



ErpaChat session on The Selection, Appraisal and Retention of Digital Scientific Data
March 16, 2004
15:00 – 16:15 GMT

Participants:

William Anderson, CODATA Data Preservation Task Group, USA
Renata Arovelius, Swedish University of Agricultural Sciences, Sweden
Melissa Cragin, University of Illinois-Champaign, USA
Joy Davidson, ERPANET, UK
Terry Eastwood, University of British Columbia, Canada
John Faundeen, U.S. Geological Survey, EROS Data Centre, USA
Meredith Lane, Global Biodiversity Information Facility, Denmark
Peter McKinney, ERPANET, UK

Joy Davidson:

Welcome to the erpaChat session on the Selection, Appraisal and Retention of Digital Scientific Data. The proceedings will be transcribed and will be available on the ERPANET site soon.

Terry Eastwood:

Regarding terminology, InterPARES is producing an interdisciplinary dictionary of terminology used by various disciplines, including scientific disciplines. The dictionary will be available at www.interpares.org.

Pete McKinney:

Ok, first, I'll introduce myself since I wasn't at Lisbon. I'm Peter McKinney and I'm the coordinator of erpanet. I'm glad you could all make it! Terry, that's a very timely point...the 1st question we had lined up is: It became clear by the end of the workshop that the understanding of the term appraisal is not universally understood. CODATA provided a glossary of key terms on the last day of the workshop to help address this problem. What other steps can we take to eradicate terminological misunderstandings?

Joy Davidson:

That sounds like a very promising tool Terry. Will it be freely available?

John Faundeen:

Terminology, or semantics, were obstacles at Lisbon....much more than I had anticipated.

Terry Eastwood:

Yes, as the dictionary develops, people will be able to consult it on the website.

Joy Davidson:

It seems to be an issue that no one really anticipated. However, now that we recognize the problem, we can start to work on a solution. Terry's description of the dictionary sounds promising.

John Faundeen:

The Society of American Archivists (SAA) has a publication that can help, too.

Joy Davidson:

Terry - will people be able to add terms to the dictionary for consideration? Will it be an ongoing project?

William Anderson:

Terry: have you thought about using a Wiki (a la Wikipedia - online collab encyclopedia) as a vehicle for discussion of terminology?

Joy Davidson:

It seems to me that getting the scientific and archival communities talking helped to illustrate the problem of semantics. So, does the dictionary you describe involve both scientists and archivists?

Terry Eastwood:

Right now only researchers in the project can add terms. By the way the SAA is about to reissue its glossary. I'll pass on the Wiki idea.

William Anderson:

I think there are ways to validate contributors; there's a lot on energy on wikis out there.

Pete McKinney:

Are there any ways you're going to try and ensure take up of the terminology Terry?

Terry Eastwood:

One focus of the InterPARES research is on scientific records, so we face the problem of terminology in science as compared to archival work, computer science and so on.

Joy Davidson:

It does seem that a collaborative effort will be necessary - involving all stakeholders. Can we tackle this at any upcoming events?

Pete McKinney:

So do you think it will every truly be a universal dictionary, employed by most?

John Faundeen:

At upcoming events and through listservers/email/etc...to cover more focus faster.

Terry Eastwood:

Good question from Pete. We need to publicize our work, and I think provide some vehicle for input outside the project. I'll pass this on to colleagues.

Terry Eastwood:

I don't think it will cover every Scientific discipline or every term one might want, but it will be an important start at taming the terminological tiger.

William Anderson:

Actually I think that "universal" is relative, but communities who need to share do develop ways to communicate.

Renata Arovelius:

The SS-ISO 15489-1 on records management may also be helpful; even the terminology used in the reference model of OAIS.

Meredith Lane:

This is certainly a problem in our area, too. Getting "uptake" by the broader community of GBIF standards and terms is a long, slow process.

John Faundeen:

Universal is not attainable, but if at least we could each understand the others' meaning, we would be further along.

Pete McKinney:

I'm just dubious over the universal part, I agree it is relative, but in past erpaChats(metadata particularly) we've run up against problems of different disciplines...

Pete McKinney:

Is the OAIS terminology agreed by everyone?

Renata Arovelius:

No, as far as I know.

John Faundeen:

OAIS is relatively new and thus, not even known by many folks.

Joy Davidson:

I agree with John – understanding the basics is necessary before we can move on to specifics.

