Research Data Policy

Version 1.3

June 11, 2014

Edited by:

Jeffery S. Horsburgh and Amber S. Jones

i. Preface

iUTAH (innovative Urban Transitions and Aridregion Hydrosustainability) is committed to an open data policy that will maximize the impact and broad use of datasets collected within iUTAH facilities and by iUTAH research teams. This policy document focuses on assisting iUTAH investigators in creating and sharing high-quality data. The iUTAH Modeling & Data Federation (MDF) is developing tools that assist iUTAH participants in discovering, visualizing, and accessing data in which they are interested.

ii. Document Contributors and Contacts

Name	Organization	Email
Jeffery S. Horsburgh	Utah State University	jeff.horsburgh@usu.edu
Amber S. Jones	Utah State University	amber.jones@usu.edu
Todd Crowl	Utah State University	facrowl@gmail.com
Courtney Flint	Utah State University	courtney.flint@usu.edu
Jim Ehleringer	University of Utah	jim.ehleringer@utah.edu
Dave Eiriksson	University of Utah	dave.eiriksson@utah.edu
Zach Aanderud	Brigham Young University	zachary_aanderud@byu.edu
Erin Jones	Brigham Young University	erinfjones3@gmail.com
Doug Jackson-Smith	Utah State University	doug.jackson-smith@usu.edu

iii. Revision History

Date	Release Version	Editors	Sections Modified	Description
12/19/2013	1.0	Jeffery Horsburgh Amber S. Jones	Initial release	Initial release
1/31/2014	1.1	Jim Ehleringer, Courtney Flint, Jeff Horsburgh and others	Many modifications made throughout after first round of reviews by data policy committee.	New version resulting from review by iUTAH data policy committee.
4/3/2014	1.2	Doug Jackson-Smith, Jeffery Horsburgh, Courtney Flint, Amber Jones	Integrating comments – many modifications throughout	New version resulting from review by iUTAH Leadership Team.
6/11/2014	1.3	Amber Jones, Jeffery Horsburgh	Integrating comments from the Management team - modifications throughout	New version resulting from review by iUTAH Management Team.

Table of Contents

1. Overview

The iUTAH Modeling and Data Federation (MDF) is a suite of hardware, software, and computer systems that provide cyberinfrastructure (CI) to support the research efforts of the iUTAH project. Tools developed by the iUTAH CI Team will support the full data lifecycle for research data collected as part of the iUTAH project. Research data are defined as "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings" by the U.S. Office of Management and Budget and includes the data as well as the metadata that define the content and context of the data (US OMB, 1999). The National Academy of Sciences, National Academy of Engineering, and Institute of Medicine (2009) describe research data:

> *It includes textual information, numeric information, instrumental readouts, equations, statistics, images (whether fixed or moving), diagrams, and audio recordings. It includes raw data, processed data, published data, and archived data. It includes the data generated by experiments, by models and simulations, and by observations of natural and social phenomena at specific times and locations. It includes data gathered specifically for research as well as information gathered for other purposes that is then used in research. It includes data stored on a wide variety of media including magnetic and optical media.*

The National Science Foundation (NSF) requires establishment of policy governing the data collected as part of NSF-funded research efforts. The guidelines in this document are not intended to replace legal and institutional requirements regarding data rights, privacy, and sharing, but are proposed and implemented to meet NSF requirements and should supplement existing institutional data stipulations.

In general, and as specified in the original Data Management Plan that accompanied the iUTAH proposal (Appendix A), all of the primary datasets generated as part of iUTAH will be made freely and publicly available using the methods for data loading and distribution developed by the iUTAH CI Team. The purpose of this document is to clarify the types of data that will be supported and provide more specific guidance and policy related to collection and publication of iUTAH datasets. The goal of this policy is to ensure that iUTAH data are available to and useful for as wide an audience as possible in a timely manner, as detailed below.

2. Roles and Responsibilities

2.1. iUTAH Data Policy Committee

A Data Policy Committee was formed to oversee the development of this policy document, to advise the CI Team on prioritizing data products to include in the iUTAH MDF, to advocate for the implementation of iUTAH data policies and practices among iUTAH participants, and to assist in resolving any conflicts that occur. The Data Policy Committee will also review, provide feedback on, and approve Data Collection Plans submitted by iUTAH researchers.

The eight members of the Data Policy Committee initially include:

- iUTAH Project Director, ex officio
- A representative from the CI Team
- A representative from each of the two iUTAH research facilities
- A representative from each the three Research Focus Areas
- An iUTAH graduate student representative.

Members serve a 1-year renewable term and may rotate on a yearly basis at the direction of the iUTAH Leadership Team. The requirements articulated in this document will be reviewed annually by the iUTAH Data Policy Committee in collaboration with the CI Team. Any changes must be approved by the iUTAH Leadership Team. Additional members of the Data Policy Committee may be added at the direction of the iUTAH Leadership Team.

