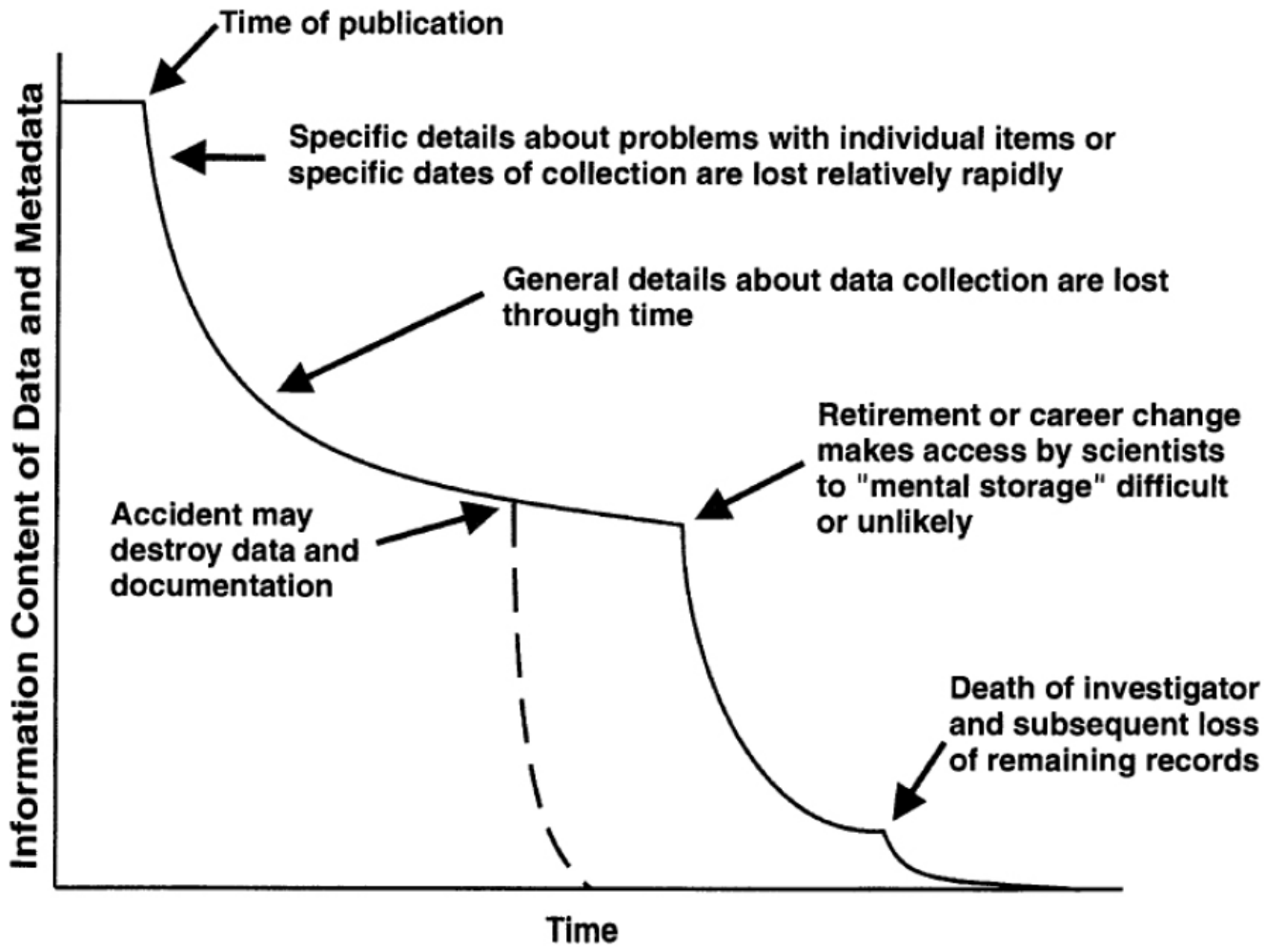
iUTAH Data Management Update

Amber Spackman Jones
Jeffery S. Horsburgh
iUTAH Summer Symposium
7-17-2015

Information Entropy



Example of the normal degradation in information content associated with data and metadata over time ("information entropy").

Michener, W.K. (2006). Meta-information concepts for ecological data management, Ecological Informatics, 1(1), 3-7, http://dx.doi.org/10.1016/j.ecoinf.2005.08.004.

Information Entropy

"Do not underestimate your ability to forget details about a study!"

Borer, E.T., Seabloom, E.W., Jones, M.B., Schildhauer, M. (2009). Some simple guidelines for effective data management. Bulletin of the Ecological Society of America 90:205-214. dx.doi.org/10.1890/0012-9623-90.2.205

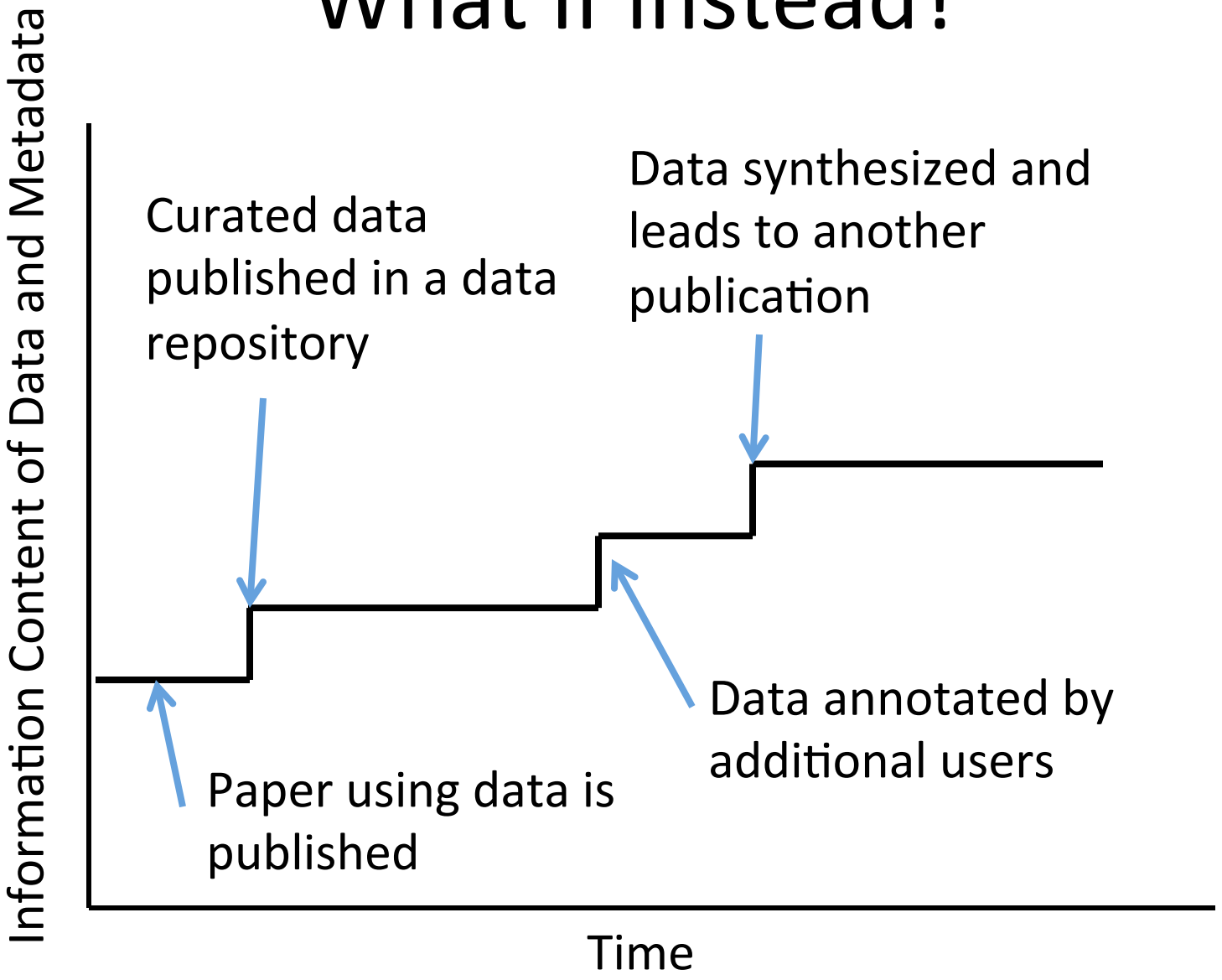
"If the information on an observation is lost, it is lost forever because it is almost impossible to measure the observation again in the original context."

