

## [Metadata for Discoverability, Accessibility, Useability, and Understanding](#) [1]



Submitted by Ted.Habermann on Mon, 2014-10-13 10:06 Tuesday, January 6, 2015 - 13:30 to 15:00

**Event:** [Winter Meeting 2015](#) [2]

**Session Type:** [Breakout](#) [3]

**Expertise Level:** [Intermediate](#) [4]

**Collaboration Area:** [Documentation](#) [5]

**Abstract/Agenda:**

The connection between metadata and data discovery has been strong since the early days of the Web. At the same time, metadata standards have included concepts that go significantly beyond discovery into the realms of accessibility, useability, and understanding. These elements generally do not make it onto lists of minimum metadata requirements and typically are only sparsely populated in many metadata collections. This means data that are discovered may not be accessible, useable, or understandable. As part of the Big Earth Data Initiative (BEDI) NASA is developing guidelines for documenting access, use, and understanding. This work is shared in the [Documentation Connections](#) [6] section of the ESIP Wiki and will be presented during this session. The next step is developing tools and processes that will facilitate effective improvement of metadata for accessibility, use, and understanding.

**Notes:**

- [http://wiki.esipfed.org/index.php/Category:Documentation\\_Connections](http://wiki.esipfed.org/index.php/Category:Documentation_Connections) [7]

Speaker 1: Curt Tilmes

(No presentation available)

- Metadata quality study: determine the characteristics of the data and contribution
- Who has done what to which dataset.
- Drive improvement

Speaker 2: John Kozimor of HDF

Presentation Slides Title: Big Earth Data Initiative Metadata Quality Study (derived from 2 white papers).

- The information is also available on ESIP wiki:

[http://wiki.esipfed.org/index.php/Category:Documentation\\_Connections](http://wiki.esipfed.org/index.php/Category:Documentation_Connections) [6]

- Many different organizations have different approaches for defining and implementing metadata.
- HDF constructed a method to compare and contrast metadata recommendations → ensemble approach (please note that the approach is still under development; ESIP community is welcome to provide feedback regarding the approach).
  - This approach looks for intersections between different metadata dialects.
  - This approach also aims to provide a quantitative way to represent the comparison among the different metadata, so that the similarities and the differences could be better understood. In return, there is also a quantifiable way to express the areas that are well understood or not well understood, so that the recommendations can be fine tuned.
  - Fundamentally, the approach would like to help answering the “why” in why use metadata and improving the recommendations of metadata, so that the diversity of users’ needs can be better met via the usage and implementation of metadata.
  - The ensemble approach method includes the following components:
    - Spiral = collection of related concepts or recommendations.
      - Spiral is meant to be efficient and “do-able”.
    - Concepts = metadata elements; ex: Catalog Services for the web, DIF.
      - Concepts could be grouped by organizations or use scenarios → these are the main drivers for grouping individual concepts together.
      - Names for the concept should be unique and be able to work across different metadata “dialects”.
- In the study cases presented, Data Discovery, Accessibility/Usability, and Understanding are the 3 areas of recommendations studied.
  - Result 1: There are a lot of capabilities of metadata that are not being utilized.
  - Result 2: The current recommendations are being summarized under the following page: [http://wiki.esipfed.org/index.php/Category:Documentation\\_Recommendations](http://wiki.esipfed.org/index.php/Category:Documentation_Recommendations) [8]

- Result 3: An example reviewed is “Data Discovery (FGDC)”:  
[http://wiki.esipfed.org/index.php/Data\\_Discovery\\_\(FGDC\)](http://wiki.esipfed.org/index.php/Data_Discovery_(FGDC)) [9]
- Result 4: Another example reviewed is “Service Discovery (SERF)”:  
[http://wiki.esipfed.org/index.php/Service\\_Discovery\\_\(SERF\)](http://wiki.esipfed.org/index.php/Service_Discovery_(SERF)) [10]
- Result 5: A third example reviewed is “Data Understanding - External References (ISO-19115-1)”:  
[http://wiki.esipfed.org/index.php/Data\\_Understanding\\_-\\_External\\_References\\_\(ISO-19115-1\)](http://wiki.esipfed.org/index.php/Data_Understanding_-_External_References_(ISO-19115-1)) [11]
- Different users will want to create/examine guidance (i.e. selection scenario) based on different sets of recommendations.
- Documentation Selection Scenarios on ESIP wiki can be found under the following link:  
[http://wiki.esipfed.org/index.php/Category:Documentation\\_Selection\\_Scenarios](http://wiki.esipfed.org/index.php/Category:Documentation_Selection_Scenarios) [12]
  - There are two scenarios presented under this wiki page.
- Quality of metadata is related to how well the users can use the selected metadata to discover, access, and understand the final information object.
  - Completeness and consistency are also important to help upholding the quality of metadata.
- Metadata Solution Concepts Study includes the following components. These comparisons help demonstrating which concept is covered well in which dialect and by which recommendation.
  - Recommendation comparison
  - Dialect comparison
  - Dialect/Recommendation comparison
- A suggestion was made from the audience to include non-mandatory concepts as well.
- What if there is a concept that one needs, but the concept is not available in the dialect that one is using?
  - This might be an issue if the dialect is not extensible.
- A suggestion from the audience is to receive which dialect covers most recommended concepts:

- ISO19115 was provided as the answer by Ted.
- Tools will be developed by the Documentation Cluster to help with analyzing and visualizing the ensemble approach.
- A suggestion from the audience is to compare data portals because this is the primary mechanism for users to engage/interact with the metadata.
  - Ted follows up by reiterating there are definitely still room for improvement, including better utilizing the capabilities of different metadata dialect and tracking provenance.
- Metadata rubrics as evaluation tool → has the following 3 components:
  - Situation: Metadata collection
  - “Tool”: Selection Scenario and Dialect
  - Result: Spiral scores for each record
- Metadata rubrics allow collection analysis.
- Example discussed include 8 different spiral (as x-axis) and scores of the spiral (as y-axis) built using 52 records and divided into 2 groups. The “global average” or the baseline for each concept is calculated based on previously selected, different records.
  - The result shows that there are 14 “bright spots,” or areas of metadata that are done well, in the records and 38 opportunities for improvement.
  - Overall, this approach identifies:
    - Specific actions to improve metadata
    - Examples the demonstrate the benefits
    - Quantitative metrics for measuring improvement
- Description Spiral provides additional information that may be searched in some text searches. It includes the following: Project Keyword, Project Keyword Thesaurus, Dataset Extent Description, Purpose, and Lineage Statement (these fields are texts, so are still searchable. They also augment information provided in the “standard” fields, such as title and abstract).

- One benefit observed is that the Project Keyword covers more than just GCMD keywords (thought to be the default). As a result, by including Project Keyword field, the “richness” of the search field could be enhanced. → this would be one “bright spot” or area that could be improved.
- On ESIP Documentation Cluster wiki, Guidance pages are provided, including dialect descriptions and crosswalks.
- UMM = Unified Metadata Mapping
  - UMM Collection: [http://wiki.esipfed.org/index.php/UMM\\_Collection](http://wiki.esipfed.org/index.php/UMM_Collection) [13]
  - UMM Granule: [http://wiki.esipfed.org/index.php/UMM\\_Granule](http://wiki.esipfed.org/index.php/UMM_Granule) [14]

### Discussions:

- A question from the audience: Can the quality of metadata be compromised after several transforms?
  - We should be able to quantitatively show which metadata elements match between the dialects and which ones do not.
  - Semantically, there should not be too much degradation.
  - “Wikis, Rubrics, Views, and Connections: An Integrated Approach to Improving Documentation” ([http://www.slideshare.net/tedhabermann/wikis-rubrics-and-views?qid=5a022745-1a78-461d-b047-04ec5cfbf8ec&v=qf1&b=&from\\_search=1](http://www.slideshare.net/tedhabermann/wikis-rubrics-and-views?qid=5a022745-1a78-461d-b047-04ec5cfbf8ec&v=qf1&b=&from_search=1) [15])
- A question from the audience: Crosswalking additional attributes might also be challenging, and therefore, at risk of data loss?
  - Non-standard metadata, such as archive metadata, is more likely to be at danger to loss data than the attributes that are currently being used.
- Accountability through metadata is going to be significant for science advancement.
- A question from the audience: Metrics - is ensemble approach looking at the actual metadata value?
  - This is a significant question because currently, a detailed description and a statement that said “No abstract available” will both get the same score (this is the

completeness issue relating to the quality of metadata). Similarly, different spellings and misspellings are also examples of inconsistency issues that impact quality of metadata.

- A suggestion from the audience: Do we need control vocabulary for most of the metadata elements?
  - Some fields will work well with control vocabulary, but others won't. However, this is also why it is important to analyze metadata, so that the areas of improvement can be identified.
  - Different levels of details can also help enrich the metadata content.
- A question from the audience: Is it possible to achieve the perfect metadata?
  - Yes, when every field of a selected, implemented metadata has meaningful and understandable content for the described object.
- A questions from the audience: How can the ECHO metadata be made available to the data owners, so that they have the opportunity to improve the metadata?
  - The feedback will be provided back to the data owners.

## Session Leads:

**Name:** [Ted Habermann](#) [16]  
**Organization(s):** [The HDF Group](#) [17]  
**Email:** [thabermann@hdfgroup.org](mailto:thabermann@hdfgroup.org) [18]

## Presenters:

**Name:** [John Kozimor](#) [19]  
**Organization(s):** [The HDF Group](#) [17]  
**Email:** [jkozimor@hdfgroup.org](mailto:jkozimor@hdfgroup.org) [20]

**Name:** [Ted Habermann](#) [16]  
**Organization(s):** [The HDF Group](#) [17]  
**Email:** [thabermann@hdfgroup.org](mailto:thabermann@hdfgroup.org) [18]