2.2. iUTAH Cyberinfrastructure Team

The management and curation of multi-investigator data has both service and research components. While it might be initially perceived that the role of the CI Team is solely to structure user-provided data for external display, this is not the case if iUTAH is to have user-friendly, integrated, and coupled data for applications across a wide range of disciplines. Development of the required repositories, protocols, and methods for enabling shared access to the full range of data created by iUTAH is part of the intellectual contribution of the iUTAH project. The role of the iUTAH CI Team includes both conducting cutting-edge, CI-related research but also assisting iUTAH investigators in structuring, organizing, and sharing high quality data. The CI Team is responsible for developing tools to assist researchers and other data consumers in discovering, visualizing, and accessing data of interest.

The CI Team will be responsive to iUTAH participants regarding the loading and dissemination of iUTAH datasets. The CI Team will ensure that iUTAH data will be centrally stored, loaded, and archived to facilitate collaboration. The CI Team will work with researchers to provide guidance for the protocol to follow to load data and metadata to the iUTAH MDF including specific workflows of particular types of data. The CI Team will also provide training and/or resources on the relevant CI tools to iUTAH participants. This may be done in conjunction with the iUTAH Education, Outreach, and Development (EOD) team. The CI Team will also establish web services for iUTAH datasets to make them accessible as part of various large-scale data repositories. Finally, although the CI Team's primary responsibility is supporting the data creation and modeling efforts of the iUTAH project, the CI Team will work to facilitate access to data external to iUTAH (e.g., data in national data repositories and those from local, state and federal agencies) under prioritization from the iUTAH Leadership Team.

2.3. iUTAH Researchers

Researchers participating in and receiving funding from iUTAH are expected to adhere to the policies in this document. The iUTAH Data Publisher's Agreement (Appendix B) specifies the expectations for researchers participating in iUTAH and submitting data to the iUTAH MDF. In general, participants should provide high quality datasets with sufficient metadata for unambiguous use and interpretation. Requirements for data and metadata may be specific to the associated discipline and the data type, but a standard set of core metadata is required for all datasets to facilitate archival, storage, and cataloging for discovery and retrieval.

Researchers must consider the submittal of datasets to the iUTAH MDF Data Publication System similar to the submittal of manuscripts to research journals. Data collection, quality assurance, quality control (QA/QC), and annotation with metadata are all the responsibility of iUTAH researchers and should be conducted in accordance with protocols developed in collaboration with the CI Team. It is acknowledged that QA/QC procedures may be different across data types within iUTAH. We expect that the QA/QC procedures employed will be consistent with common practices employed within the respective discipline. Researchers must be responsive to requests by the CI team for descriptive metadata.

Researchers have a right to expect that the iUTAH CI Team will provide clear procedures for creating metadata for and uploading their data to the iUTAH MDF and that CI Team staff will assist with data curation. Researchers will have the capability to curate their data and metadata, tools for uploading and describing their data, a secure and well-maintained repository for their data, and mechanisms for data discovery and access by others. Finally, for some data types (e.g., Type B data described in Section 5 below), researchers should have the expectation of first rights to analyze and publish those data and to be credited for their data contributions.

2.4. iUTAH Leadership Team

The iUTAH Leadership Team is primarily responsible for:

- Reviewing and approving the data management policy document
- Reviewing and approving subsequent changes to the data management policy document
- Ensuring that the data management policy document is consistent with policies at the NSF
- Enforcement of the data management policy, working with iUTAH investigators who will be submitting data to ensure that the conditions of this policy are met
- Resolving conflicts that arise in the specification of Data Collection Plans or among data collection teams (e.g., in cases where there are disputes about who should be included in a data collection team or who should have access to iUTAH datasets, etc.)

3. Data Collection Plans

3.1. Creation of Data Collection Plans

For ***all*** data creation efforts executed with ***any*** funding from iUTAH (salary, travel, sampling, equipment purchase, etc.), a brief plan for data collection should be submitted to the Data Policy Committee prior to the funding of data collection/creation. Investigators or data collection teams who are uncertain about whether a data collection plan is required should consult with the iUTAH Data Policy Committee. Data collection efforts that collect multiple subtypes of data need only submit a single plan, but each subtype of data should be described.

The Data Collection Plan should include the following information:

a. Identification of the types of data to be created (using the data typology below).
b. A brief description of the methods that will be used to create the data.
c. Identification of the data formats that will be used to store the data.
d. A brief description of the data to be created (e.g., what are the final products that will be uploaded to the iUTAH MDF for publication?). This may also include a listing of anticipated journal papers or other publications.
e. A description of who will have access to the data during data collection and how the data will be made broadly available after completion and publication.
f. Information on potential collaborators/co-authors for each data product and anticipated publications.

The investigator or data collection team must define the scope of their data collection efforts in the data collection plan. It is acknowledged that, in practice, the boundaries of when data collection begins and ends may be ambiguous and that the scope of data collection efforts may change over time as data are collected. Where there are substantive changes to the types of data, methods, data formats, products, or availability of data described in a submitted data collection plan, the investigator or data collection team should submit an updated data collection plan to the Data Policy Committee.

