



Cyberinfrastructure Update

Amber Spackman Jones

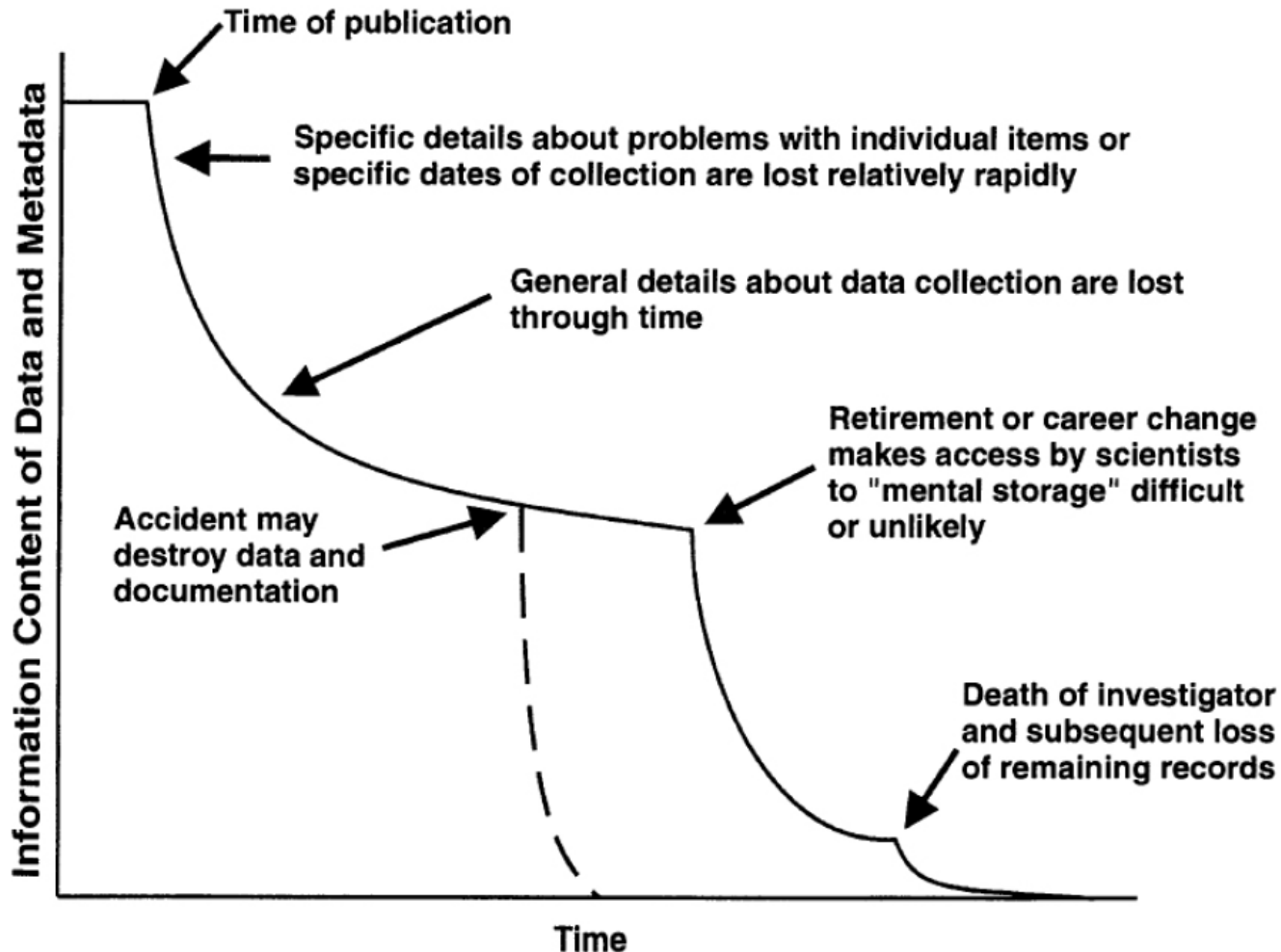
Jeffery S. Horsburgh, Anthony Castronova,
and the rest of the CI Team

iUTAH All Hands Meeting
11-21-2014

Overview

- iUTAH Data Policy Review and Update
(available at <http://data.iutahepscor.org> or <http://iutahepscor.org>)
- Data Publication Process
 - Data Collection Plans
 - Metadata Submission
 - Data Submission
- Data Publication System
Demonstration of publication and search at <http://repository.iutahepscor.org>
- Cyberinfrastructure for Social Science Data