Specht, A., Guru, S., Houghton, L., Keniger, L., Driver, P., Ritchie, E.G., Lai, K., Treloar, A. (2015). Data management challenges in analysis and synthesis in the ecosystem sciences. Science of the Total Environment. Dx.doi.org/10.1016/j.scitotenv.2015.03.092

If the rewards of the data deluge are to be reaped, then researchers who produce those data must share them, and do so in such a way that the data are interpretable and reusable by others.

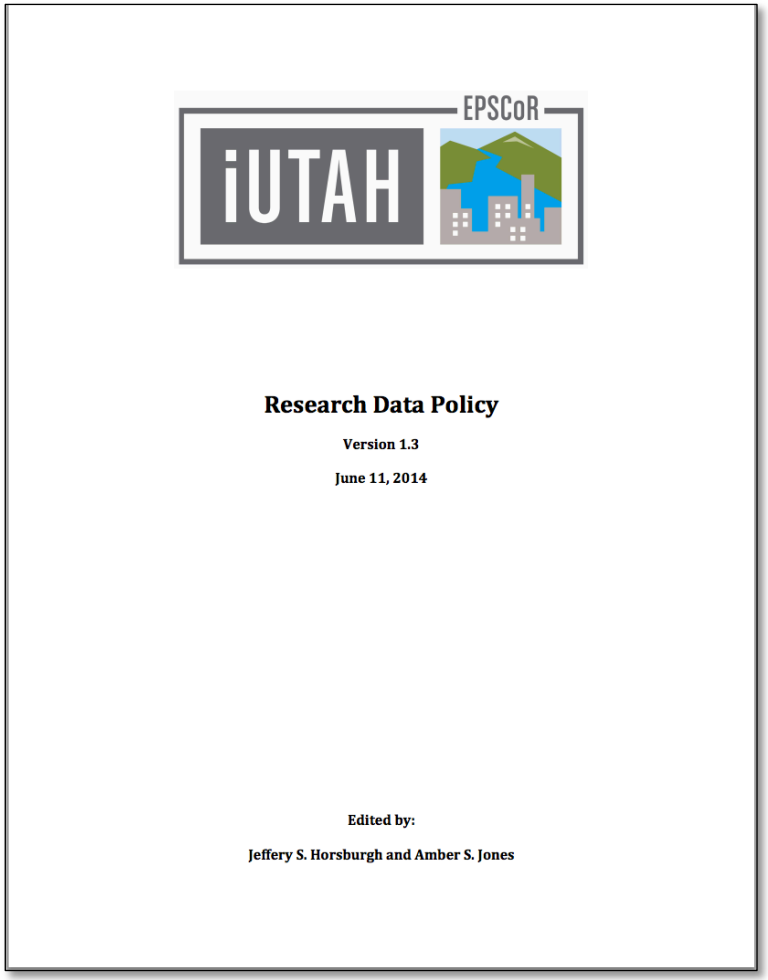
Borgman, C.L. (2012). The conundrum of sharing research data. Journal of the American Society for Information Science and Technology 63(6): 1059-1078. dx.doi.org/10.1002/asi.22634

What if instead?



iUTAH Data Policy

- **Applies to all datasets created or developed with support from iUTAH**
- Recommended reading for all investigators. Available at: http://iutahepscor.org/resources/documents/iUTAH_Data_Policy.pdf
- In general, **open data policy** to maximize the impact and broad use of datasets collected by iUTAH research teams
- Researchers should have expectation of **first rights** to analyze and publish data
- Researchers expected to provide **high quality datasets** with sufficient metadata
- Data products should to be **considered at outset** of data collection



iUTAH Data Policy: Data Typology

Timeframes

A. Primary iUTAH datasets and research products (e.g., raw and QAQC sensor data, baseline sampling datasets, general iUTAH datasets for the iUTAH community).

→ Published as soon as results are created

B. Support from iUTAH provided, but created by a specific investigator or group to support particular research question/goal.

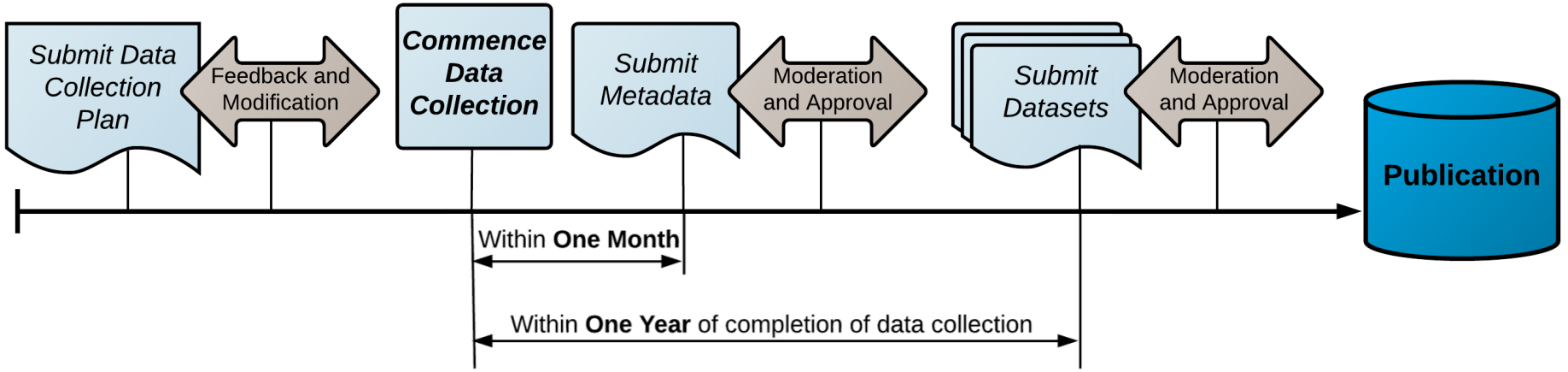
→ Finalized data submitted within one year of completion of data collection activities

