# Cyberinfrastructure Tools for Managing GAMUT Data and Infrastructure, Part 1

10/3/2013

Amber Spackman Jones

Jeffery S Horsburgh

Logan River Sites

Red Butte Sites

Provo River Sites

## MDF LoggerNet Server

Base Station

LoggerNet Backup

ODM Streaming Data Loader

.dat Datalogger Text Files

Backup Script

Backup Script

## MDF Database Server

Automated QC Checks

Logan ODM Raw Data

QC Data

Automated QC Checks

Red Butte ODM Raw Data

QC Data

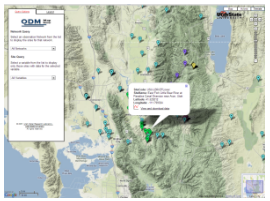
Automated QC Checks

Provo ODM Raw Data

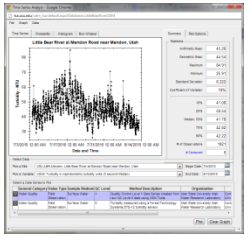
QC Data

Staging of Current Conditions

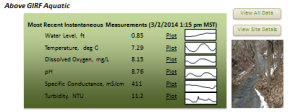
## MDF Web Server



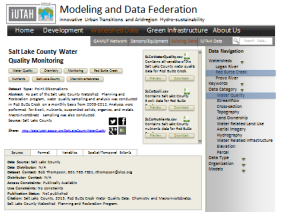
Map Based Display



Time Series Display (and other plot types)



Current Conditions



File-based Archival and Publication

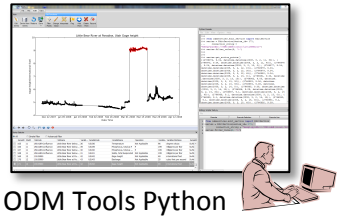
Archival of original .dat files

Backup of ODM Databases

UofU Data Store

Virtual Machine Backups

## Visualization and QC of Sensor Data Stream



ODM Tools Python

# Datalogger Program/Datalogger File Management

- Protocol document on CloudShare shared GAMUT folder: cloudshare.iutahepscor.org
- DataloggerFiles folder
  - Watershed → Site → .cr3 files
  - Change log: Name and date of versions, include changes made. One for each site.
- Data folder: prototype of archival file process/ structure
  - Watershed → Site → .dat files
  - Include date range in file name
  - Maybe should also have log file?
  - Also included metadata spreadsheets

## Planned Update

**for addition/deletion of long term variable, adding tables for adamptive sampling, etc.**

Data for 1-2 time steps may be lost in this process

Technician modifies datalogger program. Saves it as new version.

Technician documents changes and rationale in changelog file.

Uploaded to iUTAH cloud share

Technician notifies Jeff and Amber via email.

Jeff or Amber performs the following on the LoggerNet Server:

- -Downloads latest data from the site.
- -Runs the Streaming Data Loader.
- -Pauses the Streaming Data Loader.
- -Moves the old datalogger file into a backups folder.
- -Sends the new datalogger program to the site
- -Modifies the Streaming Data Loader to account for any changes to the datalogger file.
- -Restarts the Streaming Data Loader.

Jeff or Amber notifies technician that process is complete.

## Urgent Update

**for diagnostic or troubleshooting purposes**

Some data may be lost under this scenario!

iUTAH GAMUT Technicians

iUTAH CI Team

Technician modifies datalogger program. Saves it as a diagnostic version.

Technician documents changes and rationale in changelog file.

Uploaded to iUTAH cloud share

Technician notifies Jeff and Amber via email.

If available, Jeff or Amber will:
-Download data to Loggernet Server just prior to diagnostic program being sent.
-Pause Streaming Data Loader.

If there is no central data download prior to sending new program, all data not yet loaded to database will be lost.

Technician sends diagnostic program to site.

**Troubleshooting/ Diagnostic Period**

Site is taken offline. It is assumed that any data collected during this period will not be stored in the database.

Technician downloads data resulting from diagnostic program if it is to be saved.