**Name:** [Curt Tilmes](#) [21]  
**Organization(s):** [OzoneAO](#) [22]  
**Email:** [curt.tilmes@nasa.gov](mailto:curt.tilmes@nasa.gov) [23]

## Notes takers:

**Name:** [Sophie Hou](#) [24]  
**Organization(s):** [UCAR/NCAR](#) [25]

## Participants:

Sarah O'Connor, Jason Gallo, Emily Nightingale, Labit Wanchoo, Stephen Berrick, David Gallaher, Jonathan Blythe, Ed Armstong, Stephan Klene, Richard Eckerman, Walt Baskin, Pam Mlynezak, Lily Zolly, Tyler Stevens, Lisa Booker, Jennifer Wei, Ajay Krishnan, Melissa Zweng, Nate James, Brian

## Metadata for Discoverability, Accessibility, Useability, and Understanding

Published on Commons (<http://commons.esipfed.org>)

---

Johnson, Chris Crosby, William Doane, Kaijian Xu, Ellen Johnson, Sara Graves, Steve Diggs, Don Stott, Kelly Monteleone, Sophie Hou, John Kozimor, Curt Tilmes, Lindsay Powers, Beth Huffin, Steve Richards, Tiffany Mathews, Kaylin Bugbee, Rahul Ramachandra, Casandra Ladino, Steve Kepler, Yonsook Enloe, Anna Milan, Yuangie Li, Andrew Mitchell, Jeff Chen, Brent Newman, Bruce V., Byron Peters, Jacquelin (Jaci) Mize, Jim Duncan, Justin Goldstein, Katie Baynes, Peng, Robert Wolfe, Steve Aulenbach, Heather Brown, Feng Ding, Chung-Lin Shie, Dan Keyes, Archana Subramanian

**Creative Common License:** Creative Commons Attribution 3.0 License

**Teaser:** How can we improve accessibility, Usability, and understanding of environmental data

**Accepted:**

**Keywords:** [documentation](#) [26]

**Source URL:** <http://commons.esipfed.org/node/2727>

### Links:

- [1] <http://commons.esipfed.org/node/2727>
- [2] <http://commons.esipfed.org/2015WinterMeeting>
- [3] <http://commons.esipfed.org/session-type/breakout>
- [4] <http://commons.esipfed.org/taxonomy/term/261>
- [5] <http://commons.esipfed.org/collaboration-area/documentation>
- [6] [http://wiki.esipfed.org/index.php/Category:Documentation\\_Connections](http://wiki.esipfed.org/index.php/Category:Documentation_Connections)
- [7] [http://wiki.esipfed.org/index.php/Category:Documentation\\_Connections](http://wiki.esipfed.org/index.php/Category:Documentation_Connections)
- [8] [http://wiki.esipfed.org/index.php/Category:Documentation\\_Recommendations](http://wiki.esipfed.org/index.php/Category:Documentation_Recommendations)
- [9] [http://wiki.esipfed.org/index.php/Data\\_Discovery\\_\(FGDC\)](http://wiki.esipfed.org/index.php/Data_Discovery_(FGDC))
- [10] [http://wiki.esipfed.org/index.php/Service\\_Discovery\\_\(SERF\)](http://wiki.esipfed.org/index.php/Service_Discovery_(SERF))
- [11] [http://wiki.esipfed.org/index.php/Data\\_Understanding\\_-\\_External\\_References\\_\(ISO-19115-1\)](http://wiki.esipfed.org/index.php/Data_Understanding_-_External_References_(ISO-19115-1))
- [12] [http://wiki.esipfed.org/index.php/Category:Documentation\\_Selection\\_Scenarios](http://wiki.esipfed.org/index.php/Category:Documentation_Selection_Scenarios)
- [13] [http://wiki.esipfed.org/index.php/UMM\\_Collection](http://wiki.esipfed.org/index.php/UMM_Collection)
- [14] [http://wiki.esipfed.org/index.php/UMM\\_Granule](http://wiki.esipfed.org/index.php/UMM_Granule)
- [15] [http://www.slideshare.net/tedhabermann/wikis-rubrics-and-views?qid=5a022745-1a78-461d-b047-04ec5cfbf8ec&v=qf1&b=&from\\_search=1](http://www.slideshare.net/tedhabermann/wikis-rubrics-and-views?qid=5a022745-1a78-461d-b047-04ec5cfbf8ec&v=qf1&b=&from_search=1)
- [16] <http://commons.esipfed.org/node/1847>
- [17] <http://commons.esipfed.org/taxonomy/term/822>
- [18] <mailto:thabermann@hdfgroup.org>
- [19] <http://commons.esipfed.org/node/7745>
- [20] <mailto:jkozimor@hdfgroup.org>
- [21] <http://commons.esipfed.org/node/1397>
- [22] <http://commons.esipfed.org/taxonomy/term/763>
- [23] <mailto:curt.tilmes@nasa.gov>
- [24] <http://commons.esipfed.org/node/7872>
- [25] <http://commons.esipfed.org/taxonomy/term/2486>
- [26] <http://commons.esipfed.org/taxonomy/term/924>