C. Types A and B but subject to IRB restrictions.

→ Same timeframes as A and B, but may require anonymization

D. Proprietary data that may be subject to licensing, copyright, other restrictions.

Data Publication Process



Publication:

- Assigns a citation and a URL
- Data and metadata are discoverable
- Datasets are archived and curated

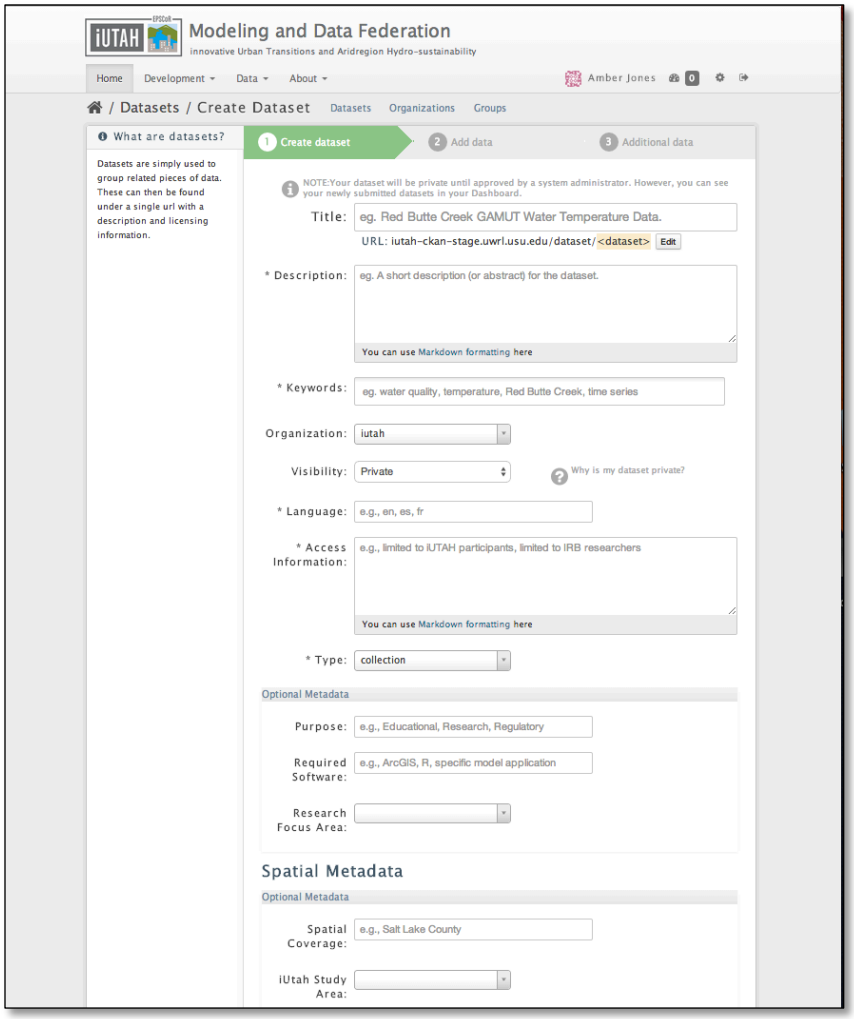
iUTAH Data Policy: Data Collection Plans

- **ALL** data creation efforts with **ANY** funding from iUTAH (salary, travel, sampling, equipment, etc.) **MUST** submit a brief plan to the Data Policy Committee **PRIOR** to funding.
- Plan should include:
  1. Identification of types of data to be collected/created
  2. Brief description of methods, data formats, and data products
  3. Timeline for data generation and expected publication
  4. Identification of who will have access to preliminary data during collection
  5. Identification of limits to access
  6. Information on collaborators/co-authors of data products or publications

Data Publication System

http://repository.iutahepscor.org

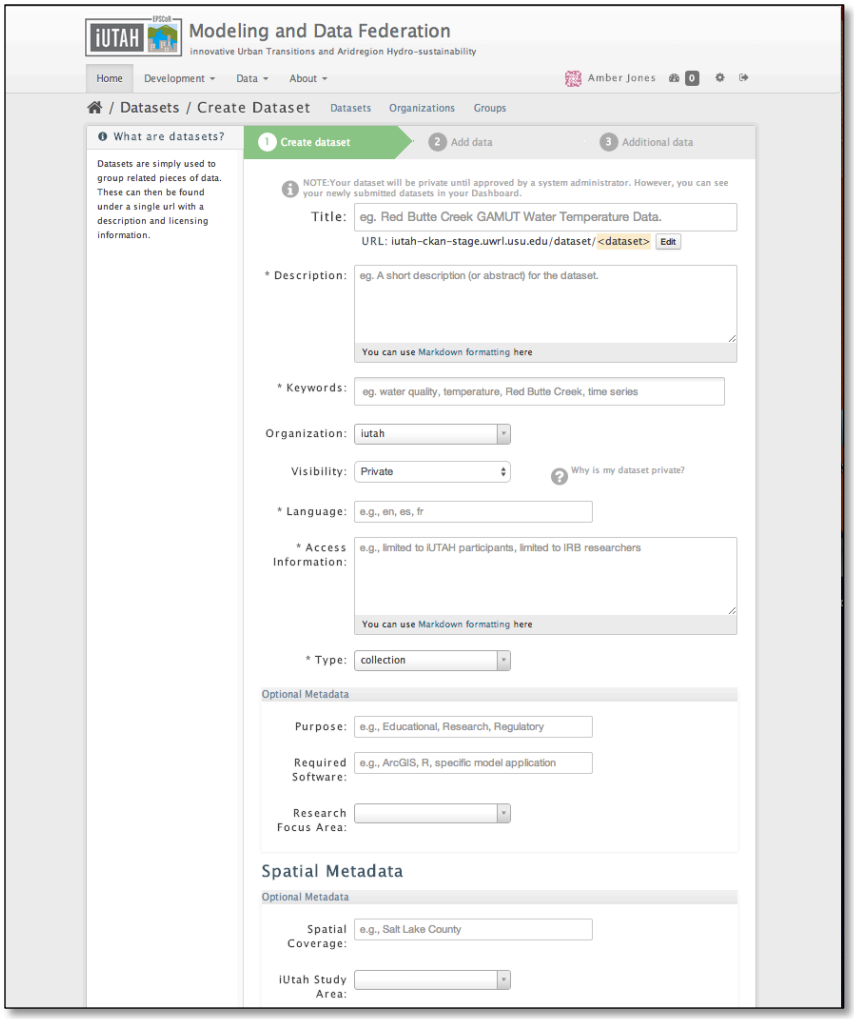
- Web-based system for iUTAH researchers to submit and publish data and models.
- System supports **curation** of datasets.
- **Integrates** the submission and presentation of data and metadata.
- Supports **discovery and access** of datasets to a wide audience.
- Supports **storage and archival**.
- Datasets are private until **approved** by a moderator.