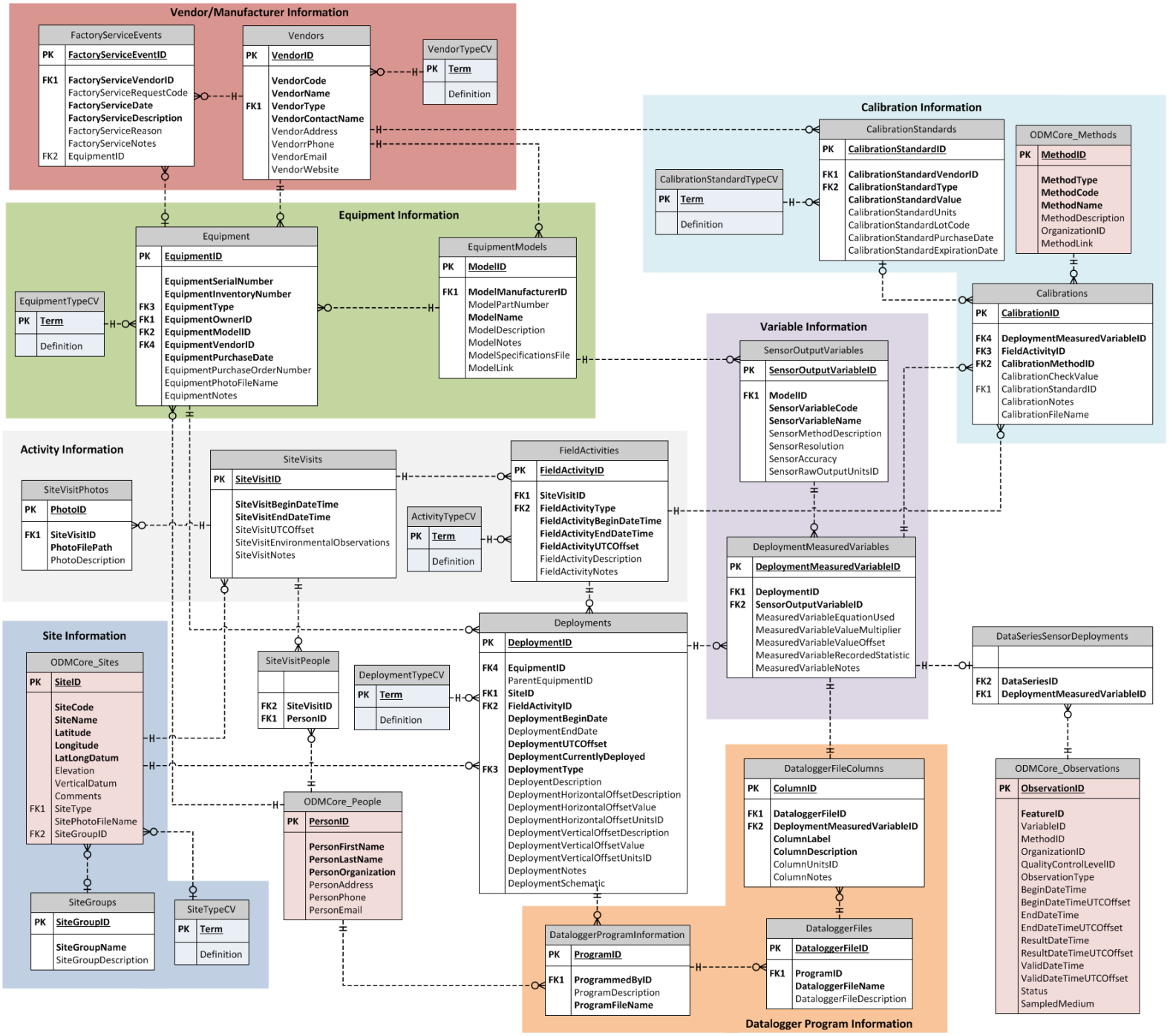
Technician sends regular program to the site.

Technician notifies Jeff and Amber that site is back to normal operation.

