Cyberinfrastructure Update

Amber Spackman Jones

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

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

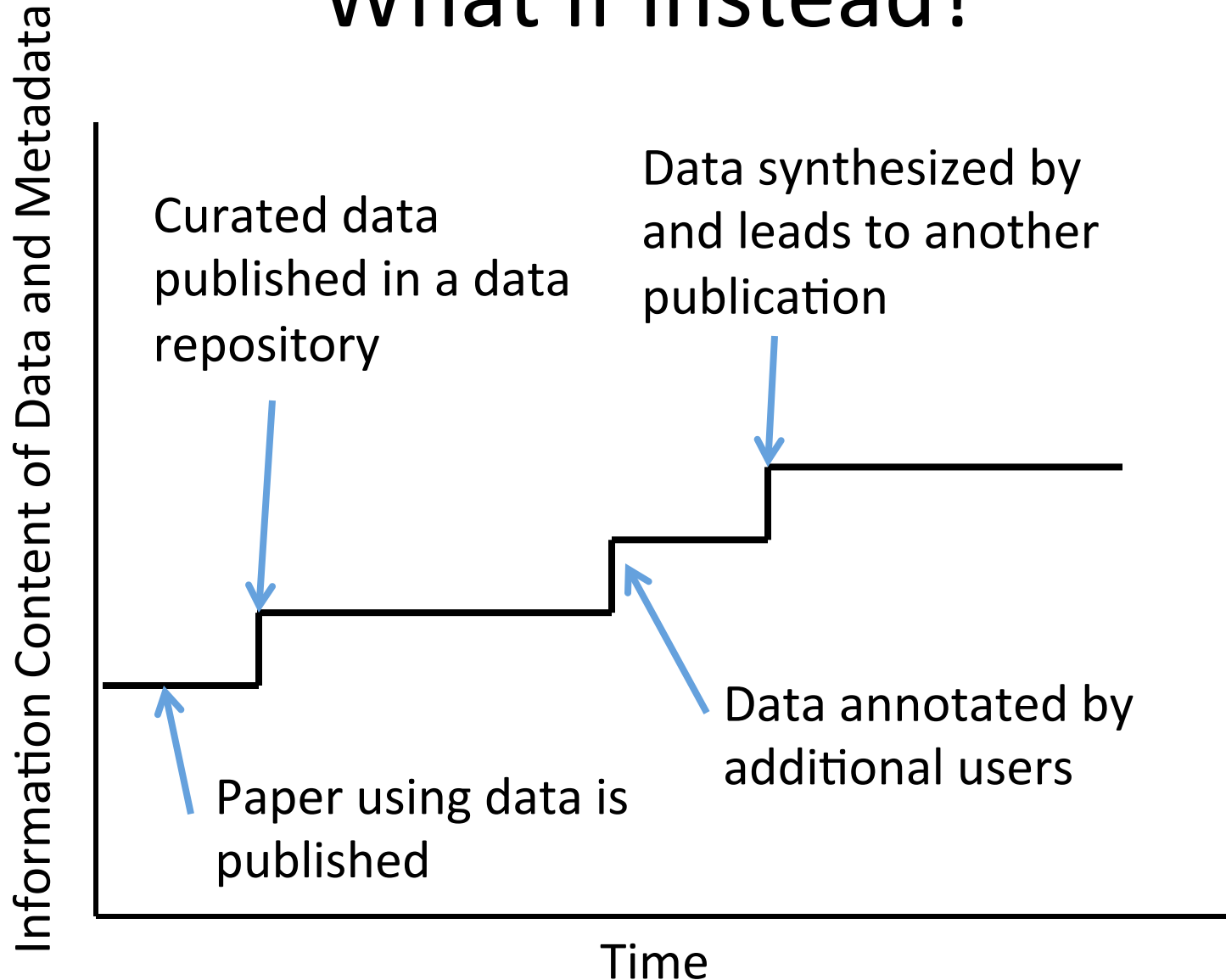- Cyberinfrastructure for Social Science Data