Data Publication System

http://repository.iutahepscor.org

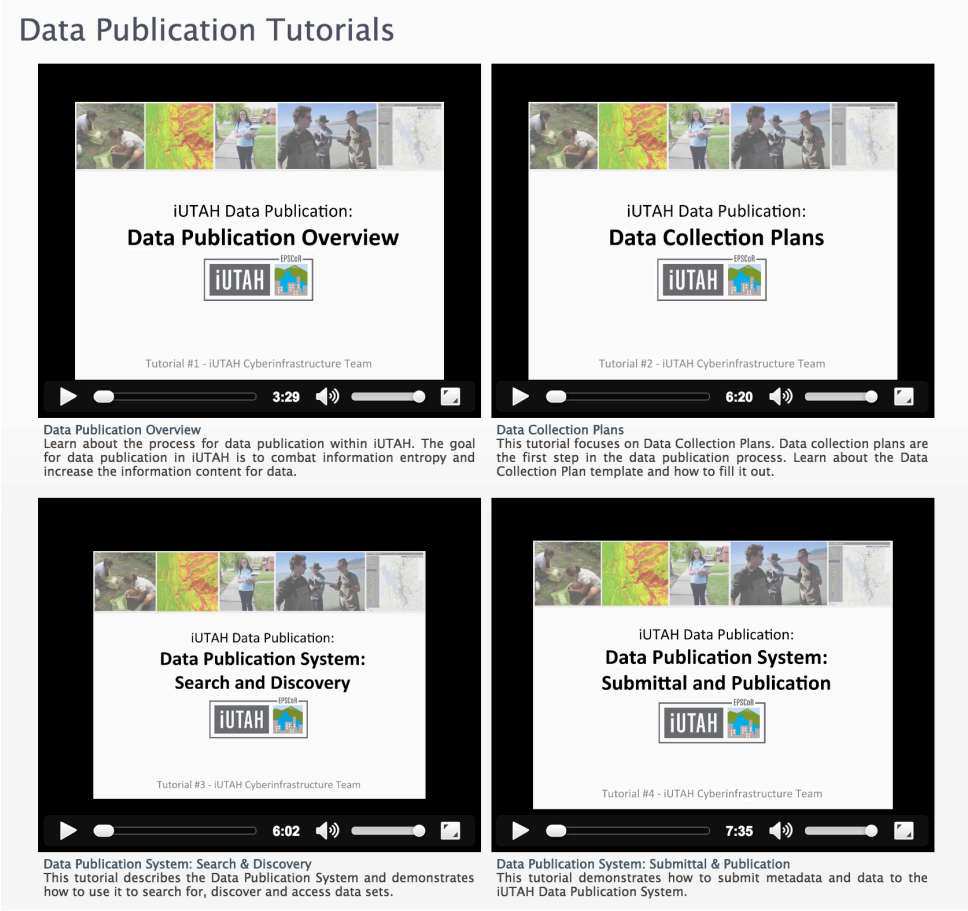
- Organized into **datasets**. Each dataset consists of multiple resources.
- Supports submission of **metadata-only** record.
- **General level metadata**. More specific metadata may be submitted as a resource.
- Metadata records submitted to **provide insight** into what work iUTAH participants are conducting.



Data Publication Tutorials

http://iutahepscor.org/data_modeling.shtml
http://data.iutahepscor.org/mdf/About/Training_Materials/

- Developed short videos to provide training on aspects of data publication.
- Four videos:
  - Overview
  - Data Collection Plans
  - Data Publication System: Search and Discovery
  - Data Publication System: Submittal and Publication



Data Submission Best Practices

Use Descriptive File Names

- Use only plain ASCII characters
- Brief, but descriptive of content
- Include a “readme” file when using many files

Archive Data in Non-Proprietary Formats

- Microsoft Excel is widely available and used now, but what about in 10 years? 20 years?
- How many other software programs can open your data?
- Will your data disappear if the file format/software become obsolete?

Data Submission Best Practices

What format to use?

- Store it in a file format that can be used by many different software programs
  - Text files – e.g., comma separated values (CSV) for tabular data
- Use a standard file format accepted by your scientific community
- Consider both format and syntax (e.g., the structure within the file)

Energy Balance of and Isolated Urban Tree

Followers
0

Follow

Organization

iUTAH

iUTAH, innovative Urb
Transitions and Aridr
Hydro-sustainability,
statewide effort dedic
maintaining and impr
water sustainability i
Funded by the... read

Social

Google+

Twitter

Facebook

License

Dataset | Activity Stream | Related | Edit

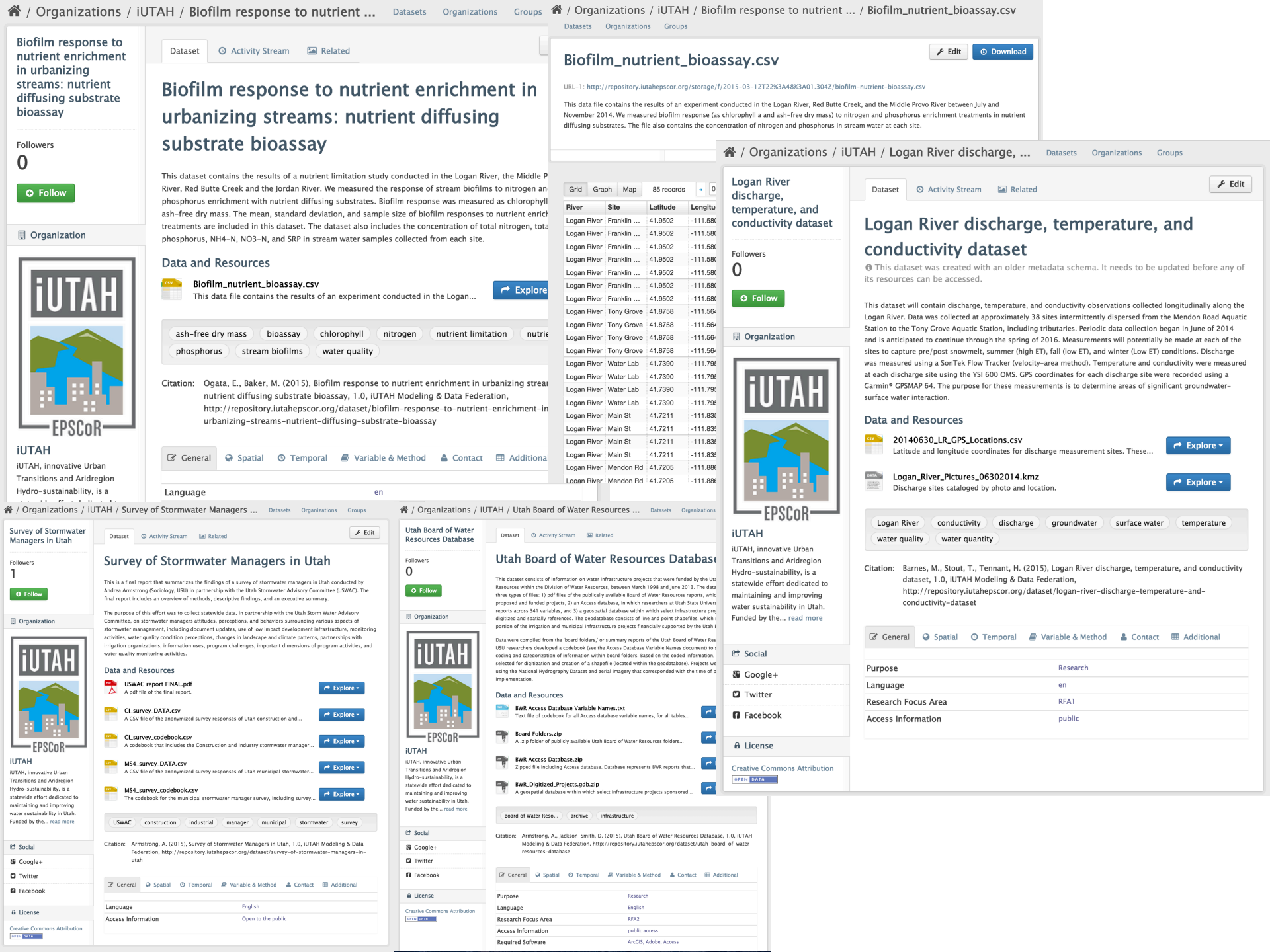
Energy Balance of and Isolated Urban Tree