Jeff or Amber restart the Streaming Data Loader for the site.

# Equipment Management

- http://data.iutahepscor.org/gamutmanagement/
- Credentials
- Web entry focused on:
  - Equipment: should (mostly) be loaded already
  - Activities:
    - Factory Service Events
    - Site Visits → Field Activities →
    - Calibrations, Deployments
  - Can also add information for Sites, Vendors, and People
- Compile list of issues/questions/feature requests (probably eventually use a Google document)


Vendor/Manufacturer Information
FactoryServiceEvents
PK FactoryServiceEventID
FK1 FactoryServiceVendorID
FactoryServiceRequestCode
FactoryServiceDate
FactoryServiceDescription
FactoryServiceReason
FactoryServiceNotes
FK2 EquipmentID
Vendors
PK VendorID
VendorCode
VendorName
FK1 VendorType
VendorContactName
VendorAddress
VendorrPhone
VendorEmail
VendorWebsite
VendorTypeCV
PK Term
Definition
Calibration Information
CalibrationStandards
PK CalibrationStandardID
FK1 CalibrationStandardVendorID
FK2 CalibrationStandardType
CalibrationStandardValue
CalibrationStandardUnits
CalibrationStandardLotCode
CalibrationStandardPurchaseDate
CalibrationStandardExpirationDate
CalibrationStandardTypeCV
PK Term
Definition
ODMCore_Methods
PK MethodID
MethodType
MethodCode
MethodName
MethodDescription
OrganizationID
MethodLink
Calibrations
PK CalibrationID
FK4 DeploymentMeasuredVariableID
FK3 FieldActivityID
FK2 CalibrationMethodID
CalibrationCheckValue
FK1 CalibrationStandardID
CalibrationNotes
CalibrationFileName
Equipment Information
Equipment
PK EquipmentID
EquipmentSerialNumber
EquipmentInventoryNumber
FK3 EquipmentType
FK1 EquipmentOwnerID
FK2 EquipmentModelID
FK4 EquipmentVendorID
EquipmentPurchaseDate
EquipmentPurchaseOrderNumber
EquipmentPhotoFileName
EquipmentNotes
EquipmentModels
PK ModelID
FK1 ModelManufacturerID
ModelPartNumber
ModelName
ModelDescription
ModelNotes
ModelSpecificationsFile
ModelLink
EquipmentTypeCV
PK Term
Definition
Variable Information
SensorOutputVariables
PK SensorOutputVariableID
FK1 ModelID
SensorVariableCode
SensorVariableName
SensorMethodDescription
SensorResolution
SensorAccuracy
SensorRawOutputUnitsID
Activity Information
SiteVisits
PK SiteVisitID
SiteVisitBeginDateTime
SiteVisitEndDateTime
SiteVisitUTCOffset
SiteVisitEnvironmentalObservations
SiteVisitNotes
FieldActivities
PK FieldActivityID
FK1 SiteVisitID
FK2 FieldActivityType
FieldActivityBeginDateTime
FieldActivityEndDateTime
FieldActivityUTCOffset
FieldActivityDescription
FieldActivityNotes
SiteVisitPhotos
PK PhotoID
FK1 SiteVisitID
PhotoFilePath
PhotoDescription
ActivityTypeCV
PK Term
Definition
DeploymentMeasuredVariables
PK DeploymentMeasuredVariableID
FK1 DeploymentID
FK2 SensorOutputVariableID
MeasuredVariableEquationUsed
MeasuredVariableValueMultiplier
MeasuredVariableValueOffset
MeasuredVariableRecordedStatistic
MeasuredVariableNotes
Site Information
ODMCore_Sites
PK SiteID
SiteCode
SiteName
Latitude
Longitude
LatLongDatum
Elevation
VerticalDatum
Comments
FK1 SiteType
SitePhotoFileName
FK2 SiteGroupID
SiteVisitPeople
FK2 SiteVisitID
FK1 PersonID
DeploymentTypeCV
PK Term
Definition
Deployments
PK DeploymentID
FK4 EquipmentID
ParentEquipmentID
FK1 SiteID
FK2 FieldActivityID
DeploymentBeginDate
DeploymentEndDate
DeploymentUTCOffset
DeploymentCurrentlyDeployed
FK3 DeploymentType
DeployentDescription
DeploymentHorizontalOffsetDescription
DeploymentHorizontalOffsetValue
DeploymentHorizontalOffsetUnitsID
DeploymentVerticalOffsetDescription
DeploymentVerticalOffsetValue
DeploymentVerticalOffsetUnitsID
DeploymentNotes
DeploymentSchematic
DataSeriesSensorDeployments
FK2 DataSeriesID
FK1 DeploymentMeasuredVariableID
ODMCore_People
PK PersonID
PersonFirstName
PersonLastName
PersonOrganization
PersonAddress
PersonPhone
PersonEmail
SiteGroups
PK SiteGroupID
SiteGroupName
SiteGroupDescription
SiteTypeCV
PK Term
Definition
DataloggerFileColumns
PK ColumnID
FK1 DataloggerFileID
FK2 DeploymentMeasuredVariableID
ColumnLabel
ColumnDescription
ColumnUnitsID
ColumnNotes
ODMCore_Observations
PK ObservationID
FeatureID
VariableID
MethodID
OrganizationID
QualityControlLevelID
ObservationType
BeginDateTime
BeginDateTimeUTCOffset
EndDateTime
EndDateTimeUTCOffset
ResultDateTime
ResultDateTimeUTCOffset
ValidDateTime
ValidDateTimeUTCOffset
Status
SampledMedium
DataloggerProgramInformation
PK ProgramID
FK1 ProgrammedByID
ProgramDescription
ProgramFileName
DataloggerFiles
PK DataloggerFileID
FK1 ProgramID
DataloggerFileName
DataloggerFileDescription
Datalogger Program Information

