**Group #1:**

**Thematic Interest Areas**: Relevancy/Ranking & Metadata

**Evaluation Form**: <http://bit.ly/esip2016_uceval_downs>

**Use Case Title**: Dataset Rice Cooker Theory

**Use Case Author**: Bob Downs

**Use Case Narrative**: An integrated data product contains inputs from both ancillary and science data sources, of which the sources of science data come from multiple sensors/platforms/projects. The problem is we would want to make some assumptions about its quality if we were going to use it. This is a problem because it's a product that contains inputs from multiple datasets. This can pertain to all types of sources including satellite, in situ, airborne, model/reanalysis data etc... The other problem is how we communicate changes to input data to the end users. Examples of changes to input data include: statistical changes, quality flag implementations, data gaps... Such changes can affect the quality of the final product even though the final product algorithm doesn't change. Chung-Lin calls this the "rice cooker" theory. We also need to consider carefully examining uncertainties and other quality issues with similar integrated products that assimilate the exact same sources of input data but have a slightly different assimilation algorithm. Converging on three fundamental problems: two groups of producers of data quality information, input producers and integrated dataset producers; how is the product distributor taking these product quality inputs and conveying the information; data distributor would also have to document any changes to the data quality information resulting from an update to a finished product.

**Domain of Interest**: All of the above

**Professional Domain of User:** Data Management

**Primary User-Stakeholder Relationship:** Human User -> Data Producer/Stakeholder

**Secondary User-Stakeholder Relationship:** Human User -> Data Distributor/Stakeholder

**Primary Scope:** Quantitative-Science (for details see: <http://bit.ly/quantsci>)

**Primary Scope Rationale:** N/A

**Secondary Scope:** Qualitative-Product (for details see: <http://bit.ly/qualprod>)

**Secondary Scope Rationale:** N/A

**Use Case Chronology:**

1. User draws attention to an issue with the output data of an integrated dataset by contacting the data distributor.

2. Data distributor self examines the issue and attempts to resolve and diagnose. 3. Data distributor then contacts the integrated data producer with the summary of the issues discovered.

4. Integrated data producer investigates all of the input data and troubleshoots the issue to find the sources of the issue.

5. Integrated data producer then contacts the input data producer to inform them of issue and request more detailed quality information be provided for the input data.

6. This will all feedback to the end data user.

7. The data distributor with then document this information for all other users to discover.

**Success Criteria:**

1. All parties involved are responding to communication.

2. All parties involved have the requisite knowledge, skillsets, and funding to resolve these issues.

3. The user makes adjustments to account for this data quality issue.

4. Established process to document and track status.

5. The distributor should have all the information needed to update or create documentation to be made known to all other data users.

6. Fundamentally, enough distinctions can be made between the quality of various input datasets in relation to the integrated dataset.

**Data Quality Keywords**: algorithm, calibration, cross-calibration, data sampling, derivatives, documentation, filtering, flags, instrument sampling, interoperability, metadata, metrics, missing data, reporting, spatial resolution, standardization, temporal resolution, workflow

**Group #2:**

**Thematic Interest Areas**: Relevancy/Ranking & Metadata

**Evaluation Form**: <http://bit.ly/esip2016_uceval_peng>

**Use Case Title**: Appropriate Amount/Extent of Documentations for Data Use

**Use Case Author**: Ge Peng

**Use Case Narrative**: Traditionally, information about a data product can be captured and provided to end-users in the forms of collection metadata (e.g., via a doi landing page) and descriptive document such as A-TBD. Many end-users do not seem to think that they are easy to understand or are turned off by the overly extended elements or information. In addition, there is always a balance to strike between all possible descriptive information about the data product and available resources. The goal is to define levels of recommended documentations based on the levels of data use needs to guide end-users on what document to look for which will also help data centers on what documentations to prepare for.

**Domain of Interest**: All of the above

**Professional Domain of User:** Data Management, Stewardship, Service Support, Users

**Primary User-Stakeholder Relationship:** Human User -> Data Distributor/Stakeholder

**Secondary User-Stakeholder Relationship:** Human User -> Data Center/Stakeholder

**Primary Scope:** Quantitative-Science (for details see: <http://bit.ly/quantsci>)

**Primary Scope Rationale:** Scientific understanding -> improved applicability and usability

**Secondary Scope:** N/A

**Secondary Scope Rationale:** N/A

**Use Case Chronology:**

1. User defines data use need level - TBD; e.g., just browsing/high-level general information; basic information/characteristics of the product; initial use/understanding of the product; targeted application; extensive application of the product.