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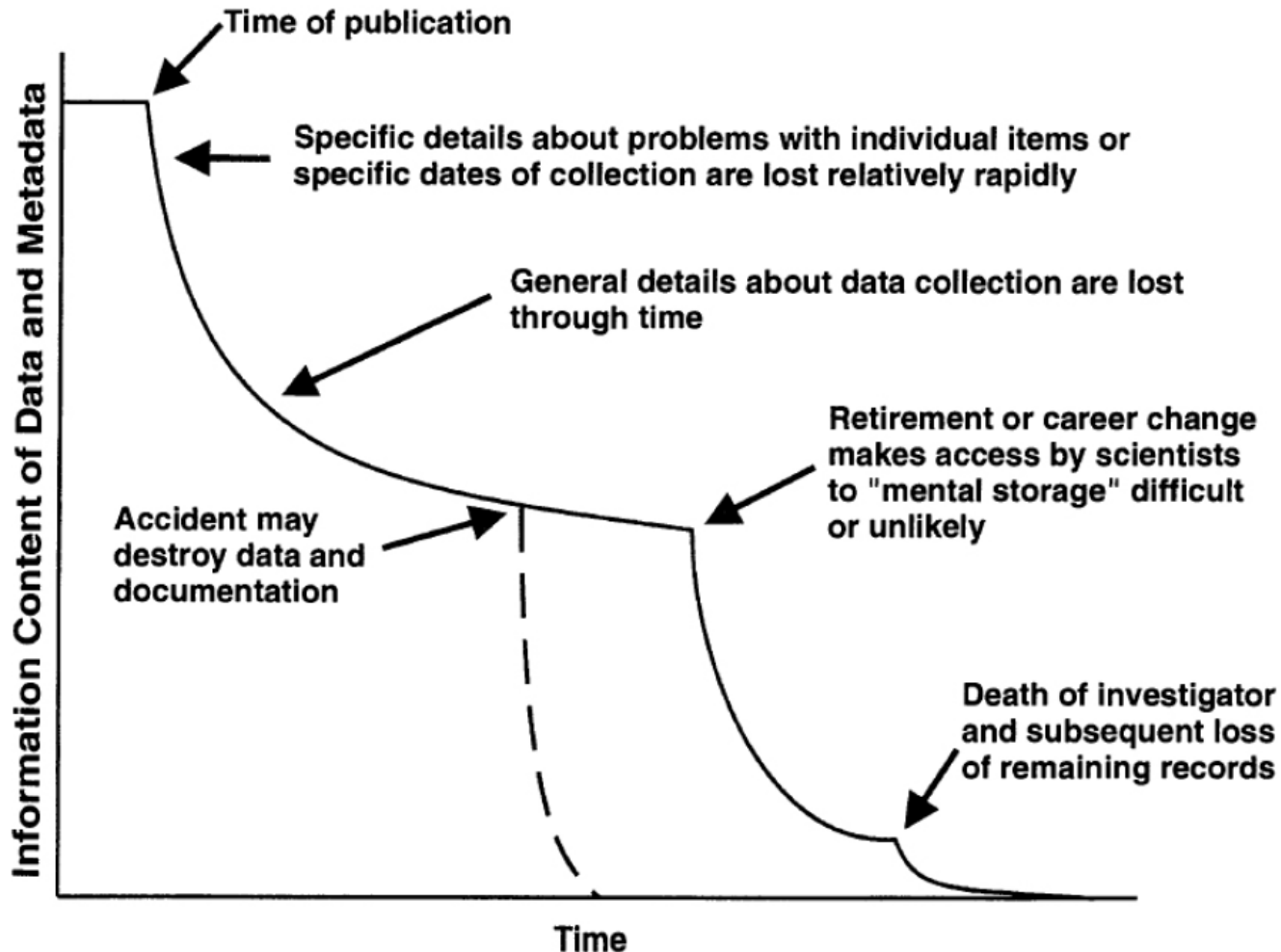
Example of the normal degradation in information content associated with data and metadata over time ("information entropy").

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?

Information Content of Data and Metadata (vertical axis label)

Time (horizontal axis label)