# Equipment Management

- http://data.iutahepscor.org/gamutmanagement/
- Menus
- Vocabularies: add here or in forms
- Workflow:
  - Add Equipment (if needed)
  - Add Sites (for Logan and Provo)
  - Add Deployments (for one site)
    - Add Site Visit → Add Deployments, repeat
    - Add Deployment Variables (on Sites page)
  - Add other Field Activities (Stage Reading, Sample Collection, Retrieval- to “un-deploy”)
    - Add Site Visit → Add Field Activity

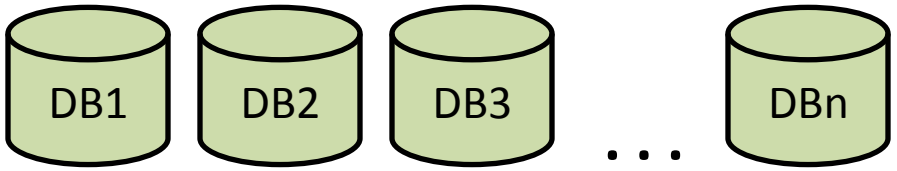
# Observations Data Model

- http://his.cuahsi.org/odmdatabases.html
- ODM is a standard structure for observations data storage and retrieval in a relational database allowing for sharing and publication
- (Should) provide sufficient metadata for observations data to be unambiguously interpreted and used
- Use SQLServer to store data and structured query language for querying (let's do a few queries)!
- Tools have been developed to use with data stored in ODM (Streaming Data Loader, ODM Tools, etc.)

