

## ESIP Preservation and Stewardship Cluster Monthly Telecon 2012-08-01

Attendees: Curt Tilmes, Sarah Ramdeen, Hook Hua, Erin Robinson, Anne Wilson, Nancy Hoebelheinrich, Helen Conover, Rama.

### Agenda:

1. [Poll](#) for determining Week, Day and Time for Monthly Telecons. To be filled out by August 15th.
2. [W3C Provenance Last Call](#)
3. ESIP Summer Meeting Review - [Meeting Notes](#)

### Notes:

A poll for determining future meeting times is now available. In order to have say, please fill it out by August 15th.

W3C provenance message - Curt wanted to talk about the email that went out. The last call working draft is out for review. For this group, we intend to develop an Earth Science extension to that. 1) Does it support our work groups and 2) will it support that type of extension - are there any restrictions that would make it difficult to do our earth science work.

Speak up on the mailing list if you are interested. Hook and Curt will be working on that with a few others. Please at least read through it and provide suggestions.

### ESIP summer meeting -

We discussed a few things relating to identifiers and next steps. Curt would like to encourage those who volunteered to be part of those activities to update some of the wiki pages and focused activities where we can drive those efforts through the mailing list.

The identifiers - need to get others to use our identifiers and other next steps now that we have developed these guidelines. There was talk of developing a paper. Other comments or suggestions?

Nancy - one issue that came up was a request for how to handle versions of data sets. There was also some discussion about semantics for identifiers, given what people saw from John Moses's work.

Curt - Landing pages - DataOne is doing something different (not including landing pages) and what should be on DOI data set pages - is there guidance we would like to offer as a group?

Rama - they are starting to implement these things now, but does not have the details currently.

Curt asked if we should create a white paper or wiki guidelines on these topics?

Rama suggested a page or two at most.

Curt would like to get in to protocols and formats, more concrete as to what should be on the landing pages. Hook mentioned that landing pages should have RDF information and should be machine readable as well. Potentially linking to provenance metadata as well. OSDD documents that describe in a machine readable way, what you are landing on.

Curt spoke with (?) where there is a structure to their landing pages. They use handles instead of DOIs but the landing pages have paths to other specific information sets. Curt would like to form a working group and survey everyone about these issues.

Hook, spoke with Bruce Volmer at GES-DISC who is handling the landing pages for the measures project. And discussed having these be automatically generated (instead of manually generated).

Curt - maybe a survey to data centers to see what the current practices are, and then we can discuss what we like best and how to improve things. Hook - also if they are handling landing pages themselves. Rama pointed out that not all groups are using landing pages.

Helen - some people might have landing pages without having DOIs and can still take part in a survey.

Curt will post some questions on the wiki page to help develop a survey. We can send it to ESIP all. Rama suggested having some sample landing pages with the survey, or at least to look at a couple of samples. Curt suggested a few possible NASA examples.

Curt - Identifiers questions. Asked Nancy if there were specific activities (papers etc) that people thought would be useful to pursue. (other than landing pages). Nancy - There were some questions about citations. And how to including identifiers in citations for granular level data.

Curt - We did talk about creating guidelines for reviewers and editors. Nancy - Brent and Deborah were going to work on guidelines for journal publishers to help them ask for the right information and they also wanted to provide guidelines for researchers in case the journals did not know what to ask for. (Enough information for both sides). Rama - asked if the information for editors would include notes for peer reviewers. Nancy agreed, that these things, and how to ask for citations for data.

Curt asked Nancy to create a wiki page to describe this issue. This will help promote discussion.

Curt mentioned the earth science extension to the W3C - asked Hook about timelines and where to get started. When and how should we address it?

Hook - it is more mature right now specifically the provenance model and ontology encoding. This along with what Rama has been working on (the PCCS specifications). Suggested a preservation ontology for the PCCS. Timing ways, the W3C is slated to complete their recommendations by March/April of next year. The content is grounded, and should not have any major changes from this point on.

Curt thinks it is better do work on this now instead of waiting until it is finalized.

Hook - perhaps doing it now or soon, can bring the efforts of this activity into the W3C community/attention. The W3C holds a lot of weight. Maybe we should also ask who the customer is, what are the use cases (from an ESIP perspective). The PCCS came to mind first to Hook. Getting the final details worked out etc. We can make this a formal data model of the PCCS. Like an end to end preservation. The W3C is generic, and the specifics (no software aspects of preservation) go beyond the context.

Curt - deliverables for this would be a Earth Science node for the W3C recommendation.

Hook - why should we use this and who are the customers - these should be asked first.

Rama - was at IGARSS last week and gave a presentation on standards. Discussed with people there about creating an ISO standard.

Hook - the ISO standards on metadata - could be considered an umbrella representation, and can cite some more specific metadata standards - more domain specific standards. (ISO 19115).

Curt - PCCS still has some questions in it that should be addressed.

Hook - an earth science provenance - the current W3C standard does not address domains. This would include notes on collections and granules. Which we have different definitions of collections and data sets then the W3C standard.

Hook asked Rama if this type of activity - defining a data model for PCCS, would be useful? Rama would have to think about ISO 19115 and a gap analysis on the PCCS the ISO standard.

Rama - PCCS gives the list of what is important.

Curt - Discussed the importance of interoperability.

Hook - another good distinction, a prov earth science - we can encode things like people who are associated and some of the data generation, who acted on behalf, or what was generated by whom. Peoples, organizations, software agents. It seems to cover some of the organization

things that PCCS is trying to tackle.

Curt -that brings up another issue, going beyond data sets, recommending identifiers for everything else. What are the next steps? Hook - how useful, we should look at use cases we want to capture, which will help us focus on what we would be supporting. Curt - what are the questions we would like to answer. Hook - who are the stakeholders we would like to support. Curt - how would we gather these? Hook - need to find out who the stakeholders are. Nancy - that might be worth a call to the ESIP all. Generally what type of use cases might be useful. Hook - a lot of that overlaps with the use cases we have already been collecting. We have to assess how useful those are to this particular effort. Most of what Rama has already done with PCCS falls into this.

Curt encourages everyone to go read those blog entries that describe the prov specifications, and any comments you can make can only improve what is there now.

Curt will send a note about making a wiki page on a couple of topics. Nancy will do one on the guidelines for reviewers/editors. These can help us produce some more guidelines to run through this process. Nancy will also contact Deborah and Brent. This can be the focus of a future telecon.

Reminder to fill out the poll for scheduling future meetings.