Terry Eastwood:

I see a possibility for ERPANET, CODATA, InterPARES cooperation. I will promote the idea at our end.

Joy Davidson:

Good idea Terry - I'll see what we can do at our end.#

William Anderson:

I think it's a good start perhaps to lay out the landscape of meanings,, say for metadata, by discipline or ...Terry, I'd like to follow up on potential CODATA collaboration.

Joy Davidson:

Luigi Fusco has asked me to pass on an invitation to a conference in Fieschi, Italy in October. The URL for more info is <http://www.congrex.nl/04a08/>

John Faundeen:
I am presenting at that conference.

Joy Davidson:
Perhaps we can agree to continue to work on this issue after the session. If you'd like to move on to the next question, I thought we might look at how we can increase the awareness of funding bodies of the importance of digital curation and preservation. Any thoughts or suggestions?

William Anderson:
Opinion or report pieces in scientific/policy outlets? I vote for the "water on a rock; steady drip" approach; review the issue constantly.

Melissa Cragin:
One way to do this is to have researchers submit grants to study some of the phenomena involved.

Renata Arovelius:
Perhaps costs for archiving/digital preservation should be specified already in the funds application.

Pete McKinney:
What we're finding in our case studies is that companies do not have digital preservation policies per se, but are relying on good curation to get them through just now. Is this a matter that digital preservation is a little bit sexier just now than curation.

Joy Davidson:
It seems that access is something that funders are willing to finance, just not for the long-term! Perhaps this is something that we can work together to promote within our respective disciplines.

William Anderson:
It's not just discipline; it's the nature of the media and technology.

Terry Eastwood:
I think that digital preservation issues are now on the radar scope of funding bodies, but I agree that they need help seeing the long-term perspective.

William Anderson:
Paper can hang around without much care.

Melissa Cragin:
I also think that the concept(s) and practice of data curation would appeal to some people in the LIS field.

Pete McKinney:

Certainly over here it is the libraries that are more on top of curation....

Melissa Cragin:

Not so sure this is the case in the US, though. And, how about the rest of the world?

Joy Davidson:

I would say that it is probably the case in Canada as well - Terry?

Terry Eastwood:

Situation in Canada much like that in the U.S. Libraries are archives have the same problem in general, but it differs in the details. Collaboration is definitely not all it should be. Funding bodies have it on the radar screen.

Joy Davidson:

The new Digital Curation Centre funded by JISC will aim to promote curation in the e-science community. It is in the very early stages but it should help to address some of these issues.

Renata Arovelius:

Definitely not in Sweden. But there are some projects on preservation metadata in connection to electronic publishing.

John Faundeen:

Digital records are my challenge. Generally, photographs and paper are not a problem.

Joy Davidson:

The workshop in Lisbon aimed to start a process of communication and collaboration between the scientific and archival communities. What are some of the next steps we can take to ensure that this process continues?

Terry Eastwood:

Not to be Johnny one note, but ERPANET, InterPARES and CODATA could work to prepare a statement to send to funding bodies, a kind of manifesto.

Joy Davidson:

I think that might be a good idea Terry.

William Anderson:

Terry: I agree; let's look at this.

John Faundeen:

I like Terry's idea. By including scientists and archivists, this might get noticed.

William Anderson:

Also, there's work to include archivists in the upcoming COATA meeting in Berlin in November.

Terry Eastwood:

I will push the idea in my realm and keep in touch. Director of InterPARES will be in Berlin in the fall.

Joy Davidson:
Perhaps some kind of task force could be set up?

Terry Eastwood:
Indeed, we should have a vehicle to forward the matter. Sorry, I have to leave for a meeting in about 3 minutes.

William Anderson:
Terry; this could be brought up with CODATA.

Joy Davidson:
Thanks for your valuable insights Terry - I'll let you know when the chat proceedings are up on-line.

Pete McKinney:
Terry, thanks, I think we could have started something important here. We'll need to keep in touch over this.

Renata Arovelius:
We have within the ICA a university section and a subcommittee on science archives. We use to work together with scientists and ICT-people.

Terry Eastwood:
Please feel free to contact me on these matters at eastwood@interchange.ubc.ca .
Bye now.

Pete McKinney:
What is the collaboration between different archival areas in the ICA?

Joy Davidson:
The ICA would be great to have on board for something like this.

Melissa Cragin:
I would be interested in hearing more about this. Perhaps, Renata, you could point me to a web site or refer me to someone?