This dataset contains energy-related measurements in and around a relatively isolated urban tree on the University of Utah campus. Several measurements techniques are used to describe the spatial distribution of surface temperature, and fluxes of sensible heat, latent heat, radiation, and moisture. The tree is a Freeman

repository.iutahepscor.org/storage/f/2015-02-26T23%3A26%3A09.192Z/.readme

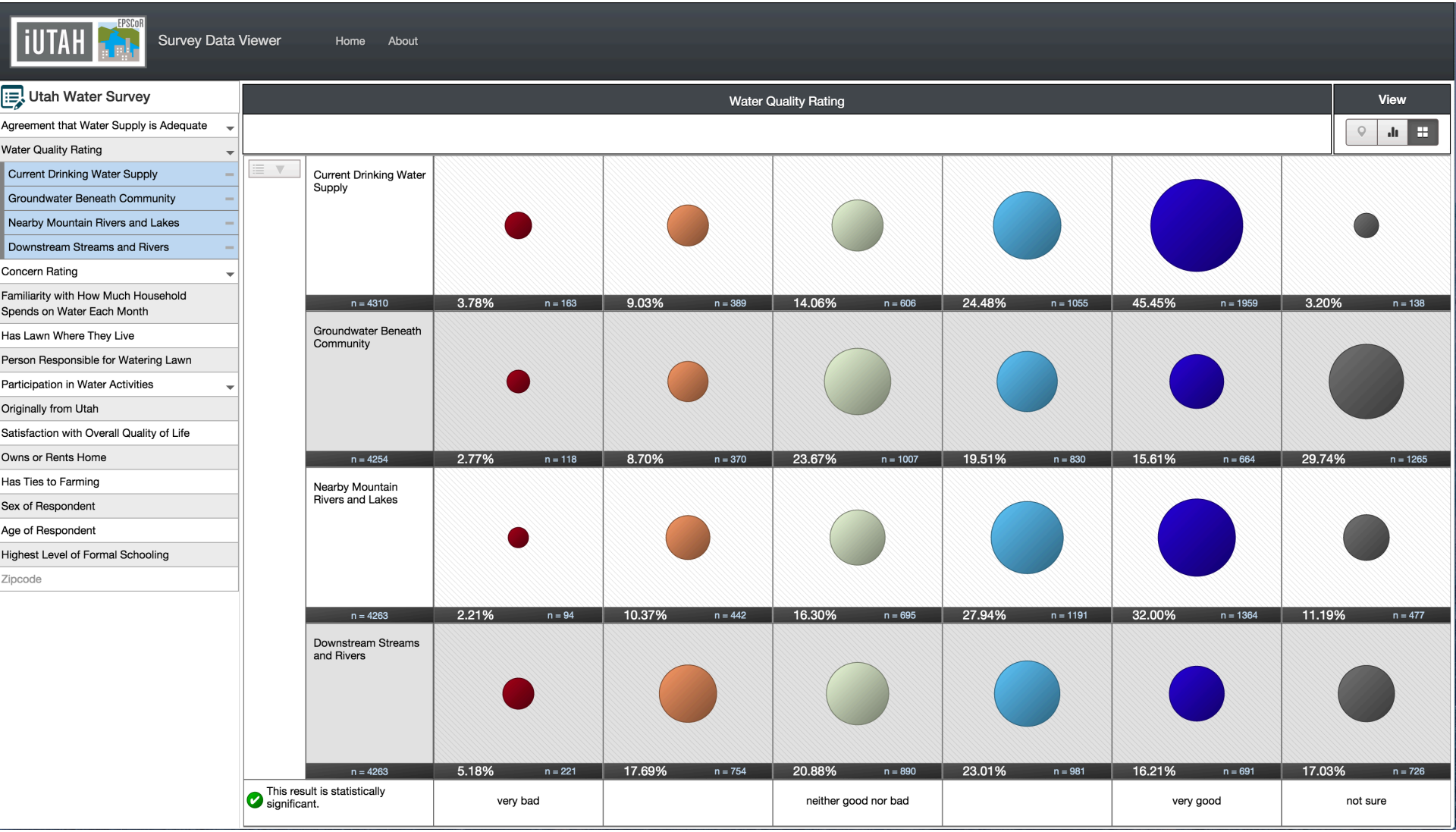
```
Isolate Tree (IsoTree) Dataset Metadata

Contact:
Brian Bailey
bbailey@eng.utah.edu

Coordinate System:
+x - North
+y - East
+z - upward
(x,y,z)=(0,0,0) @ (tree trunk, tree trunk, ground @ tree)

--------------------------------------------------------------------------------
Tower Data Tables
--------------------------------------------------------------------------------
Data table files prepended with 'Tower' correspond to measurements taken on the sonic tower.  Each file corresponds to a different day of the year. The tower was located at cooridate position (9.10 m,6.65 m,0).

Sampling Rate: 20 Hz

Equipment:
1. CSAT3: Campbell Scientific CSAT3 three-dimensional sonic anemometer. Measurements taken at heights of 1.5m, 4.0m, 7m, 10m. Manual: https://s.campbellsci.com/documents/us/manuals/csat3.pdf
2. EC150: Campbell Scientific EC150 open-path gas analyzer. Measurements taken at 4m (co-located with CSAT3). Manual: http://s.campbellsci.com/documents/us/manuals/ec150.pdf
3. thermocouple: Omega engineering type-E finewire thermocouples.
4. KH20: Campbell Scientific KH20 krypton hygrometer. Measurements taken at 4m (co-located with CSAT3). Manual: https://s.campbellsci.com/documents/us/manuals/kh20.pdf

Data Table:
Column titles in the data table are listed below along with an explanation of the value
1. year
2. day of year: Julian day of year (Jan 1 = 1, etc.)
3. time: hour and minute of measurement (local time, MDT)
4. seconds: second of measurement
5. KH20 H2O: water vapor concentration from KH20 (grams H20 per meter^3 air)
6. KH20 mV: raw KH20 voltage reading (millivolts)
7,12,17,22. x-wind @ XXm: wind speed from CSAT3 at heights of 1.5m, 4.0m, 7m, 10m in x-dir +North (meters/second)
8,13,18,23. y-wind @ XXm: wind speed from CSAT3 at heights of 1.5m, 4.0m, 7m, 10m in y-dir +East (meters/second)
9,14,19,24. z-wind @ XXm: wind speed from CSAT3 at heights of 1.5m, 4.0m, 7m, 10m in z-dir +up (meters/second)
10,15,20,25. sonic temp @ XXm: air temperature measured by CSAT3 at heights of 1.5m, 4.0m, 7m, 10m (degrees Celcius)
11,16,21,26: sonic diagnostic @ XXm: diagnostic flag from CSAT3 at heights of 1.5m, 4.0m, 7m, 10m
27. CO2 @ 4m: air CO2 concentration measured by EC150 at a height of 4m (milligrams CO2 per meter^3 air)
28. H2O @ 4m: air H2O concentration measured by EC150 at a height of 4m (grams H2O per meter^3 air)
29. gas diagnostic @ 4m: diagnostic flat from EC150 at a height of 4m
30. EC150 temp @ 4m: air temperature measured by EC150 at a height of 4m (degrees Celcius)
31. EC150 pressure @ 4m: air pressure measured by EC150 at a height of 4m (kiloPascals)
32. CO2 signal @ 4m: EC150 CO2 signal strength at a height of 4m
33. H2O signal @ 4m: EC150 H2O signal strength at a height of 4m
34,35,36,37. air temp @ XXm: air temperature from finewire thermocouples at heights of 1.5m, 4.0m, 7m, 10m (degrees Celcius)

--------------------------------------------------------------------------------
Tree Data Tables
--------------------------------------------------------------------------------
Data table files prepended with 'Tree' correspond to measurements taken at the tree.  Each file corresponds to a different day of the year.

Sampling Rate: 1 Hz

Equipment:
1. CNR1: Campbell Scientific CNR1 net radiometer. Manual: http://s.campbellsci.com/documents/au/manuals/cnr1.pdf.  Consists of four independent radiometers:
   >SW: upward- and downward-facing pyrenometers which measure shortwave radiation with wavelengths between 305-2,800 nanometers
   >LW: upward- and downward-facing pyrgeometers which measure longwave radiation with wavelengths between 5,000-50,000 nanometers
2. SQ-100: Apogee Instruments SQ-100 quantum (PAR) sensors. Manual: https://www.apogeeinstruments.com/content/SQ-100_300manual.pdf
3. LI-200: LI-COR pyranometer which measures shortwave radiation with wavelengths between 400-1,100 nanometers. Manual: ftp://ftp.licor.com/perm/env/Radiation_Sensors/Manual/TerrestrialSensors_Manual.pdf
4. leaf thermocouples: Omega engineering 40 AWG type-E thermocouples. Thermocouples were affixed to the lower side of various real and replica leaves using a combination of 3M `transpore' medical tape and Omega
engineering `Omegatherm' thermally conductive paste. The replica leaves consited of 0.5mm copper cut into the approximate shape of the real leaves, and were placed in approximately the same location and
orientation as the real leaves
```