3.2. Cases in which IRB Approval is Required Prior to Data Collection

Some data related to "human subjects" are sensitive in nature (i.e., those datasets identified as Type C below). Their collection and management will require approval from an Institutional Review Board (IRB). Efforts should be made to formulate the IRB application so that resulting data are available to the largest number of iUTAH researchers and partners as possible. All IRB documentation must be approved and in place before data collections can begin. Initial IRB applications must be submitted to the iUTAH Leadership Team and Data Policy Committee at the time of initial submission for review to alert these teams that the IRB application has been submitted. The subsequently approved applications should also be submitted to the Data Policy Committee and iUTAH Leadership Team as soon as possible, and the approved copy will replace any previously submitted

versions. Approved IRB applications must be submitted by the iUTAH Leadership Team to the National Science Foundation as part of project reporting requirements.

It is the responsibility of each researcher or data collection team to acquire IRB approval at their respective institution prior to data collection. Collaborative teams must ensure that all IRB requirements across institutions are met. Data involving human subjects are subject to data sharing obligations, but may require redaction or de-identification by removing identifying information (e.g., names, addresses, etc.) as well as removing indirect identifiers of research participants. When release of the dataset would result in deductive disclosure of participants, some portion of the dataset may be withheld from release and/or aggregations or statistical summaries of the data should be developed and released. Details of plans for data availability and anonymization and/or aggregation should be included in the Data Collection Plan.

In cases where there are questions or disputes about data access during or after data collection or the scope of the plan for broad distribution of the data (item e above), the Data Policy Committee may elevate those issues to the iUTAH Leadership Team for guidance and resolution.

4. Data and Metadata Standards

All data, both raw and derived, regardless of Type (as defined in Section 5 below) must be documented with complete metadata as defined below. Where possible and appropriate, iUTAH will make full use of existing and emerging standards for sharing environmental data. Datasets submitted to the iUTAH MDF will include, at a minimum, the standard metadata elements of the Dublin Core metadata standard (http://dublincore.org/documents/dces/). The data and metadata entry and upload functionality provided by the iUTAH MDF will facilitate entry of required and optional metadata elements. For static quantitative datasets to be published in the iUTAH MDF (such as experimental data, survey data, organizational data, etc.), submitted metadata should include clear descriptions of variable names, attributes, and value definitions. Submitted metadata should be consistent with the submitted data – i.e., if the submitted data is raw data, variable names, attributes, and value definitions should be for the raw data. Metadata describing finalized or derived data products that are different than the raw data should describe the finalized or derived data.

Time series data collected by iUTAH, such as the streaming sensor data from GAMUT sites, will be stored using the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS) Observations Data Model (ODM) (Horsburgh et al., 2008) and published in the Water Markup Language (WaterML) format using CUAHSI HIS WaterOneFlow web services (Zaslavsky et al., 2007).

Geospatial datasets published as part of iUTAH will be made available using common geospatial data formats. In some cases, web map services will be created from geospatial datasets using existing Open Geospatial Consortium interfaces such as Web Map Services,

Web Feature Services, and Web Coverage Services for easy use within Geographic Information Systems (GIS) software.

For data accessed via HTTP for download, appropriate metadata descriptions will accompany the data download. For all datasets, regardless of restrictions on access, metadata will be publicly available via the iUTAH MDF for data discovery purposes.

5. Data Typology

This policy defines a data typology to categorize datasets and other research products (e.g., including model results) created by iUTAH researchers so that specific policies related to sharing, access, and timeframes can be appropriately applied. Data created or collected under funding from iUTAH should be defined as either Type A or Type B. In addition, the dataset may require the additional stipulations of Type C. Finally, data acquired from external sources, but paid for by iUTAH may be considered as Type D data. For iUTAH data collection efforts, the data type(s) to be created should be defined in the Data Collection Plan submitted to the Data Policy Committee. The data types are defined as:

Type A - Primary iUTAH datasets or research products. These include raw and QA/QC calibrated sensor data from iUTAH facilities, baseline sampling datasets across iUTAH facilities and sites, and general datasets collected by iUTAH for the community of iUTAH participants.

Type B - Datasets or other research products for which monetary or material support was provided by iUTAH, but that are created by a specific investigator, student, or coordinated iUTAH research group to support a particular research question or goal.

Type C - Type A and Type B datasets or products that include personally identifiable information or information about human subjects/participants and are subject to IRB restrictions.

Type D – Datasets or other research products procured by iUTAH or iUTAH participants supported by iUTAH that are subject to licensing, copyright, or use restrictions/agreements from the data source that may prohibit general distribution of the data.

It is anticipated that groups of iUTAH researchers and/or partners will wish to create derived data or other research products based on datasets that have been published within the iUTAH MDF. Where the creation of these derived data products is funded by iUTAH, derived data products will fall under Type B – C and will be subject to the requirements of those data types. Where funding for the creation of derived data products is not provided by iUTAH, sharing of the data products will be at the discretion of the creators.

For existing datasets from external sources that are related to iUTAH and of high interest to iUTAH participants, efforts will be made to incorporate those data into the iUTAH MDF.

These datasets technically fit in the Type D category above. The CI Team will work with the Data Policy Committee to prioritize datasets within this category.

