Data Management Plan for the CZO- Nov. 2013

*1. Expected Data***:** Three broad data types will be collected at the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO) during the project:

- Observation data (air temperature, humidity, precipitation measurement, wind speed and direction, soil moisture content, stream flow rate, turbidity, soil temperature, soil matric potential, dielectric, and sap flow), are sent in weekly batches from sensors through data loggers, wireless hubs, and microwave transmitters to the Penn State University (PSU) campus network. Data are processed by investigators and the SSHCZO Cyberspecialist to derive CO_2 flux, eddy covariance, and hydropedologic properties.
- Physical samples (precipitation, ground water, stream water, soil, vegetation, and gamma ray counts) are collected in situ by investigators. Soil geochemistry, aqueous chemistry, stable isotope hydrology, and vegetative chemistry datasets are derived by investigators from these samples.
- Model data are produced from numerical simulations using datasets above and others as needed (e.g., LIDAR elevation data). These data are made available at the modeler's discretion.

2. Data Format: Observation data are stored in a Microsoft SQL server database. Physical sample data are recorded in Excel spreadsheet files or ASCII files, and are provided to SSHCZO data managers. Geochemical data (CZChemDB) are in Microsoft Access format. Model outputs are stored in binary formats, accompanied by instructions and/or code required. Metadata are made available in Microsoft Word and ASCII files.

3. Access to Data and Data Sharing Practices and Policies

Data Access: Data are available on request under the sharing practice defined below. A web interface for the SQL Server database is under construction, available January 2014, and will allow direct access to observation data, within embargo limitations. Sample data are available on the SSHCZO data web site (http://criticalzone.org/shale-hills/data/), and will be stored in the SQL Server database. Model outputs are available at the modeler's discretion. Datasets under embargo require a password available from the SSHCZO data managers. Once per quarter, new datasets are converted to ASCII files and are harvested by the San Diego Supercomputing Center Geoportal (http://spatial.sdsc.edu/lab/). Work is in progress for making CZChemDB available from EarthChem (http://www.earthchem.org/portal).

Definitions related to Data Sharing Practices and Policies (below):

CZO Data Products: Data collected with monetary or logistical support from a CZO. This includes use of any CZO sensors, sampling equipment, vehicles, or labor from a supported investigator, student or staff.

Private CZO Repository: Password-protected directory on the CZO data server. Files will be accessible by all investigators and collaborators in the CZO and logins will be maintained by

CZO data managers. Although data values will not be accessible by the public or ingested into any central data system (i.e. CUAHSI HIS), metadata will be fully discoverable by the public. This provides the dual benefit of giving attribution and credit to dataset creators and the CZO in general, while maintaining protection of intellectual property while publications are pending.

Dataset Creators: People responsible for dataset design, collection, analysis and quality assurance. Dataset Creators are analogous to authors of a publication, and datasets should be cited in an analogous manner following emerging guidelines described at http://www.datacite.org/whycitedata.

Data Sharing: All SSHCZO investigators and collaborators who receive CZO support agree to:

1. Share data privately within 1 year. CZO investigators and collaborators agree to provide CZO Data Products, including data files and metadata for raw, quality controlled and/or derived data, to CZO data managers within one year of collection. By default, data values will be held in a Private CZO Repository (password-protected directory/files), but metadata will be made public and will provide full attribution to Dataset Creators.

2. Release data to public within 2 years. CZO Dataset Creators will be encouraged after one year to release data for public access. Dataset Creators may choose to publish or release data sooner.

3. Request, in writing, data privacy up to 4 years. CZO PIs will review written applications to extend data privacy beyond 2 years and up to 4 years from collection. Extensions beyond 3 years should not be the norm, and will be granted only for compelling cases.

4. Consult with creators of private CZO datasets prior to use. In order to enable the collaborative vision of the CZO program, data in Private CZO Repositories will be available to other investigators and collaborators within that CZO. Releasing or publishing any derivative of such private data without explicit consent from the Dataset Creators will be considered a serious scientific ethics violation.

4. Policies for Re-Use, Re-Distribution: The following policy is also available on the SSHCZO data web site:

1. Use our data freely. All CZO Data Products except those labelled Private are released to the public and may be freely copied, distributed, edited, remixed, and built upon under the condition acknowledgement is given as described below. Most private data will be released to the public within 1-2 years, with some exceptionally challenging datasets remaining private up to 4 years. To inquire about potential earlier use, please contact the SSHCZO data management team and/or Dataset Creators. Non-CZO data products, like those produced by USGS or NOAA, have their own use policies, which should be followed.

2. Give proper acknowledgement. Publications, models, and data products that make use of these datasets must include proper acknowledgement, including citing datasets in a similar way to

citing a journal article (i.e., author, title, year of publication, name of CZO publisher, edition/version, and URL/DOI information. See http://www.datacite.org/whycitedata).

3. Let us know how you will use the data. Dataset Creators appreciate hearing of any plans to use the dataset. Consider consultation or collaboration with Dataset Creators.

*5. Archiving of Data***:** All datasets are archived electronically with assistance of the PSU College of Earth and Mineral Sciences (EMS) Information Technology Services. The SQL Server database and data files available on the web are archived according to the EMS web backup scheme (daily incremental, weekly full backups). Archive media are not currently stored off-site, but this is planned for the near future. Archive media are retained indefinitely. Hard copy notebooks are kept at discretion of each investigator, as all data obtained from samples and experiments are provided to SSHCZO data managers in electronic form. Physical samples used in collecting these data and performing experiments are currently archived but Penn State cannot guarantee archival more than five years.