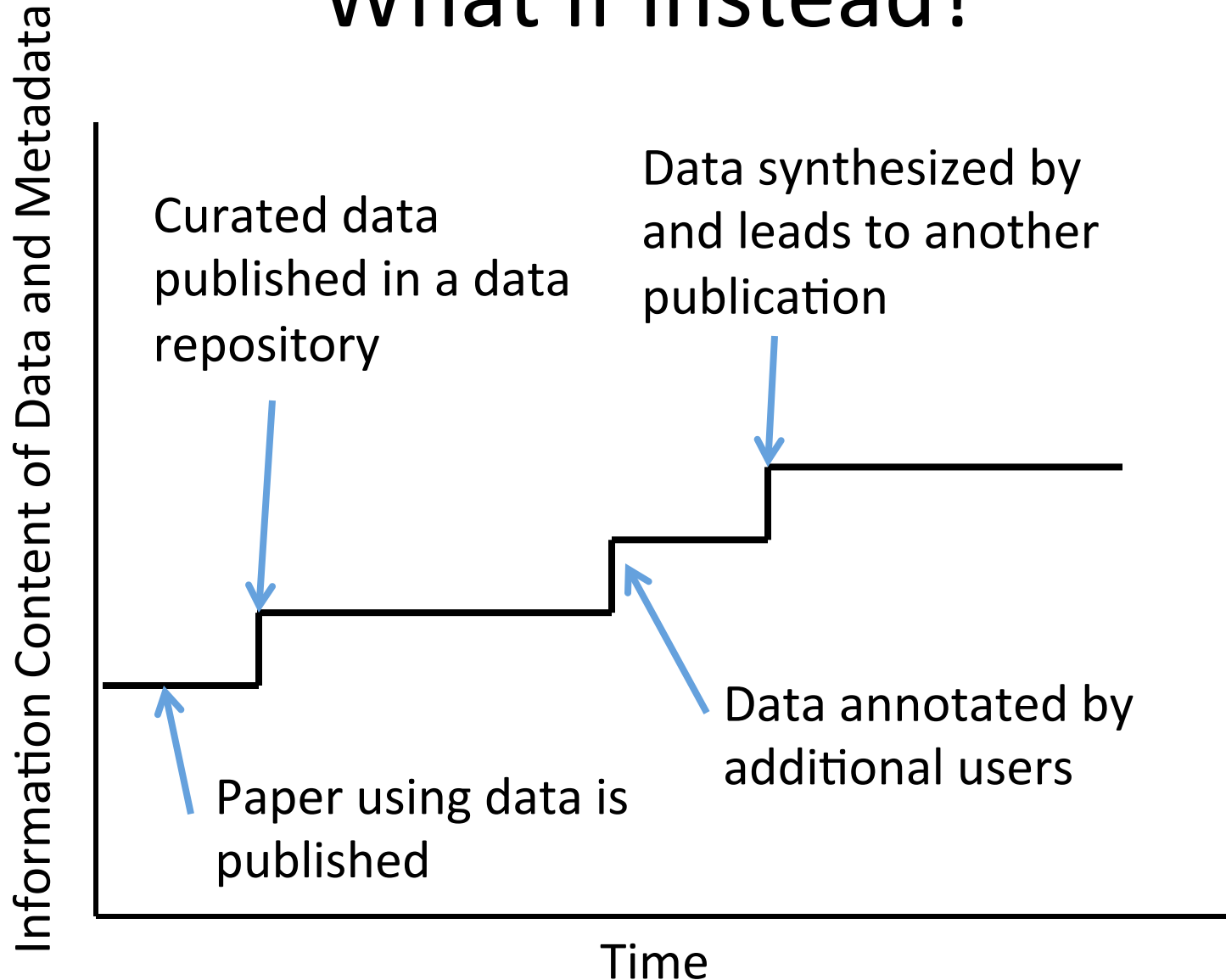
Information Entropy



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>.