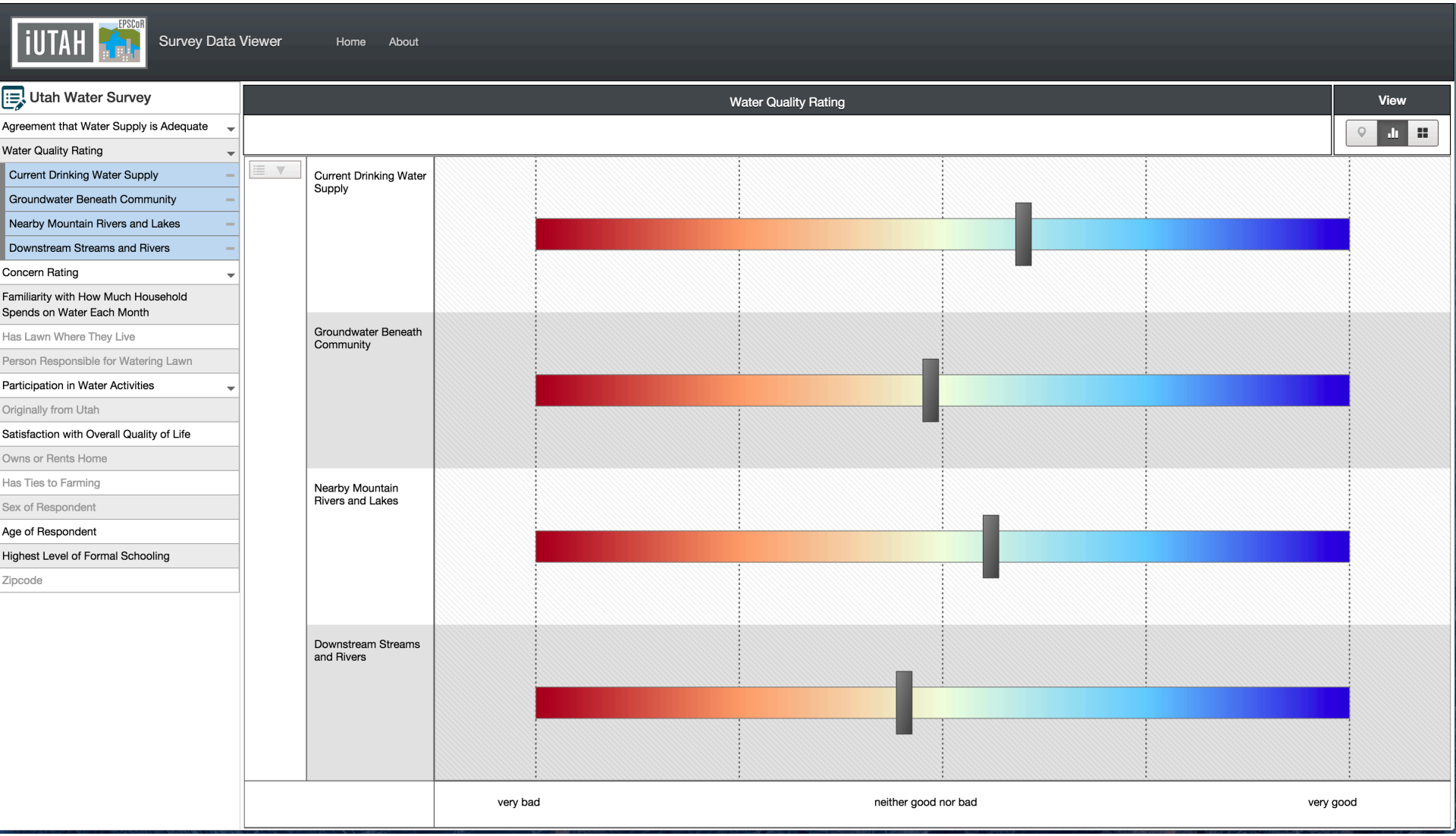
Citation: Bailey, B., Stoll, R., Pardyjak, E. (2014), Energy Balance of and Isolated Urban Tree, 1.0, iUTAH Modeling & Data Federation, http://repository.iutahepscor.org/dataset/energy-balance-of-and-isolated-urban-tree