Curated data published in a data repository

Data synthesized by and leads to another publication

Data annotated by additional users

Paper using data is published

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

EPSCoR

iUTAH

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
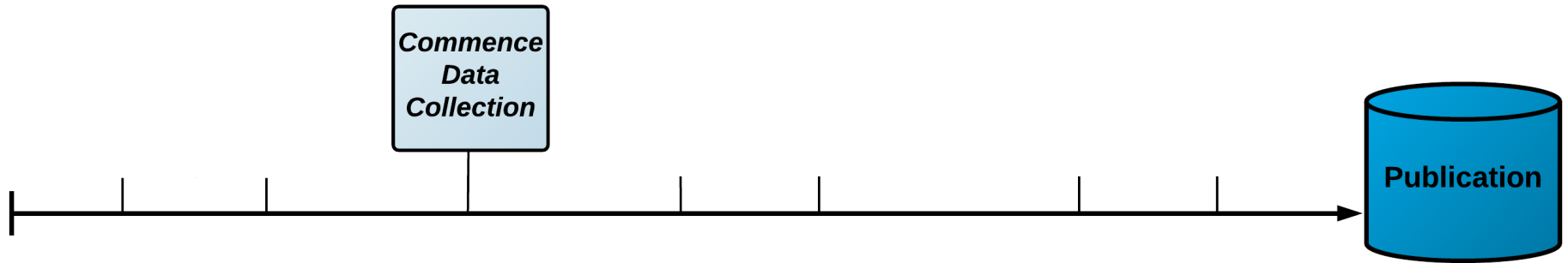
Data Publication Process
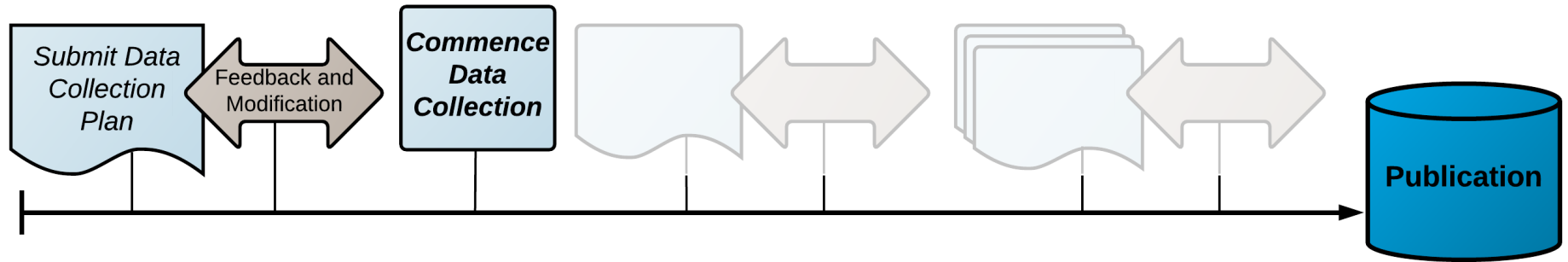
iUTAH Data Policy: Data Collection Plans

- **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

Data Product	Data Type*	Method of Creation	Resulting Data Format	Data Storage	Final Data Product	Timeframe**	Access During Collection	Access After Completion	Anonymization for IRB