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/data_modeling/data_management_Policy.html
- 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



Research Data Policy

Version 1.3

June 11, 2014

Edited by:

Jeffery S. Horsburgh and Amber S. Jones

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.

iUTAH Data Policy: UPDATE

Clarifies expectations for human subjects data:

- May require **redaction or aggregation** to prevent identification.
- If not possible, an **objective summary** of the data should be developed and released.
- Data products may include **survey instruments, summaries, or reports**.
- **IRB approvals** should be submitted (not full application).



Research Data Policy

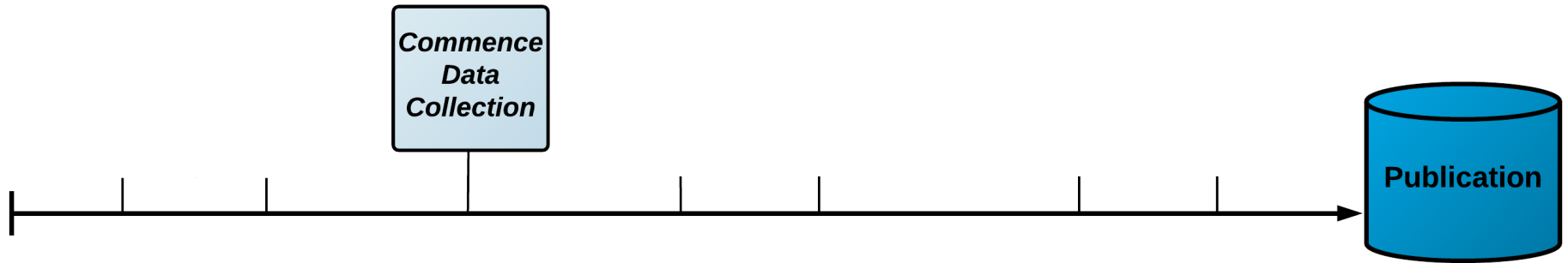
Version 1.3

June 11, 2014

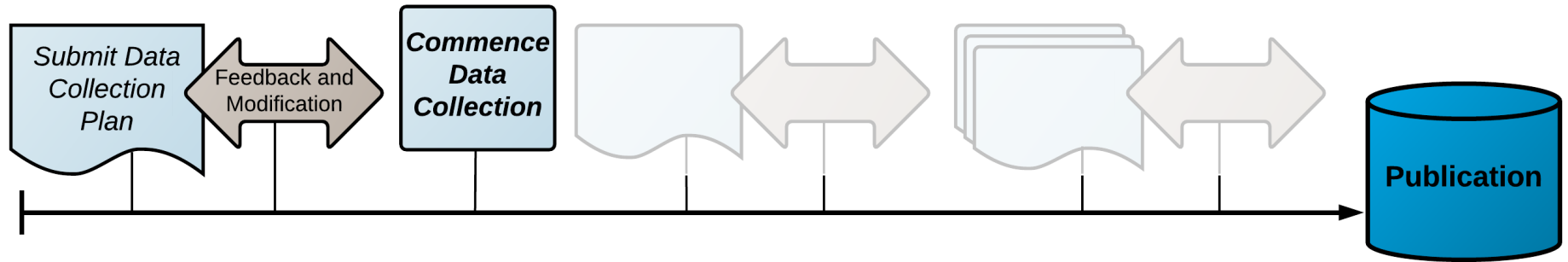
Edited by:

Jeffery S. Horsburgh and Amber S. Jones

Data Publication Process



Data Publication Process



iUTAH Data Policy: Data Collection Plans

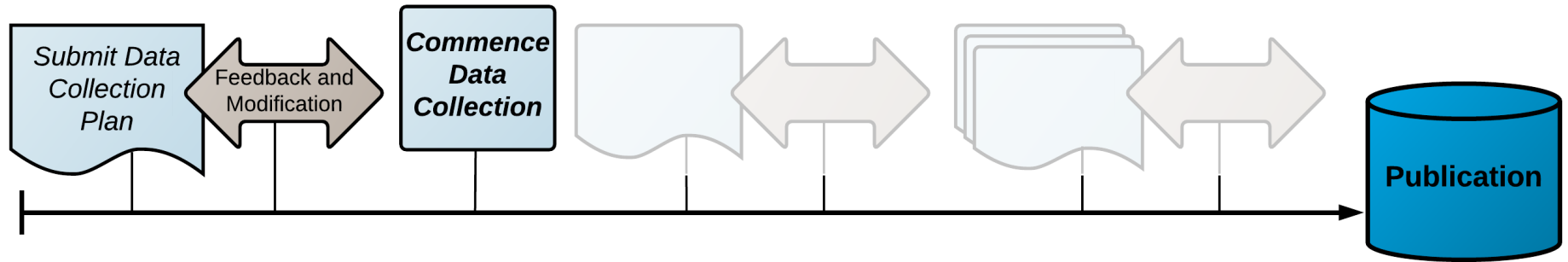
- **ALL** data creation efforts with **ANY** funding from iUTAH (salary, travel, sampling, equipment, etc.) **MUST** submit a brief plan to the Data Policy Committee **PRIOR** to funding.
- Plan should include:
 1. Identification of types of data to be collected/created
 2. Brief description of methods, 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

iUTAH Data Policy: Data Collection Plans

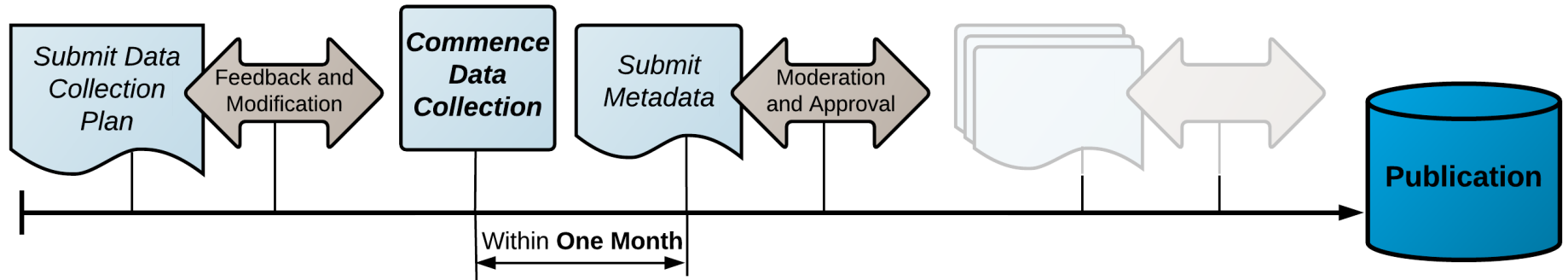
- Template should be used:
http://iutahepscor.org/data_modeling/data_management_Policy.shtml
- Submit to amber.jones@usu.edu

[illegible]