Organizations / iUTAH / Biofilm response to nutrient ...
Datasets Organizations Groups
Biofilm response to nutrient enrichment in urbanizing streams: nutrient diffusing substrate bioassay
Followers 0
Follow
Organization
iUTAH
iUTAH, innovative Urban Transitions and Aridregion Hydro-sustainability, is a
Dataset Activity Stream Related
Biofilm response to nutrient enrichment in urbanizing streams: nutrient diffusing substrate bioassay
This dataset contains the results of a nutrient limitation study conducted in the Logan River, the Middle P
River, Red Butte Creek and the Jordan River. We measured the response of stream biofilms to nitrogen an
phosphorus enrichment with nutrient diffusing substrates. Biofilm response was measured as chlorophyll
ash-free dry mass. The mean, standard deviation, and sample size of biofilm responses to nutrient enrich
treatments are included in this dataset. The dataset also includes the concentration of total nitrogen, tota
phosphorus, NH4-N, NO3-N, and SRP in stream water samples collected from each site.
Data and Resources
Biofilm_nutrient_bioassay.csv
This data file contains the results of an experiment conducted in the Logan...
Explore
ash-free dry mass bioassay chlorophyll nitrogen nutrient limitation
phosphorus stream biofilms water quality
Citation: Ogata, E., Baker, M. (2015), Biofilm response to nutrient enrichment in urbanizing strea
nutrient diffusing substrate bioassay, 1.0, iUTAH Modeling & Data Federation,
http://repository.iutahepscor.org/dataset/biofilm-response-to-nutrient-enrichment-in
urbanizing-streams-nutrient-diffusing-substrate-bioassay
General Spatial Temporal Variable & Method Contact Additional
Language en
Organizations / iUTAH / Biofilm response to nutrient ... / Biofilm_nutrient_bioassay.csv
Datasets Organizations Groups
Biofilm_nutrient_bioassay.csv
Edit Download
URL-1: http://repository.iutahepscor.org/storage/f/2015-03-12T22%3A48%3A01.304Z/biofilm-nutrient-bioassay.csv
This data file contains the results of an experiment conducted in the Logan River, Red Butte Creek, and the Middle Provo River between July and November 2014. We measured biofilm response (as chlorophyll a and ash-free dry mass) to nitrogen and phosphorus enrichment treatments in nutrient diffusing substrates. The file also contains the concentration of nitrogen and phosphorus in stream water at each site.
Grid Graph Map 85 records
River Site Latitude Longitu
Logan River Franklin ... 41.9502 -111.58(
Logan River Tony Grove 41.8758 -111.564
Logan River Water Lab 41.7390 -111.79
Logan River Main St 41.7211 -111.83
Logan River Mendon Rd 41.7205 -111.88
Organizations / iUTAH / Logan River discharge, ...
Datasets Organizations Groups
Logan River discharge, temperature, and conductivity dataset
Followers 0
Follow
Organization
iUTAH
iUTAH, innovative Urban Transitions and Aridregion Hydro-sustainability, is a statewide effort dedicated to maintaining and improving water sustainability in Utah. Funded by the... read more
Social
Google+
Twitter
Facebook
License
Creative Commons Attribution
Dataset Activity Stream Related
Edit
Logan River discharge, temperature, and conductivity dataset
This dataset was created with an older metadata schema. It needs to be updated before any of its resources can be accessed.
This dataset will contain discharge, temperature, and conductivity observations collected longitudinally along the Logan River. Data was collected at approximately 38 sites intermittently dispersed from the Mendon Road Aquatic Station to the Tony Grove Aquatic Station, including tributaries. Periodic data collection began in June of 2014 and is anticipated to continue through the spring of 2016. Measurements will potentially be made at each of the sites to capture pre/post snowmelt, summer (high ET), fall (low ET), and winter (Low ET) conditions. Discharge was measured using a SonTek Flow Tracker (velocity-area method). Temperature and conductivity were measured at each discharge site using the YSI 600 OMS. GPS coordinates for each discharge site were recorded using a Garmin® GPSMAP 64. The purpose for these measurements is to determine areas of significant groundwater-surface water interaction.
Data and Resources
20140630_LR_GPS_Locations.csv
Latitude and longitude coordinates for discharge measurement sites. These...
Explore
Logan_River_Pictures_06302014.kmz
Discharge sites cataloged by photo and location.
Explore
Logan River conductivity discharge groundwater surface water temperature water quality water quantity
Citation: Barnes, M., Stout, T., Tennant, H. (2015), Logan River discharge, temperature, and conductivity dataset, 1.0, iUTAH Modeling & Data Federation, http://repository.iutahepscor.org/dataset/logan-river-discharge-temperature-and-conductivity-dataset
General Spatial Temporal Variable & Method Contact Additional
Purpose Research
Language en
Research Focus Area RFA1
Access Information public
Organizations / iUTAH / Survey of Stormwater Managers ...
Survey of Stormwater Managers in Utah
Followers 1
Follow
Organization
iUTAH
iUTAH, innovative Urban Transitions and Aridregion Hydro-sustainability, is a statewide effort dedicated to maintaining and improving water sustainability in Utah. Funded by the... read more
Social
Google+
Twitter
Facebook
License
Creative Commons Attribution
Dataset Activity Stream Related
Edit
Survey of Stormwater Managers in Utah
This is a final report that summarizes the findings of a survey of stormwater managers in Utah conducted by Andrea Armstrong (Sociology, USU) in partnership with the Utah Stormwater Advisory Committee (USWAC). The final report includes an overview of methods, descriptive findings, and an executive summary.
The purpose of this effort was to collect statewide data, in partnership with the Utah Storm Water Advisory Committee, on stormwater managers attitudes, perceptions, and behaviors surrounding various aspects of stormwater management, including document updates, use of low impact development infrastructure, monitoring activities, water quality condition perceptions, changes in landscape and climate patterns, partnerships with irrigation organizations, information uses, program challenges, important dimensions of program activities, and water quality monitoring activities.
Data and Resources
USWAC report FINAL.pdf
A pdf file of the final report.
Explore
CI_survey_DATA.csv
A CSV file of the anonymized survey responses of Utah construction and...
Explore
CI_survey_codebook.csv
A codebook that includes the Construction and Industry stormwater manager...
Explore
MS4_survey_DATA.csv
A CSV file of the anonymized survey responses of Utah municipal stormwater...
Explore
MS4_survey_codebook.csv
The codebook for the municipal stormwater manager survey, including survey...
Explore
USWAC construction industrial manager municipal stormwater survey
Citation: Armstrong, A. (2015), Survey of Stormwater Managers in Utah, 1.0, iUTAH Modeling & Data Federation, http://repository.iutahepscor.org/dataset/survey-of-stormwater-managers-in-utah
General Spatial Temporal Variable & Method Contact Additional
Language English
Access Information Open to the public
Organizations / iUTAH / Utah Board of Water Resources ...
Utah Board of Water Resources Database
Followers 0
Follow
Organization
iUTAH
iUTAH, innovative Urban Transitions and Aridregion Hydro-sustainability, is a statewide effort dedicated to maintaining and improving water sustainability in Utah. Funded by the... read more
Social
Google+
Twitter
Facebook
License
Creative Commons Attribution
Dataset Activity Stream Related
Utah Board of Water Resources Database
This dataset consists of information on water infrastructure projects that were funded by the Uta
Resources within the Division of Water Resources, between March 1998 and June 2013. The data
three types of files: 1) pdf files of the publically available Board of Water Resources reports, whic
proposed and funded projects, 2) an Access database, in which researchers at Utah State Univers
reports across 341 variables, and 3) a geospatial database within which select infrastructure pro
digitized and spatially referenced. The geodatabase consists of line and point shapefiles, whi
portion of the irrigation and municipal infrastructure projects financially supported by the Utah
Data were compiled from the "board folders," or summary reports of the Utah Board of Water Res
USU researchers developed a codebook (see the Access Database Variable Names document) to
coding and categorization of information within board folders. Based on the coded information,
selected for digitization and creation of a shapefile (located within the geodatabase). Projects we
using the National Hydrography Dataset and aerial imagery that corresponded with the time of p
implementation.
Data and Resources
BWR Access Database Variable Names.txt
Text file of codebook for all Access database variable names, for all tables...
Board Folders.zip
A .zip folder of publicly available Utah Board of Water Resources folders...
BWR Access Database.zip
Zipped file including Access database. Database represents BWR reports that...
BWR_Digitized_Projects.gdb.zip
A geospatial database within which select infrastructure projects sponsored...
Board of Water Reso... archive infrastructure
Citation: Armstrong, A., Jackson-Smith, D. (2015), Utah Board of Water Resources Database, 1.0, iUTAH Modeling & Data Federation, http://repository.iutahepscor.org/dataset/utah-board-of-water-resources-database
General Spatial Temporal Variable & Method Contact Additional
Purpose Research
Language English
Research Focus Area RFA2
Access Information public access
Required Software ArcGIS, Adobe, Access

