

ESIP Preservation and Stewardship Committee Monthly Telecon 2013-10-02

Attendees: Bruce, Ruth Duerr, Erin Robinson, Sarah Ramdeen, Anne Wilson, Vicky Wolf, Steve Olding, Bob Downs, Mark Parsons, Marshal Ma

Agenda:

- ESIP budget request
- RDA collaboration liaison report out
- Planning for winter meeting
- Subgroup activities report outs ([see document](#) and [telecon list](#))
 - Use Case Activity
 - Data Collection Structure
 - PCCS
 - Identifiers
 - Citations
 - Data Management Training
 - Data Stewardship Principles
 - Preservation Ontology
 - Physical Object Preservation
 - Data Decadal Survey
 - Collection Structure group

Notes:

Bruce requested that the meeting he has scheduled for next week be postponed if the government shutdown continues.

Ruth asked if the budget request was submitted. Erin had received it but will double check. Ruth said both should have been turned in and would like to know if otherwise.

RDA report

Ruth reported on the RDA meeting - Ruth attended the Data publication sessions, and (???). She was surprised how quickly Journals were heading into the data citation arena. Often journals have links to the data (if you have a DOI) and ways for visualizing the data. Of importance to this group, given our data citations guidelines etc. is the factor that a number of other groups, like the DCC and the CODATA working group and others are coming up with their own statements on data citation principles. A group of members of these groups, under Paul Uh
lir, to see if they could come up with a global statement that all groups could agree too. Ruth is the rep for that, ESIP is a member of this group. A one day meeting was held to review these statements, these are high level details. Ruth shared this with us because they are intending to publish these drafts and hold a comment period on them. They would like to have each of the founding committees or groups, to make sure they vet the wording of these things and hopefully

approve them. That would require us to update the wording in some of our stewardship principles, and add a high level front to our citation guidelines. She is expecting in a month or so this document will come out and we as a community should look at it for suggested changes or ratify it, get it into a form, so that our documents can go to the assembly for approval.

Clarification was asked as far as which documents would need to be approved. This is a high level document and only principles, while our guidelines are lower level. We would want to make sure what we are doing as an Earth Science community are consistent with these high level principles, which are meant for broader communities.

If we agree with their principles, we have two documents that might be affected by that, and we would want to change our documents to be consistent - citations of the data management principles and data citations guidelines (which has a section on why you would want to cite your data).

Bruce asked if there was any international input/translation issues. Ruth noted that half or more of the organizations which have come up with these principles and have representation in this group, are not based in America. For example, CODATA. Ruth mentioned that there are some specific languages being covered, and a group from Europe and Japan will be handling these things.

As soon as Ruth hears something, she will forward it to this group for consideration through the mailing list.

Planning for the winter meeting

Ruth - What would we like to do?

Bruce mentioned the collection structure group - not sure if another telecon would be useful, but would like to have a face to face meeting about how we deal with objects that need to go into an inventory that might have written, digital or other replicas that might be indistinguishable. He can think of 3 policies for dealing with these things (1) like print copies, (2) you might have to track just number of copies (3) or perfect counts including the state of the item. Also, how do we deal with fragmentation when dealing with migration? HDF has a version 6 -what do we do with the obsolete versions when the file format is not the same? Need a distinction between textual formats (where the text is sensitive) and numerical files, where you can permute the order of the numbers etc. but you have not altered the topic. Which copy do we have? Also developing algorithms that differentiate between different copies. Which also has three different issues - how to store the representation information.

Bruce would like to have a face to face session with some concrete examples (gave a specific example). This data set is simple, stored in ASCII. Ruth mentioned she has some solutions to some of the questions Bruce mentioned and can share them later.

Ruth asked which other topics the group would like to see addressed?

Bruce mentioned how do you do a taxonomy. He has two upcoming publications on this topic. Mentioned scientifically literate people and semantic heterogeneity problems. He can submit these topics later this week.

Anne expects to propose a session related to the DDS workshop (which would occur Tuesday before the meeting) and would like one working session to digest the results of the workshop.

Ruth ask Erin how many days/how many breakout groups are there during the winter meeting?

Erin said there are 8 - because wednesday tends to be a full day, and then a plenary Thursday with a half day of sessions afterwards, and Friday people are not able to stay a full day. So 8 sessions total (1.5 hour sessions). Anne asked for clarification on total number of sessions. Erin said there tends to be 5% per time session. So looking at like 40 sessions total.

Ruth asked if Sarah wanted to have a session on the Use Case activities. Anne suggested a session as a townhall next steps and a recap of the meeting. Anne mentioned the townhall was cast as asking for feedback from the community and we should do something with it. We should suggest a townhall recap session. Going back to the Use Case activities - there was a discussion on if we should host a session and what the focus should be - creating more use cases or discussing how users would apply these, and engaging the user community. Bruce had some suggestions and discussed his background related to this topic, specifically marketing. Sarah will send something out after the call and will incorporate these suggestions.

Erin said, the call for sessions should come out in the next week or so.

Ruth thought a session on preservation, what is happening in the various RDA groups, but that would depend on the other liaisons and what they have gotten a handle on in their RDA connections. Anne and Ruth are liaisons, Denise and perhaps one other but we are not sure. It might be Curt.

Bruce had a related question, regarding RDA and another group. Ruth was able to address that - having to do with trusted repository audits and the OAIS reference model. Bob volunteered to discuss this topic at the winter meeting. Bob is part of the working group for this document, and reported back on some of this work. ISO for auditors ISO 16919. It has not been released yet.

Ruth asked if someone could talk about this as a session (90 minutes) - should it be part of a what is going on at RDA (as a segment?). Bob thought that would be great.

Mark joined the call and added in some clarification on the topic from the RDA view point. He also mentioned a preservation group at RDA.
RDA/WDS Joint Group on repository audits etc.

<https://www.rd-alliance.org/internal-groups/repository-audit-and-certification-ig.html>

Preservation e-Infrastructure group:

<https://www.rd-alliance.org/internal-groups/preservation-e-infrastructure-ig.html>

Ruth thought a session on RDA and how it impacts ESIP would be a good thing to do.

Ruth mentioned Thorny Staples - who gave an overview of Fedora, $\frac{3}{4}$ years ago. He has since moved to the smithsonian, and he is putting together a system for researchers to manage data, using derivative of Islandora tailored to scientific communities. He would like to come back to ESIP and give a demo of those developments. Would this group be interested in seeing something like that? Anne mentioned interest in this topic. Ruth said he would be very happy to be invited. Bob agreed that it would be a good idea, convenient with him being in DC.

This brings us to 6 sessions along with a planning session.

Islandora was discussed in more details, particularly the background for the development of the system and how it would be implemented. Mark Legit (sp?) was mentioned by Mark as someone else who works with this software.

Ruth thought we could continue the conversation next month and thanked everyone for coming.