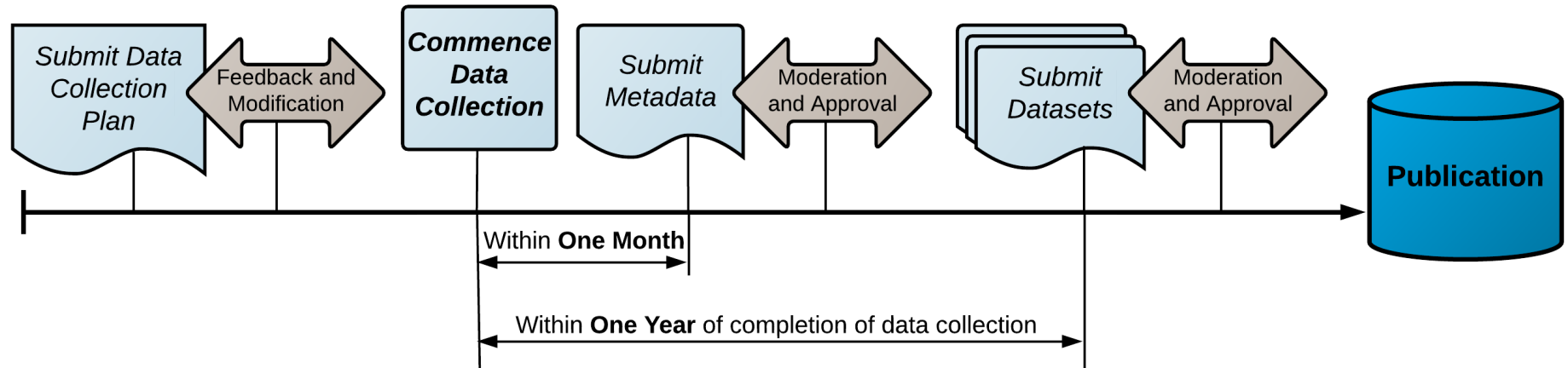
Data Publication Process



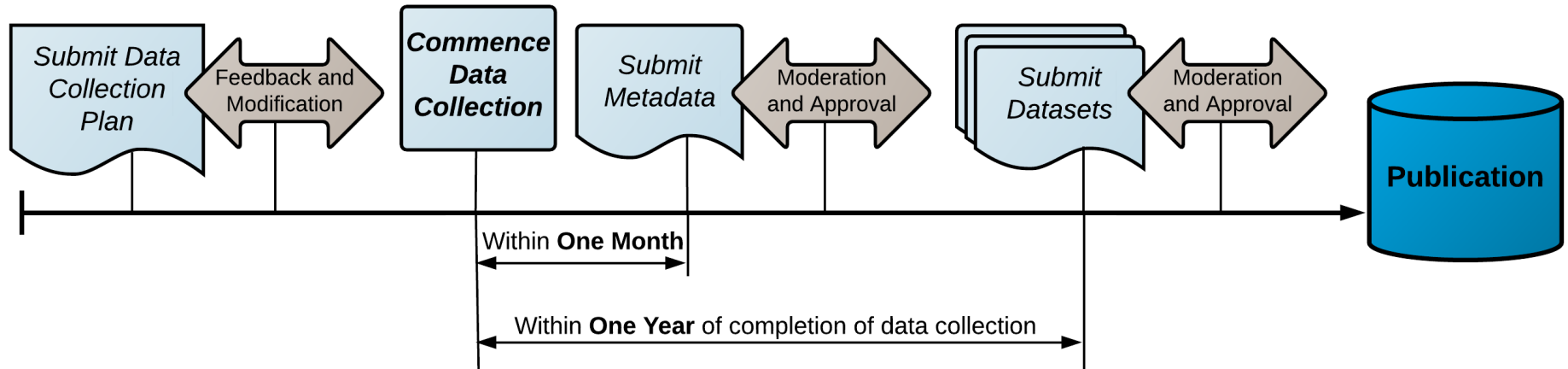
Data Publication Process



Data Publication Process



Data Publication Process



Publication:

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

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.

The screenshot shows the 'Create Dataset' form in the iUTAH Modeling and Data Federation system. The form is divided into several sections:

- What are datasets?**: A sidebar note explaining that datasets are used to group related pieces of data and can be found under a single URL with a description and licensing information.
- Create dataset**: The main form area, which includes:
 - Title**: A text field with the example 'eg. Red Butte Creek GAMUT Water Temperature Data.' and a URL field with the example 'URL: iutah-ckan-stage.uwrl.usu.edu/dataset/<dataset>' and an 'Edit' button.
 - Description**: A text area with the example 'eg. A short description (or abstract) for the dataset.' and a note 'You can use Markdown formatting here'.
 - Keywords**: A text field with the example 'eg. water quality, temperature, Red Butte Creek, time series'.
 - Organization**: A dropdown menu with 'iutah' selected.
 - Visibility**: A dropdown menu with 'Private' selected, and a link 'Why is my dataset private?'.
 - Language**: A dropdown menu with 'e.g., en, es, fr' selected.
 - Access Information**: A text area with the example 'e.g., limited to iUTAH participants, limited to IRB researchers' and a note 'You can use Markdown formatting here'.
 - Type**: A dropdown menu with 'collection' selected.
- Optional Metadata**: A section with fields for:
 - Purpose**: A text field with the example 'e.g., Educational, Research, Regulatory'.
 - Required Software**: A text field with the example 'e.g., ArcGIS, R, specific model application'.
 - Research Focus Area**: A dropdown menu.
- Spatial Metadata**: A section with fields for:
 - Spatial Coverage**: A text field with the example 'e.g., Salt Lake County'.
 - iUtah Study Area**: A dropdown menu.

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.

The screenshot shows the 'Create Dataset' form in the iUTAH Modeling and Data Federation system. The form is divided into several sections:

- What are datasets?**: A sidebar note explaining that datasets are used to group related pieces of data and can be found under a single URL with a description and licensing information.
- Create dataset**: The main form area, which includes:
 - Title**: A text field with the example 'eg. Red Butte Creek GAMUT Water Temperature Data'.
 - URL**: A text field with the example 'iutah-ckan-stage.uwrl.usu.edu/dataset/<dataset>' and an 'Edit' button.
 - Description**: A text area with the example 'eg. A short description (or abstract) for the dataset.' and a note 'You can use Markdown formatting here'.
 - Keywords**: A text field with the example 'eg. water quality, temperature, Red Butte Creek, time series'.
 - Organization**: A dropdown menu with 'iutah' selected.
 - Visibility**: A dropdown menu with 'Private' selected, and a link 'Why is my dataset private?'.
 - Language**: A text field with the example 'e.g., en, es, fr'.
 - Access Information**: A text area with the example 'e.g., limited to iUTAH participants, limited to IRB researchers' and a note 'You can use Markdown formatting here'.
 - Type**: A dropdown menu with 'collection' selected.
- Optional Metadata**: A section with fields for:
 - Purpose**: A text field with the example 'e.g., Educational, Research, Regulatory'.
 - Required Software**: A text field with the example 'e.g., ArcGIS, R, specific model application'.
 - Research Focus Area**: A dropdown menu.
- Spatial Metadata**: A section with fields for:
 - Spatial Coverage**: A text field with the example 'e.g., Salt Lake County'.
 - iUTAH Study Area**: A dropdown menu.

Survey Data Viewer

Water
Nov. 5, 2014

Thinking about the next 10 years / in this valley, how concerned are you about each of the / followin...

