

Stings from The Hive: Challenges of Launching the University of Utah Data Repository

Daureen Neddill, Rebekah Cummings and Harish Maringanti



Challenge #1 – Will It Fit?

Background: Sofia (from the Samvera Framework) was selected as the platform for the data repository.

Case 1: The platform (Sofia) includes metadata terms that are useful in describing cultural heritage content (digital collections) and institutional repository material. Those labels don't work well for describing research data sets – e.g. use of the term "language".

Result: The Tech Team attended the Data Team ingestion meetings with each researcher. This allowed them to see firsthand any issues with terminology.

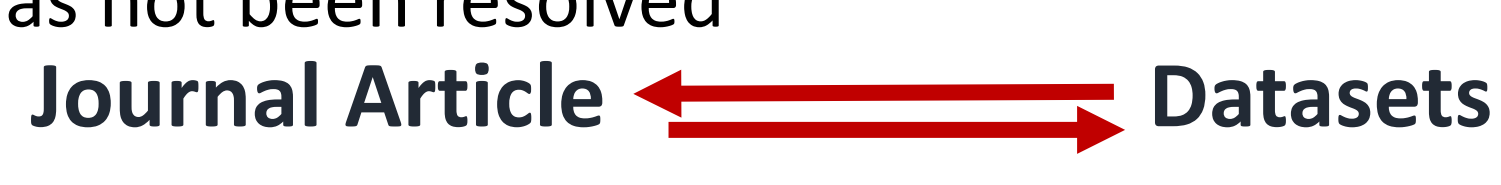


The Tech Team
Bryson Benton, Leah Donaldson, Amanda Crittenden, Brian McBride and Jacob Reed

Challenge #2 – Can We Connect?

Background: Researchers submitted datasets to obtain the DOI required by publishers. The DOI and citation for the article would be available upon publication. This information needs to be submitted to The Hive so that the article and the dataset are linked. How do we obtain this information?

Result: Has not been resolved



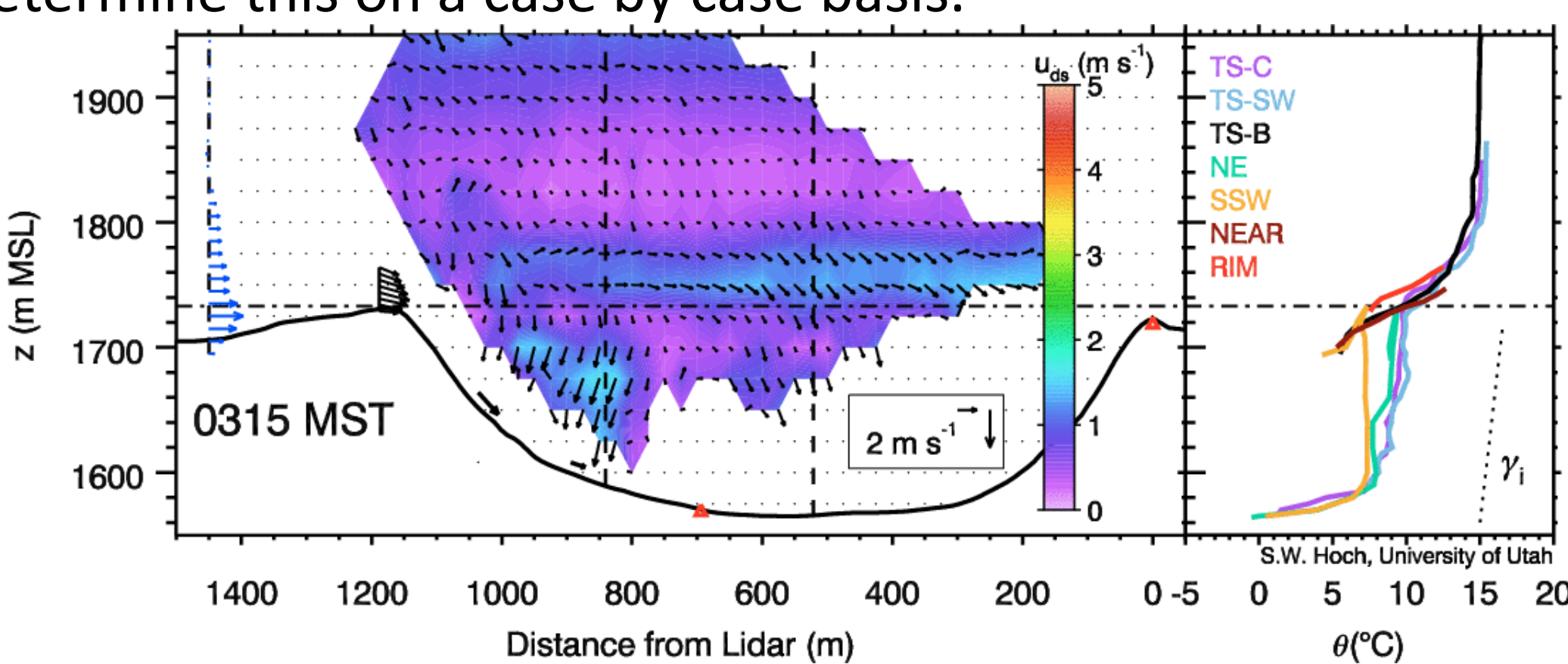
Challenge #3 - Which Repository?

Background: USpace is the University's Institutional Repository launched in 2004 for submission of preprints, posters, thesis & dissertations, and institutional records. USpace administrators allowed users to deposit small datasets. With the launching of The Hive, it was expected that researchers would deposit datasets into The Hive and gray literature into USpace.

Case 1: A researcher was receiving multiple requests for charts developed on slides from openly available data. He wanted the slides distributed via The Hive.

Case 2: A researcher wanted an interactive image submitted to The Hive with the data residing in a federal repository. It was supplemental material for a published article.

Result: Should processed data in the form of charts, images graphs, etc. be submitted to The Hive or to USpace? Case 1 the slides were not submitted and Case 2 the image was submitted. Presently, we will still determine this on a case by case basis.



Challenge #4 - No Data, Just Metadata?

Background: Two researchers with large datasets (<TB) were not looking for data storage, but rather metadata records with PIDs linking to datasets wherever they resided.

Case 1: A researcher with a project funded as a line item in the US DOT budget requested that his database be included in The Hive via a PID and metadata to (1) increase visibility and (2) because it was university research.

Case 2: A researcher with the project dataset hosted on servers at CHPC wanted a PID for a subset of the dataset for a publication. The metadata would include links to the entire dataset.

Result: The pilot project was expanded to include these types of projects. Discoverability would increase but there is less control over the persistence of links and long-term data availability.

Challenge #5 – Do You Know Your Rights?



Background: Because The Hive is a data repository specifically for the broad dissemination and preservation of the university's intellectual contributions, it was decided that the Creative Commons licenses of CC BY, CC BY NC, CCO, and Public Domain would be offered to our dataset submitters.

Case 1: Researchers submitting software asked about using other licenses, e.g. Apache License, GNU GPL, etc.

Result: An additional option was added to accommodate non-Creative Commons licenses, "Other – See README file for more information."