William Anderson:
Renata: CODATA is very interested in following up w ICA on archiving.

Renata Arovelius:
Yes, of course. We will have a seminar in Vienna in August. I am also currently the chair of the subcommittee.

Melissa Cragin:
What is the ICA, please?

Renata Arovelius:
International Council on Archives.

Melissa Cragin:
Thanks.

William Anderson:

Renata: is there info available on the August meeting?

Renata Arovelius:

Yes: please look at www.ica.org .

William Anderson:

Thanks.

Joy Davidson:

It is quite good to know about related events and conferences. Perhaps we could keep each other informed and encourage participation?

Renata Arovelius:

We are discussing now legal issues in connection to records of science. There are big differences in legislation, and as a consequence in handling of the material.

John Faundeen:

Also, how long to keep them at the point of origin before transferring to our National Archives.

William Anderson:

As well as issues of open access to publicly funded scientific data.

Renata Arovelius:

There are differences here as well. We don't need to transfer university records to National Archives in Sweden. Access is ruled by the Freedom of Information Act.

Joy Davidson:

The Freedom of Information Act will have big implications for public information held in the UK as well starting in 2005.

William Anderson:

But how will records be maintained and accessible over the long term?

Renata Arovelius:

Yes, a good question. Migration? We must have a strategy for digital preservation.

John Faundeen:

And appraisal as not every notebook is necessary for preservation and archiving forever.

Joy Davidson:

Renata: Is this something that you are currently investigating at your organisation?
Very good point John! We can't save everything!

William Anderson:

We heard this point (always worth repeating) several times in Lisbon.

Renata Arovelius:

Yes. We have started a project (together with scientist) on digital preservation. We are waiting for decision about funding. If positive, we will test 6 different databases with DSpace.

Melissa Cragin:

In addition to preservation, we will need modified practices of materials organization for use by multiple communities. I ought to have said, "access and use."

Joy Davidson:

Renata - There is a test project under way in the UK that uses ePrints to harvest metadata about institutional research projects. It is called eBank UK and you can find info on it at <http://www.ukoln.ac.uk/projects/ebank-uk/>

Renata Arovelius:

Thanks!

Melissa Cragin:

Sort of a follow-up to John, I do think we also need to understand more from the work-practices end of things. For example, what are the kind of activities (individual) scientists are currently engaged in to support archiving? of their own or their lab's materials...

Joy Davidson:

I am not sure about this, but I imagine it would be quite sporadic and depend on the individual institution.

Renata Arovelius:

It depends on. If there are longitudinal projects the scientists are very interested in archiving

William Anderson:

Do studies exist? there have been social studies of scientific practice ...

Meredith Lane:

Regarding individual scientists, unless their employers insist on such activities, they usually don't do it.

Joy Davidson:

I agree Meredith - I think that is fair to say outside the scientific community as well.

John Faundeen:

The US had a resource called the National Media Laboratory that may becoming back into existence, at least for civil government agencies. This entity provided excellent support on materials related to archiving and specifically helped us design and build ovens for m

Renata Arovelius:

We have special courses for post-graduated students in archiving and handling of scientific materials.

Joy Davidson:

Pedro Fernandes mentioned the bio-informatician training at the Gulbenkian. This hybrid professional sounds very promising.

William Anderson:

I think that scientists don't get much training in what's needed to preserve data for longer than the research project.

Joy Davidson:

I think that you are right. It is something that affects all disciplines and is probably not given enough attention.

William Anderson:

disciplines vary of course; astronomy is long-term based.

Renata Arovelius:

Yes, so the archivists have to tell why is it important.

Joy Davidson:

I think we have covered quite a bit today. Are there any other questions that anyone would like to ask? Any other comments?

Joy Davidson:

If not, perhaps we can end the session by recognizing that a lot more work needs to be done. I think that by continuing to work together we can make some good headway into solving some of the problems.

Renata Arovelius:

Just many thanks for the opportunity to participate and arranging the chat.

John Faundeen:

Yes, Thank you.

Melissa Cragin:

Yes, thank you!

Joy Davidson:

I will make sure that everyone knows when the proceedings are online. Please feel free to contact me with any other ideas or comments

William Anderson:

Joy, is this transcript going to be available?

Joy Davidson:

The transcript will be posted to the ERPANET site in the near future. I'll email everyone to let them know.

Joy Davidson:

I think that the dialogue has been very helpful and thank all of you for taking part. Your input has been terrific.