2. User selects the level of documentations based on a consistent framework.

3. User reads the appropriate documentations to get appropriate product information.

4. User uses the product properly and efficiently.

**Success Criteria:**

Defining Data Use Need and Documentation Maturity Levels - consistent framework so users know what types of documentations to turn to for their use needs.

**Data Quality Keywords**: documentation, interoperability, metadata, metrics, standardization, web services

**Group #3:**

**Thematic Interest Areas**: Relevancy/Ranking & Metadata

**Evaluation Form**: <http://bit.ly/esip2016_uceval_hou>

**Use Case Title**: Understanding and Identifying Datasets using SBC LTER Data Portal

**Use Case Author(s)**: Margaret O'Brien / Sophie Hou

**Use Case Narrative**: Through using the SBC LTER's "Browse/Search Data" capability (<http://sbc.lternet.edu//data/dataCollectionsPortal.html>), a scientist/researcher is able to identify the dataset(s) of interest based on the metadata/documentation provided and to retrieve the associated data file(s) accordingly.

**Domain of Interest**: Biology, Ecology, Ocean

**Professional Domain of User:** Scientist/Researcher

**Primary User-Stakeholder Relationship:** Machine User -> Data Producer/Stakeholder

**Secondary User-Stakeholder Relationship:** Human User -> Project/Stakeholder

**Primary Scope:** Quantitative-Science (for details see: <http://bit.ly/quantsci>)

**Primary Scope Rationale:** The user (scientist/researcher) will use the evaluation criteria of how well the quantified scientific result is to his/her own study interest as the primary motivation to review a particular dataset.

**Secondary Scope:** Qualitative-Product (for details see: <http://bit.ly/qualprod>)

**Secondary Scope Rationale:** The user (scientist/researcher) will use the procedural or descriptive attributes of the dataset to help determine if the dataset is of interest.

**Use Case Chronology:**

1. User accesses the "Browse/Search Data" interface at SBC LTER site (<http://sbc.lternet.edu//data/dataCollectionsPortal.html>)
2. Using the pre-defined categories or the "Show All Data Collections" capability, user is able to review the initial descriptions of the datasets.
3. By selecting a dataset of interest, user is able to review the dataset's full metadata in detail.
4. If user determines that the dataset is of interest, s/he downloads the data files that are associated with the dataset.

**Success Criteria:**

1. User is able to access the SBC LTER data collections.
2. User is able to understand the charactersitics of the datasets and determine if the dataset is of interest by reviewing the metadata provided by SBC LTER.
3. User is able to download the desired data files successfully.

**Data Quality Keywords**: accessibility, documentation, metadata, missing data

**Group #4:**

**Thematic Interest Areas**: Calibration/Validation & Flags/Indicators

**Evaluation Form**: <http://bit.ly/esip2016_uceval_duerr>

**Use Case Title**: Citizen Science

**Use Case Author**: Ruth Duerr

**Use Case Narrative**: A researcher working with the public (typically some community or a group of people with an interest in the researcher's topic) is accumulating data acquired by those people. Depending on the user the quality of the data acquired may vary - the researcher needs mechanisms to allow them know what data to trust.

**Domain of Interest**: Atmosphere, Biology, Climate, Cryosphere, Geology, Ecology, Hydrology, Land, Ocean, All of the above

**Professional Domain of User:** Scientist/Researcher

**Primary User-Stakeholder Relationship:** Human User -> Data Producer/Stakeholder

**Secondary User-Stakeholder Relationship:** Machine User -> Data Producer/Stakeholder

**Primary Scope:** Quantitative-Science (for details see: <http://bit.ly/quantsci>)

**Primary Scope Rationale:** N/A

**Secondary Scope:** N/A

**Secondary Scope Rationale:** N/A

**Use Case Chronology:**

The researcher has a system that allows "citizens" to enter information about something (ranging from identification of galaxy types when shown a picture, to identification of bird types, to measurements of water temperature, rain amount, etc.). Given that the training and education of these "citizens" is an unknown and these are all entries from volunteers; the question is how can the researcher establish which user data is better/worse than the norm.

Step 1: User X provides their data (perhaps a single record) to the system.

Step 2: Researcher/system evaluates new data (based on previous entries, comparison to other user entries, etc.) and places quality statement on this data

**Success Criteria:** Algorithm or mechanism to evaluate the data entry exists and has been implemented.

**Data Quality Keywords**: trustability