6. Timing of Metadata and Dataset Submission and Availability

A metadata record for any datasets falling into the categories given by the data typology above (Type A – D) should be created and submitted to the iUTAH MDF within one month of the onset of data collection. Contact information for the CI Team and instructions for how to submit metadata will be available via the iUTAH Modeling & Data Federation website http://data.iutahepscor.org.

Metadata records will be reviewed by the iUTAH CI Team and will be made public and discoverable within one month of submission. General access to iUTAH datasets will follow a time frame specific to each data type. This policy establishes timing of data availability as outlined below:

Type A - Automated data streams from iUTAH facilities will be streamed directly into live databases and will be made available online in near real time. Quality controlled and derived data products from iUTAH facilities will be published within six months of data collection. All other primary datasets will be published within 3 months of the time they become available (e.g., as soon as results are created).

Type B - Finalized data will be submitted to the iUTAH MDF within one year of the completion of data creation activities. Students collecting data with iUTAH support must submit their finalized data as a condition of their thesis/dissertation defense. For long running data creation activities (i.e., efforts that last longer than one year), the following will be required:

a. The initial metadata description will be reviewed and updated at least once per year.
b. Intermediate data sets will be submitted for archival storage at least every 6 months. These data will not be published or released until the dataset is finalized by the data creator.
c. Finalized data will be submitted within one year of collection or by the end of the project, whichever comes first.

Type C - Type C datasets will be subject to time requirements described for Type A and Type B datasets. However, they may require the additional step of anonymization or aggregation with methods described in the Data Collection Plan.

Type D - Type D datasets will be published as soon as possible (within three months) and to the greatest extent allowable by the licensing, copyright, and/or data use agreements under which they were created/procured. Some Type D datasets may be permanently restricted and/or have regulated access limited to identified groups via password or other protections.

Data creators and collectors should have the reasonable expectation for the first rights to analysis and publication. The iUTAH Leadership Team may approve the suspension or extension of time frames for data publication given extenuating circumstances. ***In particular, graduate students and their advisors may apply for a time extension pending completion of analyses for theses and dissertations***. Additionally, datasets that do not fall into one of the categories above will be reviewed by the Data Policy Committee to determine the appropriate timeframe for publication. Appendix C contains a sample data management workflow that illustrates the order and timing of operations related to submitting datasets to the iUTAH MDF.

7. Storage and Archival

The iUTAH Project has made large investments in hardware and resources for the archival of datasets created by iUTAH researchers. All datasets loaded to the iUTAH MDF, datasets stored in shared working spaces, and datasets generated by environmental sensors will be archived on servers located within the Enterprise Data Center at Utah State University. Those machines have redundant backups in storage resources that have been established at the University of Utah. At minimum, datasets will be stored and archived for the duration of the iUTAH project. Additional support will be contingent on funding.

Additionally, where appropriate, iUTAH datasets will be registered with national data repositories such as the CUAHSI HIS, DataONE, or others. For example, streaming datasets from iUTAH facilities will be published using CUAHSI HIS WaterOneFlow web services that will be registered with the CUAHSI Water Data Center. This will broaden the impact of iUTAH datasets by making them discoverable through the data discovery tools provided by these national repositories.

8. Curation

The goal of iUTAH dataset curation is to ensure that data are available and can be reused in the indefinite future by iUTAH investigators and partners. This objective is supported by ensuring that metadata descriptions are thorough and complete. All datasets submitted to the iUTAH MDF will be subject to an informal review process to verify the completeness of the metadata and the integrity of the dataset. This moderation will be completed by members of the CI Team. Additionally, data curation includes quality control of environmental sensor data and anonymization of datasets containing sensitive information.

With regard to the data derived from environmental sensors within iUTAH facilities, the raw data will not be modified (unless due to an equipment configuration error). To ensure that the raw data are as reliable as possible, procedures for data quality assurance will be established and followed by the teams managing the iUTAH facilities and data collection activities. When accessing the raw data, users will be notified that the raw data are provisional. To account for sensor malfunction, anomalies, and drift, quality control procedures will also be established and followed. Quality controlled sensor data will be

made available within six months of the collection of raw data. Additionally, relevant statistics, plots, and summaries over time of the sensor data will be generated and published. When made available by the creators of the datasets, derived results of sensor data will also serve to increase the value of those data.

The value of datasets containing sensitive or personally identifiable information may be increased with secondary analyses, so forethought should be exercised (i.e., in the data collection plan) to:

1. Where possible, plan to make the primary data available to the largest number of iUTAH researchers as possible
2. Where possible, design and structure the study and data collection instrument(s) such that required redaction of the data does not make it useless to a broad audience
3. Provide detailed metadata (even in excess of the required metadata) so that the context is clear to other researchers.

9. Data Access, Use, and Citation

The primary method for accessing data published within the iUTAH MDF will be web-based and integrated with the iUTAH MDF website at http://data.iutahepscor.org. Because of the variety in types and formats of data collected and served by the iUTAH MDF and the fact that iUTAH data are intended for various audiences, there will be multiple channels for data presentation and dissemination. In general, access to datasets with any restrictions will require authentication and authorization on any servers and web portals that provide data access. All metadata submitted to the iUTAH MDF will be public to ensure that all datasets are discoverable, even if access to particular data is restricted to particular users. Submitted metadata will identify appropriate individuals who can be contacted for potential access to restricted data.