Survey Data Viewer:

http://data.iutahepscor.org/surveys/




iUTAH
EPSCoR
Survey Data Viewer
Home
About
Utah Water Survey
Agreement that Water Supply is Adequate
Water Quality Rating
Current Drinking Water Supply
Groundwater Beneath Community
Nearby Mountain Rivers and Lakes
Downstream Streams and Rivers
Concern Rating
Familiarity with How Much Household Spends on Water Each Month
Has Lawn Where They Live
Person Responsible for Watering Lawn
Participation in Water Activities
Originally from Utah
Satisfaction with Overall Quality of Life
Owns or Rents Home
Has Ties to Farming
Sex of Respondent
Age of Respondent
Highest Level of Formal Schooling
Zipcode
Water Quality Rating
View
Current Drinking Water Supply
Groundwater Beneath Community
Nearby Mountain Rivers and Lakes
Downstream Streams and Rivers
very bad
neither good nor bad
very good

iUTAH EPSCoR Survey Data Viewer

Home About

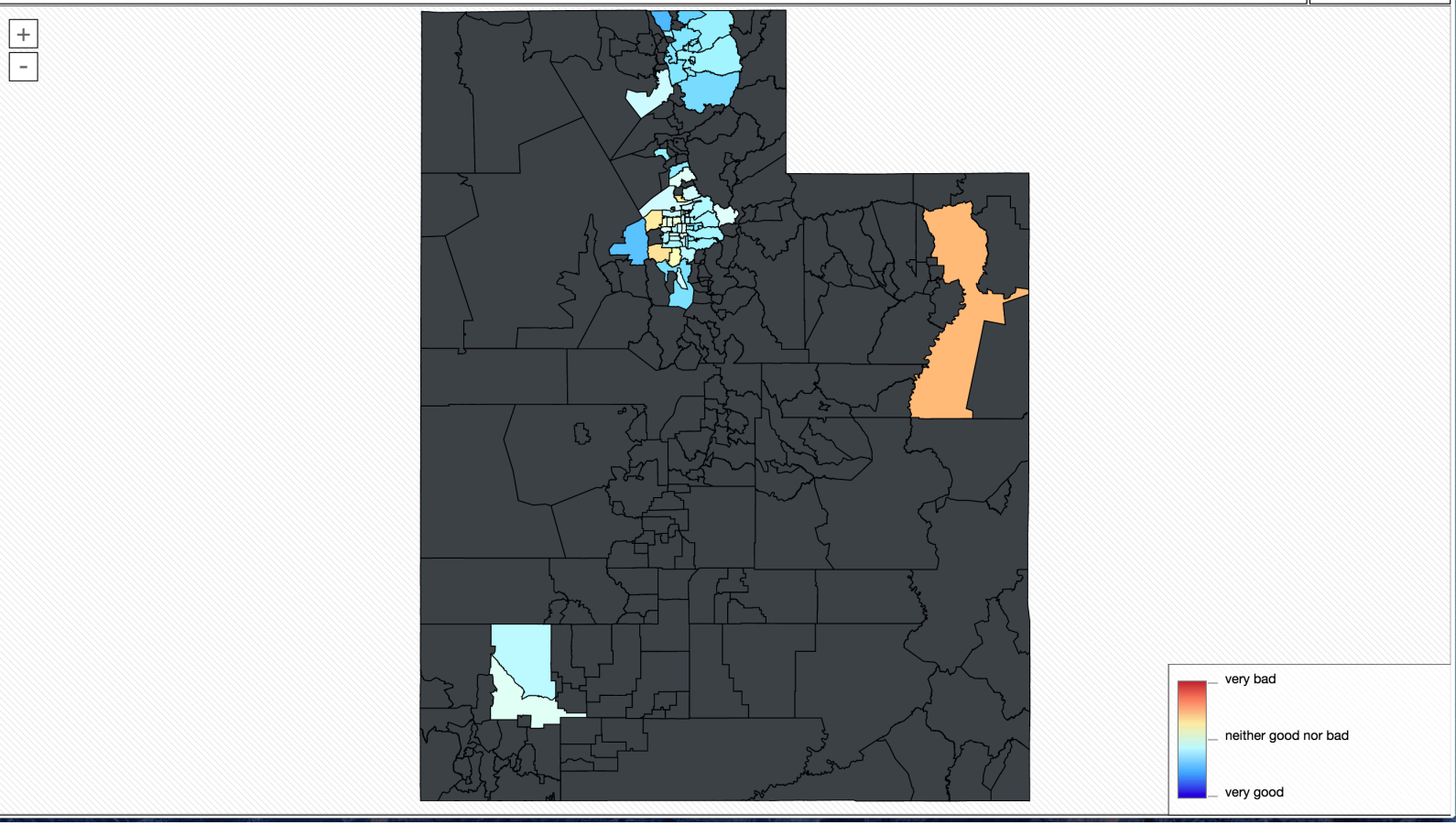
Utah Water Survey

- Agreement that Water Supply is Adequate
- Water Quality Rating
  - Current Drinking Water Supply
  - Groundwater Beneath Community
  - Nearby Mountain Rivers and Lakes
  - Downstream Streams and Rivers
- Concern Rating
- Familiarity with How Much Household Spends on Water Each Month
- Has Lawn Where They Live
- Person Responsible for Watering Lawn
- Participation in Water Activities
- Originally from Utah
- Satisfaction with Overall Quality of Life
- Owns or Rents Home
- Has Ties to Farming
- Sex of Respondent
- Age of Respondent
- Highest Level of Formal Schooling
- Zipcode

Water Quality Rating

Current Drinking Water Supply

View



Questions?

Amber Jones
amber.jones@usu.edu
Jeff Horsburgh
jeff.horsburgh@usu.edu

http://data.iutahepscor.org