# Server-Based RDBMS

- Database Server


  - Databases

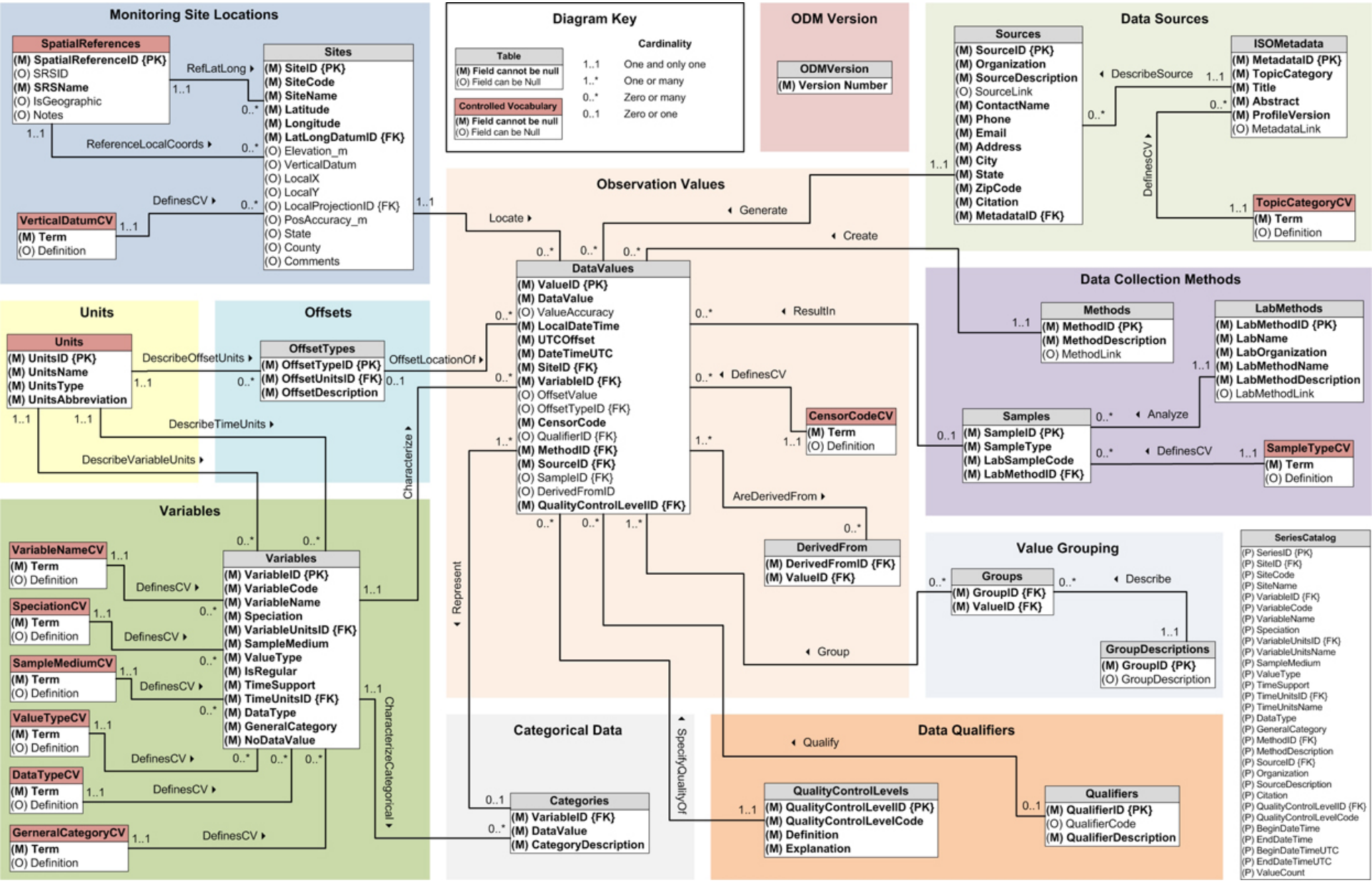


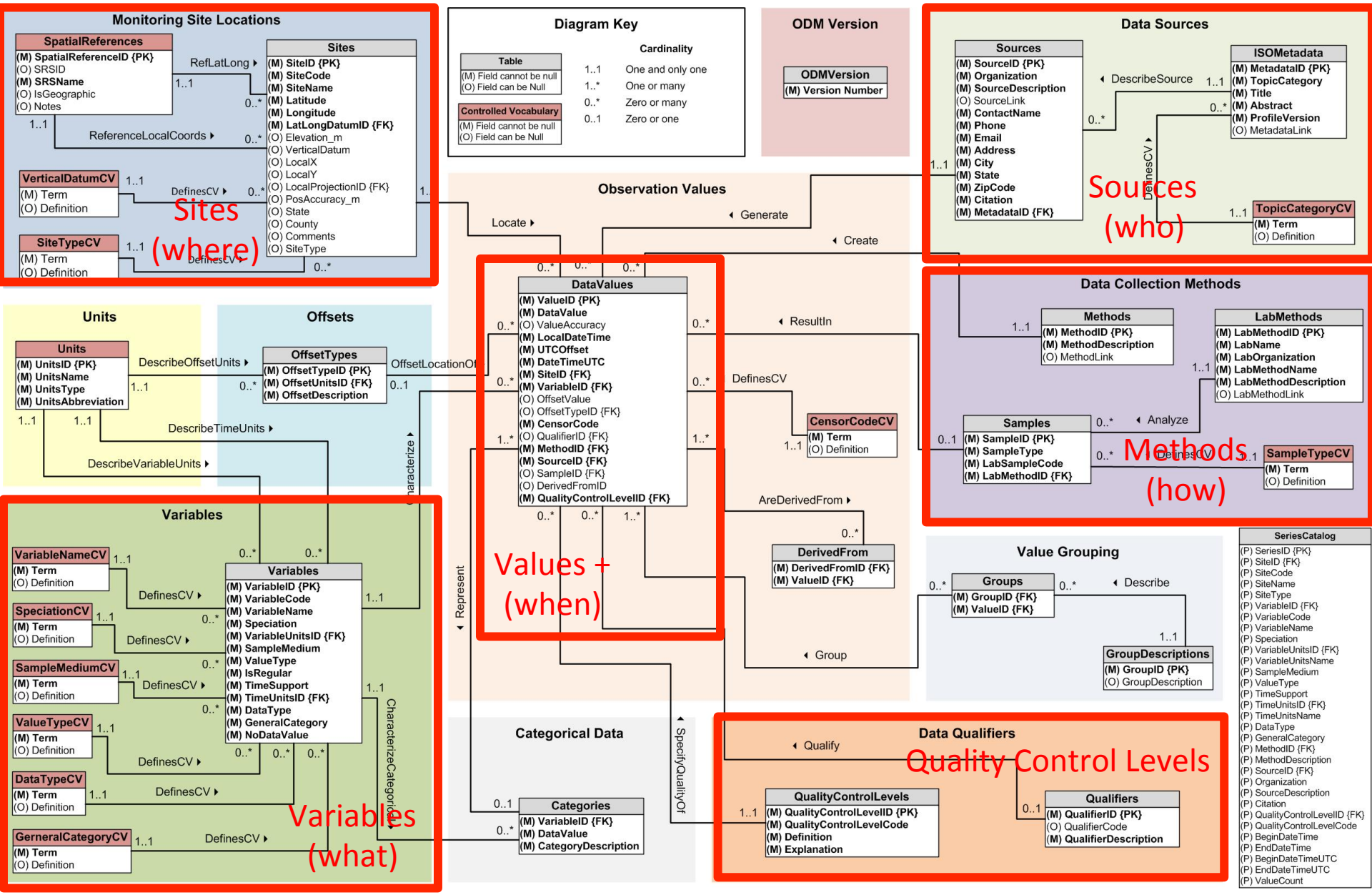

    - Tables

| Site | Variable | Date | Value |
|---|---|---|---|
| 1 | Temperature | 8/2/2007 14:00 | 12.4 |
| 1 | Temperature | 8/2/2007 14:30 | 12.7 |
| 1 | Temperature | 8/2/2007 15:00 | 13.1 |