Data from streaming environmental sensors will be presented via online graphics and plots as well as through web services that can be accessed through writing computer code to retrieve the data in standards compliant file formats - e.g., Water Markup Language (WaterML) for streaming sensor data. Geospatial data will be presented via file based downloads as well as via online maps for selected datasets. For other datasets, efforts will be made to present the data content with graphics and maps, where possible. In all cases, available spatial, temporal, and other contextual metadata will be displayed and will be available for download. It is expected that the CI Team will work with the iUTAH EOD team to identify appropriate methods for presenting iUTAH data to stakeholders with various levels of technical expertise and need.

Accessing published data and research projects will be subject to the iUTAH Data Use Agreement in Appendix D of this document. Reuse of iUTAH data is encouraged, and acceptable uses of iUTAH data include any research, educational, governmental, recreational, or other not-for-profit purpose. All data and research products published by the iUTAH MDF will include appropriate attribute and citation information and will be

citable for use in peer-reviewed publications. The citation will identify and acknowledge the creator(s) and collaborator(s) as well as iUTAH as the data source. Digital Object Identifiers (DOIs) or other resolvable identifiers will be acquired for a select number of iUTAH datasets for which there is high interest. For other datasets, a URL will be assigned so that there is a permanent, online point of access.

For data products that may not be immediately disseminated to the general public, groups of researchers may wish to share derived data, model input packages, or simulation results within their group before they are published for wide distribution. CI for private research group working areas will be available, and it is anticipated that access will be at the discretion of the members of the group to protect the intellectual property that they create.

Options for access to datasets external to iUTAH include establishing services for retrieval, retrieving the data and storing them locally, incorporating the data dynamically, or statically pointing to it in its original location. In general, the preferred method will be to use existing web services or data services for retrieving data. External datasets will be subject to the data access and use agreements from the original data source.

10. Enforcement

It is acknowledged that there are situations in which it may be difficult for investigators to be in compliance with this policy. Some of these situations include:

1. Inadequate resources for data QA/QC or data upload resulting from underestimating time or resources needed to complete these activities.
2. Lack of skill and training in data curation.
3. Failure of proposed data collection methods or apparatus to produce the anticipated results.
4. Refusal of investigators to share data within the timelines prescribed above.

As described above, it is the responsibility of the iUTAH Leadership Team to monitor iUTAH-funded data collection activities and ensure that metadata and data are provided in compliance with this policy. The iUTAH Leadership Team may work in collaboration with the leads of the Research Focus Areas, EOD teams, and the CI Team to identify and resolve any potential issues. Resolution of potential issues falling in categories 1 and 2 may involve allocation of additional time for data collection/documentation/delivery or additional training or assistance from the CI Team. Potential issues of type 3 will be evaluated and may require a new Data Collection Plan that identifies how issues with data collection methods will be mitigated if data collection is to continue. Issues of type 4 may result in elimination of future funding or removal of individuals from the project. Failure to resolve issues of any of the types above may result in the iUTAH Leadership Team withholding funding for subsequent data collection activities.

10.1. Resolution of Conflicts

Unresolved disagreements between the Data Policy Committee and data collectors about the necessity of creating and content of a DCP will be elevated to the iUTAH Leadership Team for resolution. As this policy will be retroactive, ongoing and previous data collection efforts will be required to submit a data collection plan that addresses the points above.

In cases where disagreements occur within collaborative data collection teams regarding procedures regulating access to data collected using iUTAH resources, these instances can be elevated to the iUTAH Leadership Team for resolution.

References

DataONE Project Team. 2011. DataONE structure and potential partnership as a member node. Accessed: September 14, 2011. Available from: http://www.dataone.org/content/dataone-structure-and-potential-member-node.

de la Beaujardiere , J. (editor) 2006. Open GIS web page map server implementation specification, OGC implementation specification OGC 06-042, version 1.3.0, http://portal.opengeospatial.org/files/?artifact_id=14416, pp.

Horsburgh, J.S., D.G. Tarboton, D.R. Maidment, and I. Zaslavsky. 2008. A Relational Model for Environmental and Water Resources Data. Water Resour. Res. 44:W05406.

Horsburgh, J.S., D.G. Tarboton, M. Piasecki, D.R. Maidment, I. Zaslavsky, D. Valentine, and T.Whitenack. 2009. An integrated system for publishing environmental observations data. Environmental Modelling & Software 24(8):879-888.

Horsburgh, J.S., D.G. Tarboton, D.R. Maidment, and I. Zaslavsky. 2010a. Components of an integrated environmental observatory information system. Computers & Geosciences 37(2):207-218.