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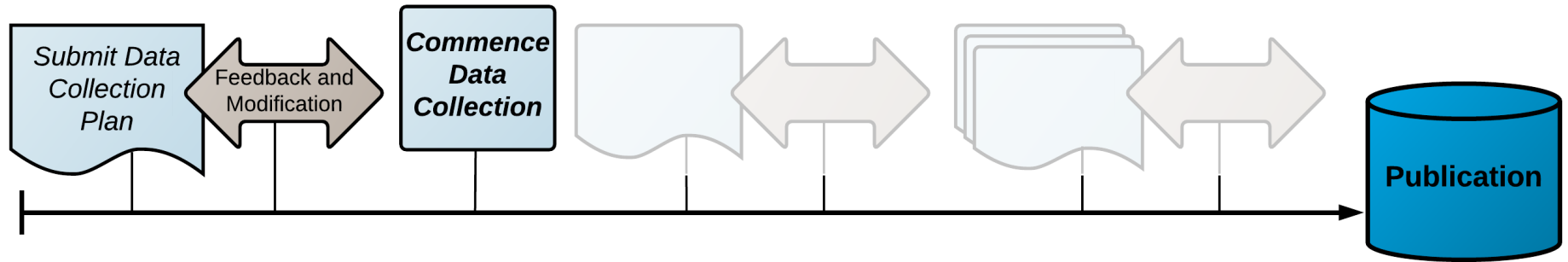
Data Publication Process
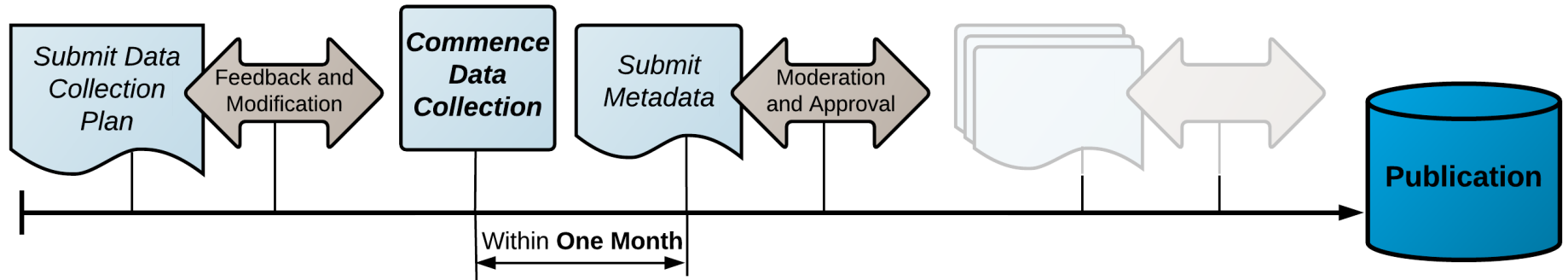
Data Publication Process
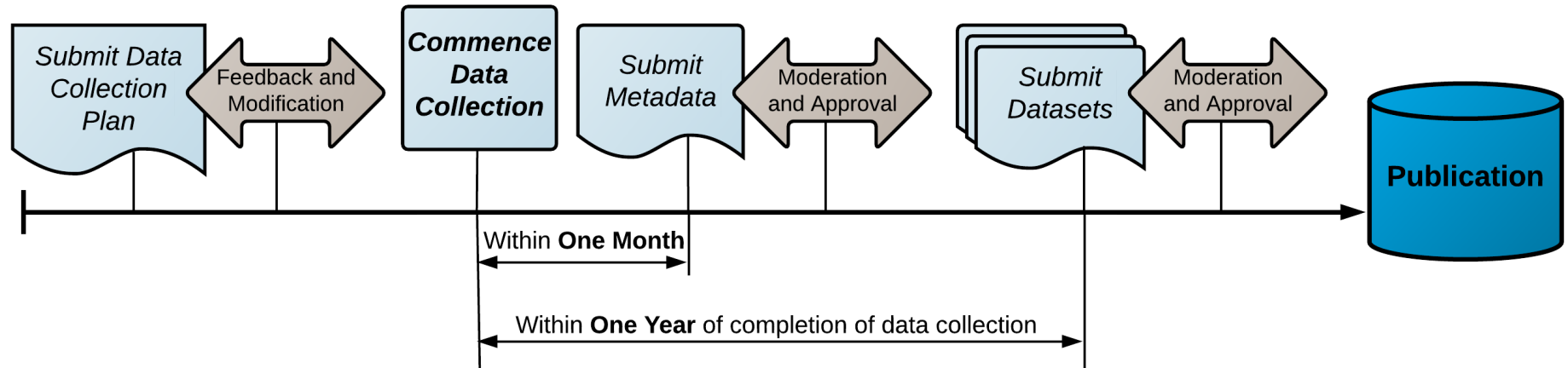
Data Publication Process
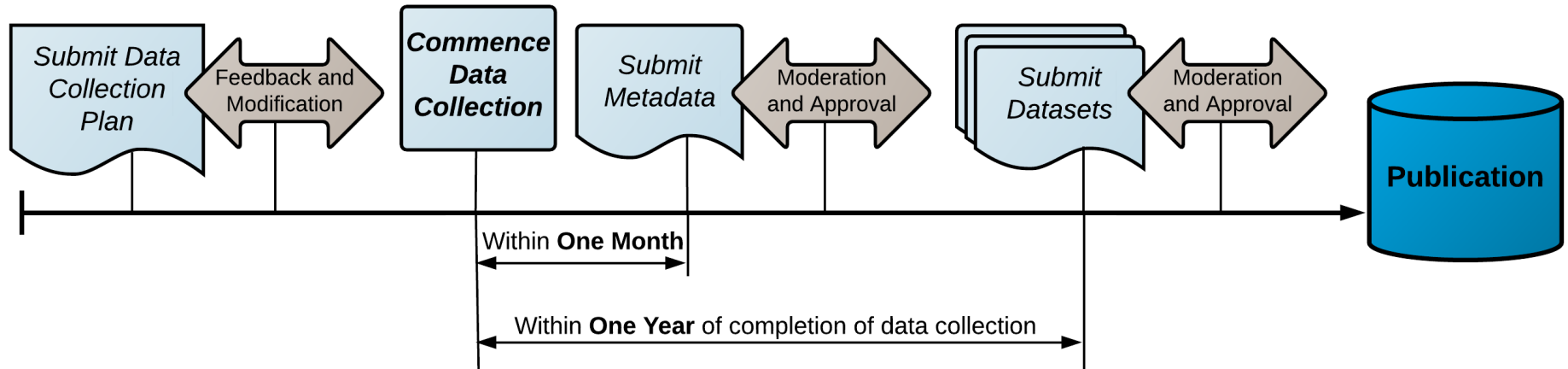
Publication:
- Assigns a citation and a url
- Data and metadata are discoverable
- Datasets are archived and curated

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.

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.

Survey Data Viewer

Survey Data Viewer

Questions?

Jeff Horsburgh
jeff.horsburgh@usu.edu

Amber Jones
amber.jones@usu.edu