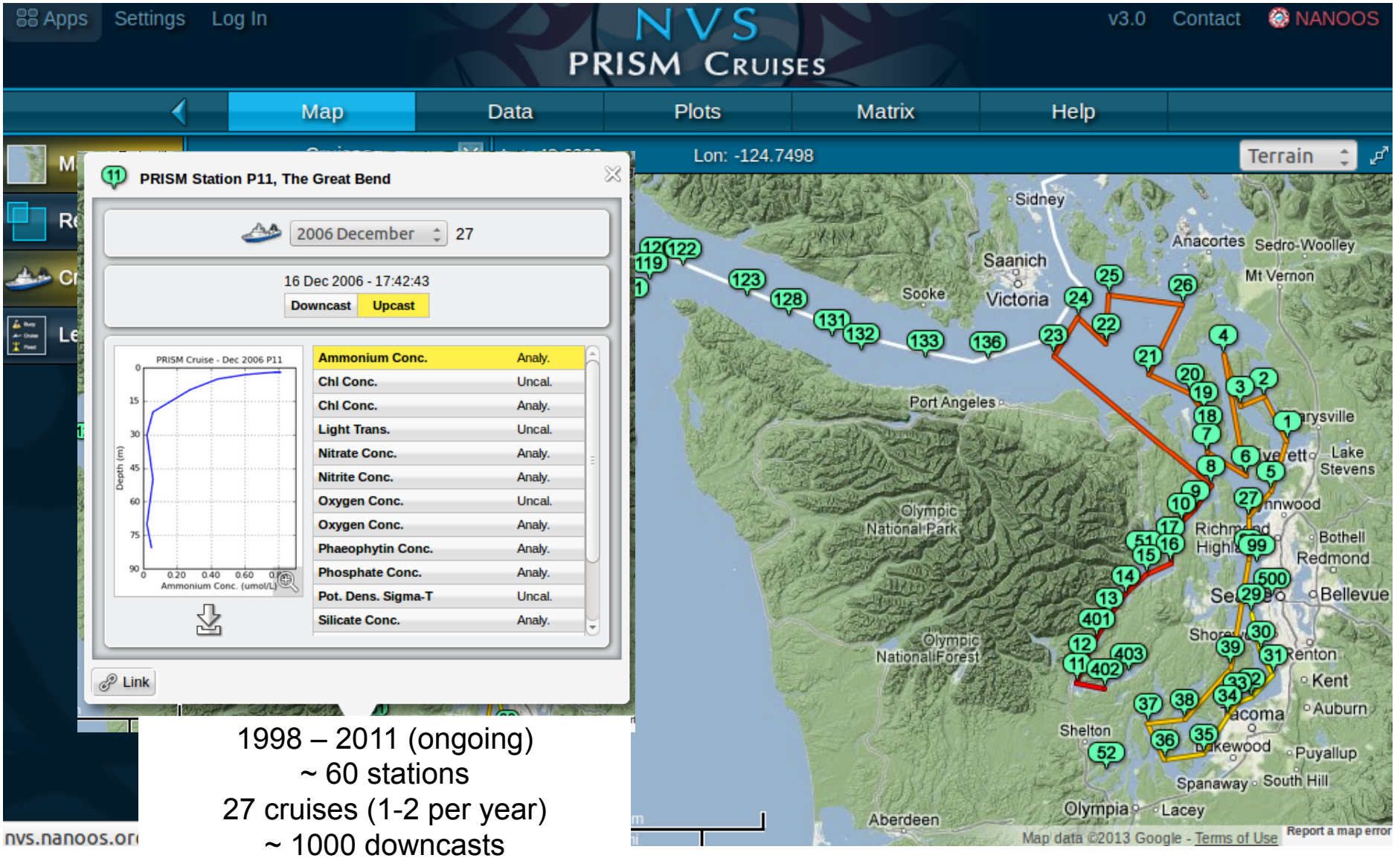
By shovel or auger:
100-300 g in ziplock bag

By core:
Plastic core tube sectioned and sub-sampled in lab

Lab Processing



Oceanographic data (depth profiles at repeat stations): "PRISM" Cruises, Puget Sound



1998 – 2011 (ongoing)
~ 60 stations
27 cruises (1-2 per year)
~ 1000 downcasts
~ 700 upcasts & bottle collections

Emilio Mayorga, University of Washington / NANOOS, 2013-3-12

EarthChem Use Cases

- Volcanic Gas Emissions
  - DECADE project as part of the Deep Carbon Observatory
  - Collaboration with DCO Data Science (Peter Fox, RPI), Global Volcanism Program (Liz Cottrell, Smithsonian), Tobias Fischer & Laura Clor at UNM, and international investigators in the DCO
- Composition and fluxes of volcanic gases
  - CO2, SO2, CH4, etc.

NSF Scientific Software Integration

BiG CZ SSI project (2014-2015):

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

Other ODM2 Adopters

- CUAHSI Water Data Center
- CUAHSI Hydroshare
- Interest by:
  - DataOne and ecoinformatics community
  - Open Geospatial Consortium (OGC) for "Specimen" profile of "Observations & Measurements"

Thank You

Funded by the
National Science Foundation
EAR 1224638
EAR 1332257
ACI 1339834

ODM2 is on GitHUB:
https://github.com/UCHIC/ODM2