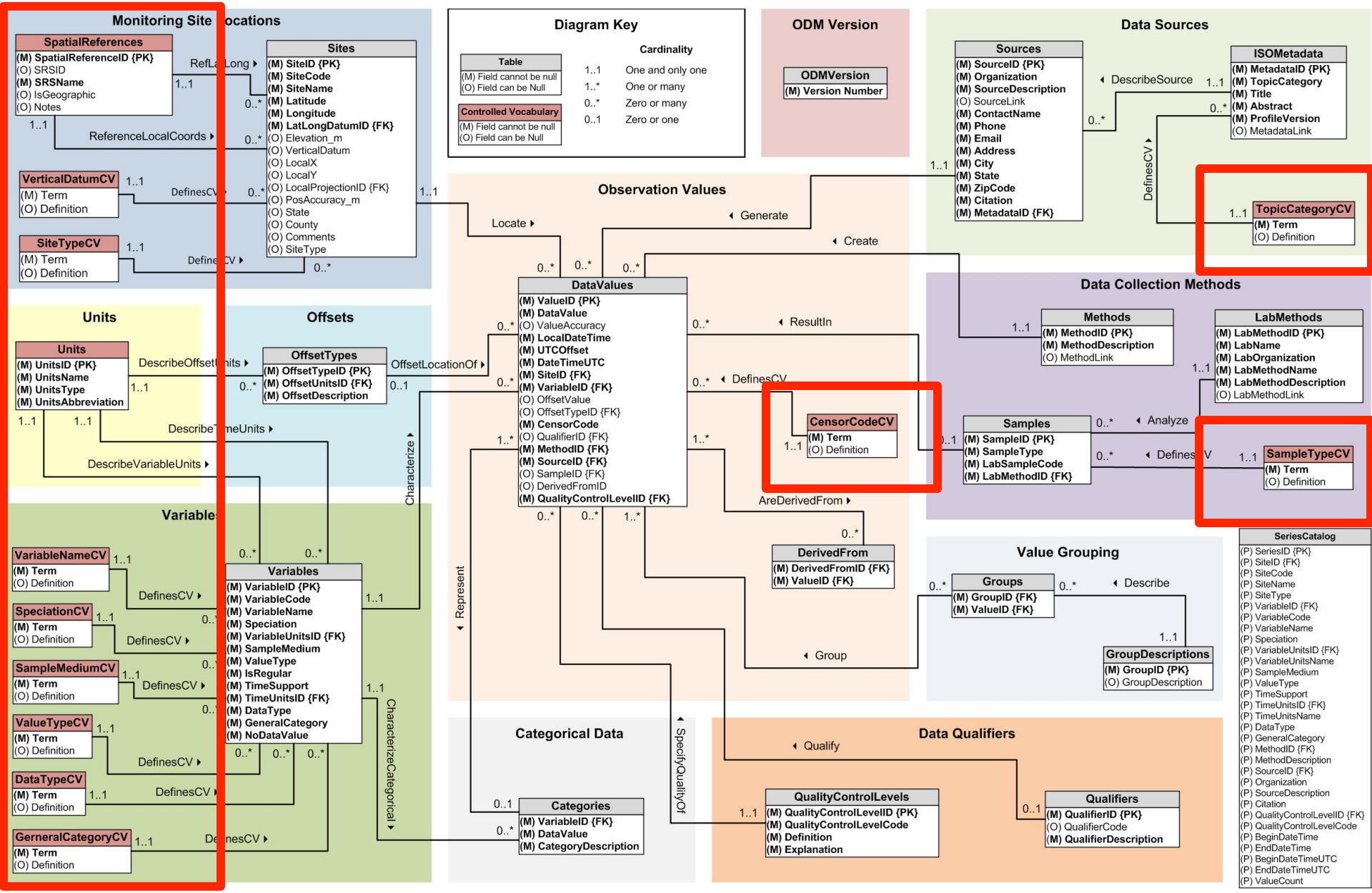
      - Records

| | | | |
|---|---|---|---|
| 1 | Temperature | 8/2/2007 14:00 | 12.4 |


Monitoring Site Locations
SpatialReferences
(M) SpatialReferenceID {PK}
(O) SRSID
(M) SRSName
(O) IsGeographic
(O) Notes
RefLatLong
ReferenceLocalCoords
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
DefinesCV
VerticalDatumCV
(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
Analyze
Samples
(M) SampleID {PK}
(M) SampleType
(M) LabSampleCode
(M) LabMethodID {FK}
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
GenreralCategoryCV
(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) 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