Horsburgh, J.S., D.G. Tarboton, K.A.T. Schreuders, D.R. Maidment, I. Zaslavsky, and D. Valentine. 2010b. Hydroserver: A Platform for Publishing Space-Time Hydrologic Datasets. in 2010 AWRA Spring Specialty Conference Geographic Information Systems (GIS) and Water Resources VI. Orlando Florida, American Water Resources Association, Middleburg, Virginia, TPS-10-1.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age; Committee on Science, Engineering, and Public Policy (COSEPUP); Policy and Global Affairs (PGA); Institute of Medicine (IOM); National Academy of Sciences. 2009. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. National Academies Press. Washington DC. http://www.nap.edu/catalog.php?record_id=12615

United States Office of Management and Budget (US OMB). 1999. OMB Circular A-110, "Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations", http://www.whitehouse.gov/omb/circulars_a110#36

Vretanos, P.A. (editor) 2010. Open GIS web feature service 2.0 interface standard. OGC implementation standard OGC 09-025r1 and IOS/DIS 19142, version 2.0.0, http://portal.opengeospatial.org/files/?artifact_id=39967, pp.

Whiteside, A. and J.D. Evans. 2008. Web coverage service (WCS) implementation standard. OGC implementation standard OGC 07-067r5, version 1.1.2, http://portal.opengeospatial.org/files/?artifact_id=27297, pp.

Zaslavsky, I., D. Valentine, and T. Whiteaker. 2007. CUAHSI WaterML, Open Geospatial Consortium Discussion Paper OGC 07-041r1. Accessed. Version 0.3.0:[Available from: http://portal.opengeospatial.org/files/?artifact_id=21743.]

Appendix A
iUTAH Data Management Plan Included in Original Proposal
(November 2011)

Types of Data: The iUTAH Modeling and Data Federation (see CI plan) will support the full data life cycle – with a focus on assisting iUTAH investigators in creating and sharing quality-controlled data and developing tools that assist researchers and other data consumers in discovering, visualizing, and accessing data in which they are interested. iUTAH data will consist of existing datasets from various data collection agencies (letters included) as well as new field and laboratory measurements related to social, hydroclimate, and ecological aspects of the water system. We will collect time series of observations from in situ sensors at fixed monitoring locations (e.g., flow and water quality gages, weather stations, atmospheric flux sites, etc.) as well as observations derived from field measurements and water quality samples. Geospatial datasets will include land use, terrain, hydrography (including natural, agricultural, and urban water distribution and drainage systems and impoundments), remote sensing vegetation imagery, and other landscape/geographic features describing our watershed observatories. Ecological data will include observations of water quality, ecosystem metabolism, aquatic biota, and riparian vegetation. Data defining the social and engineered aspects of the system will consist of agricultural and urban water supply and demand records, surface and groundwater withdrawals and deliveries, wastewater discharge, water conservation, agricultural and urban landscape irrigation rates, agricultural crop types and rotations, parcel level information on land use patterns, historic information on dynamics of water law and policy, and information about stakeholder goals, water management actions, and constraints. Finally, we anticipate that project investigators will share derived datasets resulting from analysis or modeling exercises using combinations of the above data types.

Development of the required repositories, protocols and methods for enabling shared access to the full range of iUTAH data is part of the intellectual contribution of this project. Development efforts will be devoted to creating repositories and/or web services that provide easy access to existing datasets and those created by this project. Where possible, we will use existing web services (e.g., those available already through the CUAHSI HIS) or develop data services that interface with existing, Internet-based systems for retrieving the primary data that we require (e.g., the USGS Seamless Data Server for the National Land Cover Dataset or National Elevation Datasets). The services that we develop will increase the availability of these datasets to data consumers both within and outside of our consortium.

Other Research Products: iUTAH will collaborate with the private environmental engineering and sensor community to produce a series of workshop and user manual materials related to designing watershed observatories, best practices for the implementation, operations and management of complex watershed observatory facilities. We will produce training materials for faculty, graduate students and technicians designed to maximize the usability of watershed observatories and complex, interdisciplinary data sets. All of these materials will be made publicly available on the iUTAH website and

published in appropriate journals and trade literature. We will also communicate closely with NEON personnel with regards to our experiences, findings and best practices. Training materials, models and course materials will be made available through our interactive 'situation room' sessions and will be made more widely and generally available through our museum activities (see External Engagement). Data and Metadata Standards: iUTAH will make full use of existing and emerging standards for sharing environmental data. Time series data collected by this project will be stored using the CUAHSI HIS Observations Data Model (ODM) (Horsburgh et al., 2008) and published in Water Markup Language (WaterML) format using CUAHSI HIS WaterOneFlow web services (Horsburgh et al., 2009; Zaslavsky et al., 2007). As a new, international standard for hydrologic data exchange, WaterML 2.0 is being developed through the Hydrology Domain Working Group of the Open Geospatial Consortium (OGC) and the World Meteorological Organization. For sharing geospatial datasets, we will use existing OGC standard interfaces such as Web Map Services (de la Beaujardiere, 2006), Web Feature Services (Vretanos, 2010), and Web Coverage Services (Whiteside and Evans, 2008). We will also provide standard HTTP and FTP access to download datasets that are more easily used as files rather than services, but will ensure that published data are accompanied by appropriate metadata descriptions either delivered by the data service or accompanying the data download. Metadata descriptions for all datasets will conform to appropriate ISO, EML, and/or FGDC metadata specifications. By using these accepted standards for data interfaces and format encodings, we will ensure that iUTAH data are interoperable with existing data repositories such as the CUAHSI HIS and DataONE, as well as other standards-compliant data systems in the earth sciences.

