Next steps for the Data Stewardship Committee’s Use Case Activity

Discussion and brainstorming session

Sarah Ramdeen

Wednesday January 8th

Data Stewardship Committee

Dupont room
Agenda

• Background and overview of Use case activities
• Recap AGU outcomes
  • Presentation
  • Town hall
  • Poster (which will be up at the poster session this afternoon)
• Who is the audience of this activity?
• How do we publicize or publish these activities to get more engagement
• Discussion on next steps within ESIP
Background

- PCCS started in 2011 out of work of NASA and NOAA researchers within ESIP
- Focuses on categories and content based on the data life cycles of remote sensing missions
- Generalized to cover broader set of activities – but contain information needed for long term understandability and usability of earth science data products.
- Our cases will act as specifications for developing PCCS-based systems.
  - Examination of the PCCS to determine if any additions are needed to cover the various use cases, and also provide rationale and indicate priorities for preservation.
  - Should demonstrate the perspectives of the data archivist, data user, and the data consumer.
Provenance and Context Content Standard

PCCS provides a framework for 'what' must be captured or preserved as opposed to describing only 'how' it should be done. The standard currently provides data managers with content items aligned to eight key categories:

1. Preflight/Pre-Operations: Instrument/Sensor characteristics including pre-flight/pre-operations performance measurements; calibration method; radiometric and spectral response; noise characteristics; detector offsets
2. Products (Data): Raw instrument data, Level 0 through Level 4 data products and associated metadata
3. Product Documentation: Structure and format with definitions of all parameters and metadata fields; algorithm theoretical basis; processing history and product version history; quality assessment information
4. Mission Calibration: Instrument/sensor calibration method (in operation) and data; calibration software used to generate lookup tables; instrument and platform events and maneuvers
5. Product Software: Product generation software and software documentation
6. Algorithm Input: Any ancillary data or other data sets used in generation or calibration of the data or derived product; ancillary data description and documentation
7. Validation: Record and data sets
8. Software Tools: product access (reader) tools.
Overview of Use case activities

• Each case should:
  • Identify what PCCS content items are required and why

• Mapping of use cases to content items helps with the rationale and priority for preserving the materials

• Effects the development of the PCCS
  • If no use cases are identified for some of the items
    • Look harder for use cases, or
    • Eliminate item from PCCS

• http://wiki.esipfed.org/index.php/Preservation_Use_Case_Activity
Current categories

Creating a data set:
- Releasing a data set*
- Publishing a data set *
- Creating a long term trend data set from multiple data sets*
  - A research 100 years in the future is examining the historical record.
- Giving credit to people involved in the data set
- Describing appropriate uses of a data set
  - Asserting quality of data set
  - Intellectual property rights
  - Policies
  - Creating citations for published data set
  - Validation of data

Using a data set:
- Obtaining Data*
  - Discovering data*
    - Preliminary search*
    - Choosing a data set from multiple similar choices*
    - Assessing data
- Citing use of a data set
- Applications for data
  - Analysis of new version of data set
  - How do the improvements affect our use of the data?
  - Sharing data set for collaboration
  - Comparing multiple data sets
  - Reproducing a dataset*
  - Verification of an experiment

* Indicates an existing case study
Last summer and use case template

• Outcomes: Created a template:
  1. Summary
  2. Objective/Context
  3. Actors
  4. Sequence of events
  5. PCCS artifacts

• Notes from the last session:
  • http://commons.esipfed.org/node/1470
Recap AGU outcomes - Presentation

Applying the Emerging PCCS to Physical Objects in a Core Repository

• Use Case: Applying PCCS to a Core Repository
  • Geological Survey of Alabama (GSA) houses cores, cuttings, and other physical samples collected from oil and gas wells drilled in the state

• Motivations
  • Better use of resources
  • Interoperability (and therefore potential for data use and reuse) increases
  • Discoverability increases with standardization

• Outcomes of the application and the talk
  • Possible publication – but where would it be most effective?
Recap AGU outcomes – Town Hall

Charge to Panelists, address one or more:

• A specific example of data stewardship need/issue/concern/opportunity within your expertise

• Barriers or areas for improvement if we could better engage across audience

• In your efforts, what can you already provide to cross-disciplinary ventures?
Recap AGU outcomes – Town Hall

Key topic points:
• Introduction included brief overview of why standards fail
• Overview from panelists talk
  • Mark Parsons – Unanticipated user; providing context
  • Kerstin Lehnert – long tail, need a demand from the community for standards, repositories etc.
  • Sky Bristol– ancillary data, are we capturing everything we should?
  • Karl Benedict – small research – getting it in the preservation process
  • Chris Lenhardt – Structured and unstructured data
• Closing by Anne Wilson introducing the Data Study efforts by ESIP
Recap AGU outcomes - Poster

• Overview
  • Provided background information on the PCCS Use case activities
  • Outlined the template with definitions and examples
  • Call to action – attempt to engage new users and contributors and increase awareness of these activities

• Results
  • Lots of foot traffic, new audiences reached.
  • Is now a concise document to point to for introduction to this activity
  • Won an OSPA (!)
Who is the audience of this activity?

• From the data creator side
  • Who are working with PCCS in ESIP?
  • Who might be the people outside of ESIP who would or are using PCCS?

• From the user side
  • Who are using the information captured in the PCCS?

• Who might be interested in creating use cases

• Who might be interested in using the use cases
How do we publicize or publish these activities to get more engagement

• Brainstorming – how do we get people aware of these materials?
• If we were to write up the results of the town hall and these activities – what venue would be best for publishing these results?
• Overall suggestions?
Discussion on next steps within ESIP

- Creating more documentation
- Editing existing use cases
  - To better fit the template and to fill out missing information
- Creating new use cases
- Making others aware of these materials