Respondant Percentage

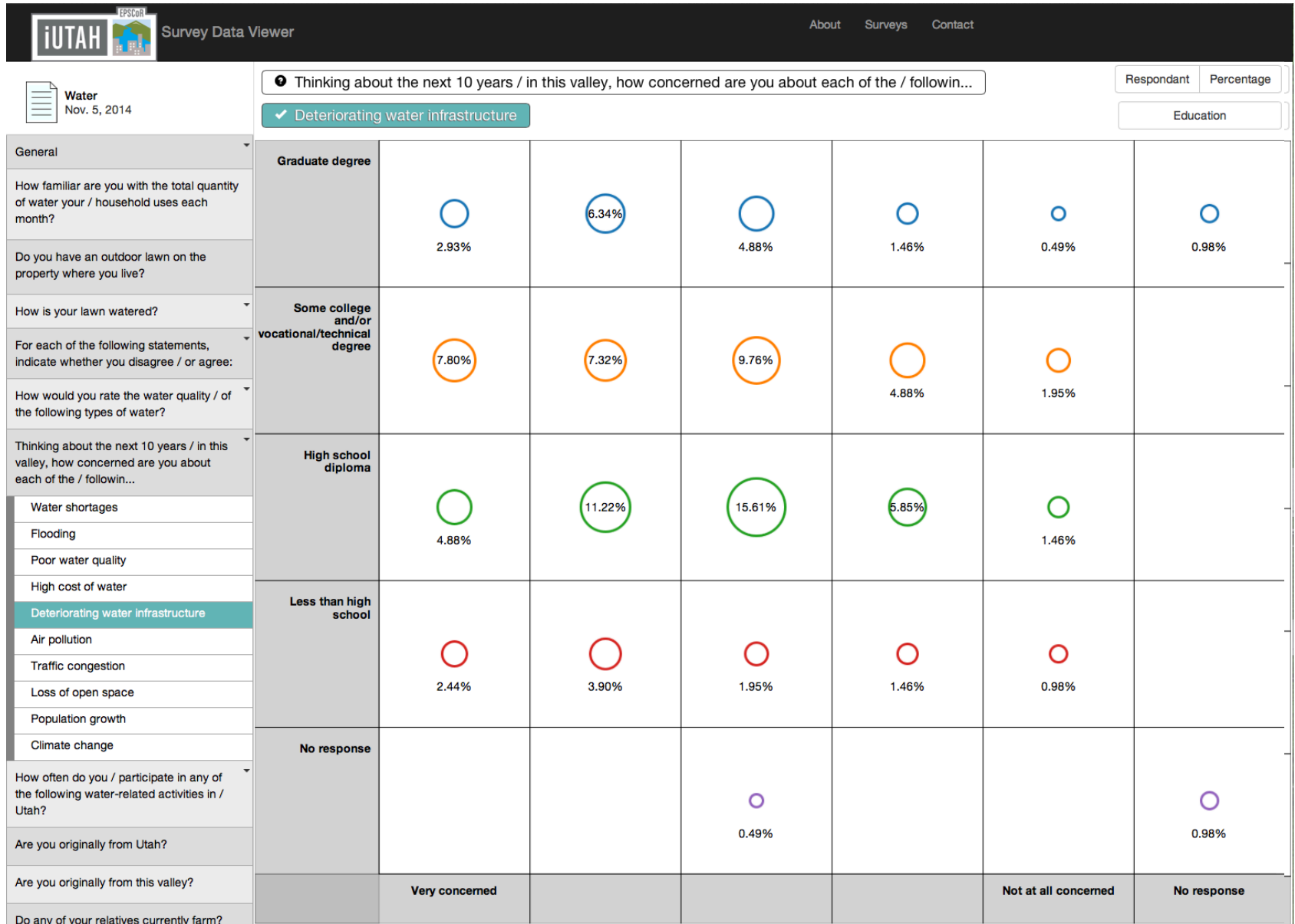
Water shortages

Age

General
How familiar are you with the total quantity of water your / household uses each month?
Do you have an outdoor lawn on the property where you live?
How is your lawn watered?
For each of the following statements, indicate whether you disagree / or agree:
How would you rate the water quality / of the following types of water?
Thinking about the next 10 years / in this valley, how concerned are you about each of the / followin...
Water shortages
Flooding
Poor water quality
High cost of water
Deteriorating water infrastructure
Air pollution
Traffic congestion
Loss of open space
Population growth
Climate change
How often do you / participate in any of the following water-related activities in / Utah?
Are you originally from Utah?
Are you originally from this valley?

60 and over						
	8.78%	7.80%	3.41%	0.49%	2.44%	0.98%
50-59	5.37%	6.34%	2.93%	1.46%	0.49%	
40-49	3.90%	4.39%	5.37%	2.44%	1.46%	
30-39	5.85%	6.34%	3.90%	0.98%	0.98%	
18-29	4.39%	9.76%	4.88%	1.95%	1.46%	
No response			0.49%			0.98%
	Very concerned				Not at all concerned	No response

Survey Data Viewer





Questions?

Jeff Horsburgh

jeff.horsburgh@usu.edu

Amber Jones

amber.jones@usu.edu