Policies for Data and Research Products: Data collection, Quality Assurance/Quality Control, and annotation with metadata will be the responsibility of iUTAH researchers, but will be conducted in accordance with protocols to be developed by Consortium members. All of the primary datasets collected as part of this project will be made freely and publicly available using the data and metadata standards described above. The primary data and research product dissemination method will be web-based. Finalized datasets and research products will be published within one year or as soon as they are certified as publication ready. Up to the point of publication, access to primary datasets may be limited to iUTAH researchers and data managers. All data and research products published by iUTAH will include appropriate attribute and citation information, and accessing published data and research products may include agreement to an access/use agreement specified by the author and that is in conformance with established practices and policies to be agreed upon by the collaborating Universities. This will ensure that intellectual rights of iUTAH investigators/data publishers are protected while granting redistribution rights to iUTAH and its archiving collaborators for purposes of long-term data sharing.

Derived datasets and results from model simulations will be handled differently. Our goal is to promote collaboration among iUTAH researchers. Groups of researchers may wish to share derived data, model input packages, or simulation results within their group before they are published externally. We will provide private project collaboration spaces accessible only to project participants within which these activities can take place. Additionally, execution of models, particularly within HPC environments, may result in

large quantities of results, only a fraction of which may be considered publishable by the researcher. As such, we will establish the facilities required to publish model simulation results, enabling individual researchers to select and publish their results as they see fit. It is anticipated that access to the contents of private research group working areas will be at the discretion of the members of the group to protect the intellectual property that they create.

As a general policy, all source code developed by this project will be open source and will be distributed under an open source license to be agreed upon by Consortium members. We will use open source code repositories for our software development, which will enable us to coordinate our activities across multiple Universities and to engage developers and contributors from outside of the immediate project team who wish to contribute.

Plans for Archiving Data: We will establish new storage resources at USU and UU for this project that will be used for both temporary scratch space for HPC and other model simulations and for permanent archival and publication of finished datasets and research products. We will leverage storage space at the UU that will be purchased under the recent Track 2 award. A portion of this storage space will be redundant, providing backup capabilities between USU and UU to ensure that important data resources are highly available and protected. Researchers will have the option of moving a subset of their model results to archival storage resources for the purpose of curation and long-term access. We anticipate that researchers may want to curate both simulation results and the input data packages from which they were generated so that their results can be reproduced. Primary datasets, model input packages, and simulation results will be published, cataloged, and archived on production servers that we will establish within the USU and UU data centers. We will support at least two methods of data publication: 1) we will assist iUTAH researchers in publishing data as web services using technologies from the CUAHSI HIS HydroServer software (Horsburgh et al., 2009; Horsburgh et al., 2010a; Horsburgh et al., 2010b); and 2) we will establish web service interfaces for iUTAH to become a DataONE Member Node (DataONE Project Team, 2011), enabling curated iUTAH data to be indexed by the NSF-funded DataONE DataNet project. By doing so, iUTAH can participate in the robust, distributed DataONE network and take advantage of the indexing, archival, and discovery services that DataONE provides. Our efforts to link with both the CUAHSI HIS and DataONE systems will broaden the impact of published iUTAH data resources and encourage their use by a broader scientific community. We anticipate that curated data products published in the iUTAH modeling and data federation will be citable for use in peer-reviewed publications.

Appendix B
iUTAH Data Publisher's Agreement

To support a collaborative research environment, the iUTAH Modeling & Data Federation (MDF) publishes data generated by the iUTAH research community and makes these data available on the Internet. The iUTAH MDF also supports fulfillment of the National Science Foundation's data sharing policy (https://www.nsf.gov/bfa/dias/policy/dmp.jsp) and the provisions of the iUTAH Data Management Plan. Your participation will encourage scientific inquiry, enable new research exploration, and facilitate education by providing the scientific community and iUTAH partners with relevant, easily accessible data. Shared data will be made available to Data Users according to the standard iUTAH Data Use Agreement.

By your acceptance of these terms, you state that ***you are the creator of the data and have the right*** to publish the data. You further agree to the following provisions for ***yourself and any collaborators with whom you coordinate*** the submission of these data:

1. As the Contributor, I am solely responsible for the integrity of these data put forward for submission.
2. I will use data upload capabilities provided by the iUTAH MDF to submit my data.
3. I will provide current contact information, including name, organization, phone number, e-mail address, and acceptance of the standard iUTAH Data Use Agreement upon submitting my data.
4. I will provide a minimum set of associated metadata to sufficiently describe submitted data and to ensure that shared data are meaningful and useful to the scientific community and iUTAH partners.
5. I agree to work with the iUTAH MDF staff to ensure that these data are provided in readable formats.
6. I agree to make these data publicly available in accordance with an agreed upon timeline with no restrictions for use other than those stated in the standard iUTAH MDF Data Use Agreement.
7. I understand that upon uploading data to the iUTAH MDF, iUTAH will index metadata describing my dataset to support rapid data discovery via the iUTAH Metadata Catalog.
8. I understand that iUTAH will maintain a logging system that tracks downloads of these data for the purpose of reporting data usage and download statistics to the National Science Foundation.