Monitoring Site Locations
SpatialReferences
(M) SpatialReferenceID {PK}
(O) SRSID
(M) SRSName
(O) IsGeographic
(O) Notes
RefLatLong
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
Sites (where)
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
Sources (who)
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}
Values + (when)
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
Analyze
Samples
(M) SampleID {PK}
(M) SampleType
(M) LabSampleCode
(M) LabMethodID {FK}
DefinesCV
SampleTypeCV
(M) Term
(O) Definition
Methods (how)
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
CharacterizeCategories
Variables (what)
Categorical Data
Categories
(M) VariableID {FK}
(M) DataValue
(M) CategoryDescription
SpecifyQualityOf
Value Grouping
Groups
(M) GroupID {FK}
(M) ValueID {FK}
Describe
GroupDescriptions
(M) GroupID {PK}
(O) GroupDescription
Data Qualifiers
Qualify
QualityControlLevels
(M) QualityControlLevelID {PK}
(M) QualityControlLevelCode
(M) Definition
(M) Explanation
Qualifiers
(M) QualifierID {PK}
(O) QualifierCode
(M) QualifierDescription
Quality Control Levels
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

# Controlled Vocabularies



# Quick Summary: Formulating a SQL Statement

| | |
|---|---|
| 1. Identify the field(s) containing the source data | **SELECT** Field_1, Field_2, Field_n<br>**SELECT *** returns all fields<br>-Can also **INSERT**, **UPDATE**, **DELETE** |
| 2. Identify the table(s) where the fields are located | **FROM** Table_1 |
| 3. Specify criteria to narrow the results | **WHERE** Field_1 = SomeCriteria **AND/OR** Field_2 = SomeCriteria<br>-Can also use **>**, **<**, **<=**, **>=**, **<>**, **LIKE**, **IN/NOT IN**, **BETWEEN** |
| 4. Determine the order to present records in the results | **ORDER BY** Field_1 **ASC** |
| 5. OTHER FUNCTIONALITY:<br>a. Can aggregate results.<br><br>b. Can select from more than one table. | **MIN**, **MAX**, **SUM**, **AVG**, **COUNT**<br>**GROUB BY** groups records into sets for aggregation.<br><br>Use **JOIN** |

# Query Examples

```
SELECT * FROM Sites
WHERE Longitude < -111.8

SELECT * FROM Sites WHERE SiteID = 1 AND SiteID = 2
SELECT * FROM Sites WHERE SiteID = 1 OR SiteID = 2

SELECT AVG(DataValue) FROM DataValues
    WHERE SiteID = 1 AND VariableID = 1 And DataValue <> -9999

SELECT SiteID, AVG(DataValue) FROM DataValues
    WHERE VariableID = 66 AND QualityControlLevelID = 0
    GROUP BY SiteID

SELECT * FROM DataValues
    WHERE (SiteID = 3 OR SiteID = 10) AND VariableID = 59
    AND LocalDateTime >= '9/20/2013' AND LocalDateTime < '9/25/2013'
    ORDER BY SiteID, LocalDateTime ASC

SELECT COUNT(*) FROM DataValues
WHERE SiteID = 1 AND VariableID = 9 AND DataValue <> -9999

SELECT MAX(DataValue) AS MaxTemp, MIN(DataValue) AS MinTemp FROM DataValues
WHERE SiteID = 3 AND VariableID = 57 AND DataValue <> -9999

SELECT MAX(LocalDateTime) AS LastDateTime FROM DataValues
WHERE SiteID = 3 AND VariableID = 57
```

# Data Access

- First iteration of web access:

http://data.iutahepscor.org/odmmap/

http://data.iutahepscor.org/odmtsa/

- ODM Tools: http://his.cuahsi.org/odmtools.html
  - Downloadable program for data visualization and editing
  - Working on new version…
- Can make direct connection to database via R or Matlab and run queries in preferred environment
- Thoughts about how often to check data, etc?
- Ideas for automated QC checks?