Catalina - Jemez CZO Sample and Data Management

Matej Durcik, Jon Chorover, Tom Meixner, Jennifer McIntosh, Craig Rasmussen







Data Management Infrastructure



Publishing Data and Metadata Tools



Data Storage Standards

- Database
 - Time series data (ODM1, MySQL)
 - Water chemistry data (ODM1, MySQL)
 - Geochemistry data (CZchemDB)
- Spatial Datasets
 - GIS standards for data and metadata
 - Vector data (ESRI shapefile, KLM, GeoJson)
 - Raster data (GeoTIFF, IMG, ASCII...)
 - Geo-database (Oracle)
 - Hydrography (ArcHydro, NHD)
- Files
 - All raw data are archived in original formats in multiple locations (server, external drives)
 - Time series and chemistry data are archived in CZO
 Display File v1 and MS Excel files





THE UNIVERSITY

OF ARIZONA .

Time Series and Water Chemistry Database



CUAHSI ODM: http://his.cuahsi.org/odmdatabases.html







Collected Datasets

- Climate
 - 5 met station (10 and 30 min)
 - 3 flux tower (30 min)
 - 13 rain gages (30 min)
- Soil data
 - 12 pedons in the Jemez and 4 in Catalinas (15 min)
 - VWC and temperature at 3 depth
 - Water potential at 2 depth
 - Redox potential at 4 depth
 - CO_2 and O_2 at 4 depth
 - Soil water samplers at 3 depth
- Soil data in the Catalinas
 - 19 pits VWC at various 2-4 depths (10 mins)
- Ground water level
 - 15 piezometers (30 min)
 - 4 in the Jemez (30 min)
- Solid samples
 - Soil and rocks
 - Drilling cores









Collected Datasets Cont.

- Sap Flows
 - 17 trees in Catalinas (30 min)
 - 16 trees in Jemez (discount.) (30 min)
- Stream flow
 - 10 flumes (15 min)
- Water chemistry
 - ~ 500 samples/year
 - 64 variables
- LiDAR datasets
 - 3 airborne (~900 GB)
 - 1 ground (~100 GB)
- GeoSpatial Data
 - Locations and boundaries
 - DEMs and derive datasets (TWI, Slope, Aspect, ...)
- Phenocam photos and data
 - 7 CAMERAS (1 hr; ~100GB/yr)
 - Snow depth
 - Greenness
- Model Datasets
 - EEMT









Data Publishing and Sharing

- CZO Database
 - More than 59 million values
 - 13.7 million values collected in 2014
- Published
 - 41 datasets available on the CZO website
 - Time series data (28 datasets)
 - Water Chemistry data (4 datasets)
 - Soil Geochemistry (1 dataset)
 - Geo-spatial and LiDAR (8 datasets)
 - CZO Display File v1 and standard geo-spatial formats
 - Updated 1-4 times per year, based on availability
 - 3 LiDAR datasets in the OpenTopography
 - More than 17 million values in the CZO Central
 - 2 data services (czo_ariz and czo_catalina)
 - CUAHSI HydroClient (<u>http://data.cuahsi.org/</u>)
- Data usage
 - From 6/1/2014 to 6/1/2015
 - 766 unique users
 - 6876 data files downloaded from published datasets





OF ARIZONA .

Data Publishing and Sharing Cont.

- Cross-CZO Sharing
 - Scientist to Scientist
 - Exchanging ideas and data
 - Published data on the website and CZO Central
- Cross-CZO Publications
 - 17 from 71 research papers (acknowledged NSF CZO projects)
 - Used datasets in these studies
 - Lidar
 - Water chemistry
 - Flux tower
 - Met data
 - Snow depth
 - Soil geochemistry
 - Soil moisture
 - Stream flow







Next Steps

- Publish new datasets
 - Geochemistry
 - Mt. Bigelow Flux tower and ZOB datasets
 - Pheno cam photos and data
- Transition from
 - ODM1 to ODM2
 - Data Display File v1 to v2







Questions

How can CZ DATA people help with depositing data and getting a DOI in a stable public repository to publish in open access journal? [CZ Data is not/cannot be a stable public repository because it does not have permanent funding (according to Emilio)] If CZ DATA does not have capacity to help with this - what are possible alternatives to minimize redundancy?

How can CZ DATA people help to register samples to get unique International Geo Sample Number (IGSN)? Is it possible to register measuring location not only samples?

When will the final version of YODA templates and tools will be released?

What is a time frame for transition from ODM1 to ODM2?

How often are the Data Visualization Tools updated? Could you add watershed boundaries associated with regions? Only data for the Catalinas (czo_catalina) are visualised. Could you add the webservice for the Jemez (czo_ariz) as well?

Are you planning web-seminar or training sessions about new data format, ODM2 and tools?

Can you show us how to use the CZOdata tool for solid phase data (i.e., soil chemistry/mineralogy results)? CZchemDB was replaced a new system of solid phase data incorporation, but it is not clear whether that is ready for use at this time. We would like to discuss the template in detail in the Webex as this is a pressing need for our CZO.







Questions Cont.

How easy is ODM2 to use for solid samples such as soil or rock samples?

Could you link the data tools (visualization and search portal) to the CZO data webpages?

More generally, is there a listing or site with templates currently available for CZO data ingestion?

For entering new publication in the CZO website, could you add option datasets with list of all published datasets? This will be very helpful and more practical for all CZOs to link publications with data rather than it is done now.