Appendix C
Example Data Management Workflow

The following is an example data management workflow that is included here to provide guidance to iUTAH researchers who are creating data.

1. Data Collection Plan submitted to Data Policy Committee. Feedback provided if needed.
2. Data collection/creation commences.
3. Submit Metadata to iUTAH MDF Data Publication System.
    a. Initial metadata submitted according to the standard metadata format **within one month of the onset of data collection/creation.**
4. The metadata record will be held in a staging area, and the CI team will work with the data creator/submitter to resolve any questions regarding the metadata and to determine the best format for storing/serving the dataset.
5. After approval by the CI Team, the metadata record will be published in the iUTAH MDF and will be discoverable by potential data users **within one month of submission**.
6. Submit data to iUTAH MDF.
    a. Type I datasets to be widely generally available (e.g., GAMUT data) are submitted as soon as they are available, ideally **at the same time as the metadata are submitted.**
    b. Type II datasets created by graduate student research totally or partially funded by iUTAH should be submitted **within one year of the completion of data collection/creation or prior to the defense of the student thesis/dissertation** (whichever comes first).
    c. Type II datasets created by iUTAH faculty members and/or technicians should be submitted **within one year of the completion of data collection/creation.**
    d. Data subsets/updates for long running datasets should be submitted at least every 6 months.
    e. Normal timelines for data submission may not apply if a student and/or advisor has applied for and received extension or suspension of normal time frames.
7. Submitted datasets will be held in a staging area, and the CI team will review submissions to ensure that they are complete, interpretable, and consistent with the submitted metadata. The CI team will work with submitters/creators to acquire additional metadata, etc. if needed
8. Datasets will be published in the iUTAH MDF within one month of successful submission.

Appendix D
iUTAH Data Use Agreement

This document outlines the provisions of a non-exclusive license for use of data shared through the iUTAH Modeling & Data Federation. Consistent with the objectives of iUTAH, the goal of the iUTAH Modeling & Data Federation is to make data originally acquired by iUTAH available to the community for further study. By receipt and use of data from the iUTAH Modeling & Data Federation, you agree to the following provisions for yourself and any collaborators with whom you share these data:

1. Free use of iUTAH Data: All iUTAH Data Products* except those labeled Restricted** are released to the public under a Creative Commons Attribution copyright license (http://creativecommons.org/licenses/by/3.0/us/) and may be freely copied, distributed, edited, and otherwise modified under the condition that you give acknowledgment as described below. Non-iUTAH data products, such as those produced by state and federal agencies have their own use policies that should be followed.

* iUTAH Data Products are defined as data collected with any monetary or logistical support from iUTAH.
** Restricted data are defined as data that cannot be released publicly due to privacy granted by human subjects legislation or other concerns. To enquire about potential use of restricted data, please contact us.

2. Data Guarantee: These data and metadata are provided by the iUTAH Modeling & Data Federation and the data contributors "as is." The Data User holds all parties involved in the production or distribution of the data and metadata harmless for any damages resulting from its use or interpretation.

3. Publication / Acknowledgement of Data Use: Acknowledgement of iUTAH and the data provider(s) is expected as standard practice in scientific publication or presentation of findings based upon these data. For iUTAH Data Products, the Data User should acknowledge the institutional support and funding award for the iUTAH project in any publication where the data contributed significantly to its content. For example:

"Data were provided by the iUTAH project and were accessed through the iUTAH Modeling & Data Federation. Significant funding for collection of these data was provided by the National Science Foundation (NSF EPS - 1208732)."

Whenever practical, the individual data providers should be acknowledged, including citing datasets as follows (see http://www.datacite.org/whycitedata):

Creator(s), year of publication, title of dataset, name of publisher (iUTAH Modeling & Data Federation), edition or version, and URL or other identifier.

When impractical to do so because of the number of datasets used, use of data from the iUTAH Modeling & Data Federation should be acknowledged.

Data Users agree to provide to iUTAH a bibliographic citation of the final published presentation or article for inclusion in the iUTAH literature archive. I will submit this information via the iUTAH website at http://iutahepscor.org/contact_us_form.php.

4. Redistribution of Data: Redistribution of original data is permitted so long as the data are redistributed under the same terms and conditions as described in this Data Use Agreement. Data derived from original data may be distributed under terms and conditions established by the creators of such derived data. Users must comply with the terms and conditions of use set by the creators of these derived data.

5. Collaboration: iUTAH researchers have released data in the spirit of open scientific collaboration. Data Users are strongly encouraged to consider consultation, collaboration, and/or co-authorship with dataset creators.

By accepting this Data Set, the Data User agrees to abide by the terms of this agreement. The Data Owner shall have the right to terminate this agreement immediately by written notice upon the Data User's breach of, or non-compliance with, any of its terms. The Data User may be held responsible for any misuse that is caused or encouraged by the Data User's failure to abide by the terms of this agreement.