RDA VP16

Breakout 1

Monday, 9 November 2020/5:00 - 6:30 PM UTC

Physical Samples and Collections in the Research Data Ecosystem: Toward A Global and Multi-disciplinary Network of FAIR Physical Samples

Meeting Title: Toward a global and multi-disciplinary network of FAIR physical samples

Group Page: Physical Samples and Collections in the Research Data Ecosystem IG

P16 Session Page:

https://www.rd-alliance.org/plenaries/rda-16th-plenary-meeting-costa-rica-virtual/harmonising-multidisciplinary-vocabularies

Session Presentations:

https://docs.google.com/presentation/d/1qBiPKUx23bZMM2hBcBTxtTz_DYL0h6TK/edit#slide=id.g9eee8b607a_0_137

Attendees:

Please fill in your name, affiliation and email

No	Name	Affiliation	Email	Sample types you deal with
1	Lesley Wyborn	NCI, AuScope, ARDC	lesley.wyborn@anu.edu.au	Geological samples, environmental samples
2	Sarah Ramdeen	SESAR; Columbia University	sramdeen@ldeo.columbia.edu	GEO
3	Kerstin Lehnert	Columbia University; IGSN e.V.,	lehnert@ldeo.columbia.edu	ESES

		www.geosamples.org (SESAR)		
4.	Esther Plomp	Delft University of Technology	e.plomp@tudelft.nl	Archaeology, human tissue
5.	Val Stanley	Oregon State University / Marine and Geology Repository	val.stanley@oregonstate.edu	Marine sediment cores, terrestrial rock drill cores, rock hand samples
6.	Alex Hardisty	Cardiff University	hardistyar@cardiff.ac.uk	Natural sciences, DiSSCo
7.	Felix Ernst	Karlsruhe Institute of Technology	felix.ernst@kit.edu	Various (mainly just interested attendee)
8.	Maggie Hellström	Lund University, ICOS & ENVRI-FAIR	margareta.hellstrom@nateko.lu.se	My RI's data center works with info on samples from Earth & environmental sciences
9.	Federica Burla	Elsevier	f.burla@elsevier.com	Scientific Editor at Data in Brief (interested attendee)
10.	Catherine Patterson	Getty Conservation Institute	cpatterson@getty.edu	Geo, Museum materials (diverse types); new to RDA
11.	Rebecca Koskela	RDA-US	rebecca.koskela@rda-foundation.org	(just an interested attendee)
13.	Kirsten Elger	GFZ Potsdam	kelger@gfz-potsdam.de	Rock samples, drill cores and related samples, bio samples, etc
14.	Helen Glaves	BGS/ RDA TAB	hmg@bgs.ac.uk	Involved with IGSN (issuing agent) (geoscience samples) / RDA TAB observer
15.	Wouter Addink	Naturalis/ DiSSCo/ BDI IG	wouter.addink@naturalis.nl	natural science collections
16.	Dorothea Strecker	HU Berlin, re3data COREF	dorothea.strecker@hu-berlin.de	Interested in representing PIDs for physical samples in re3data

17	Jessica Salas	McGill University	yessica-elena.salas-navarro@mail.mcgill.ca	rocks/ water/soil samples
18	Markus Kubin	Helmholtz-Zentrum Berlin / HMC	markus.kubin@helmholtz-berlin.de	Chemical samples, solids, crystals, proteins
19	Gerrit Günther	Helmholtz-Zentrum Berlin / HMC	gerrit.guenther@helmholtz-berlin.de	Chemical samples
20	Marie-Claude Deboin	CIRAD (French Research Institute on Tropical Agriculture)	marie-claude.deboin@cirad.fr	Wood & Plants Samples
21	Anne Cambon-Thomsen	CNRS, Inserm and Univ of Toulouse, france	anne.cambon-thomsen@univ-tlse3.fr	Human biological samples collections (biobanks), patients and healthy populations -Ethical aspects
22	Lars Möller	PANGAEA	lars.moeller@pangaea.de	Biogeochemical& geochemical samples in the marine/coastal domain
23	Stéphanie Cheviron	University of Strasbourg, France	scheviron@unistra.fr	None, but as a research data librarian I work with researchers who do manage botanical, geological and paleontological samples
24	Claudia Bauzer Medeiros	University of Campinas	cmbm@ic.unicamp.br	At the moment, none, but have worked with: plants (and herbaria), insects (flies and butterflies), cancerous tissues, and have helped researchers with data extracted from indirect physical evidence of presence of specific species (e.g., vomit or footprints). Lots of this comes from ecological studies but also taxonomic research, and curation of physical samples in my university's Zoology museum
25	Ted Habermann	Metadata Game Changers	ted.habermann@gmail.com	Collection metadata in the USGS National Digital Catalog and other metadata standards

26	Faisal Fadlelmola	University of Khartoum, Sudan and H3ABioNet Sudan Node	faisal.mohamed@hotmail.com	Genomic, health and Bioinformatics data
27	Neil Davies	University of California Berkeley	ndavies@berkeley.edu	Samples in all domains (on Earth anyway and perhaps a meteorite or two)
28	Mogens Thomsen	Inserm, University of Toulouse	mogens.thomsen@univ-tlse3.fr	Human biological samples

AGENDA:

1. **Welcome (10 minutes)**
    - *Logistics & overview*
    - *Welcome to new co-chair Esther Plomp, Panelists representing ebrief presentation of her essay that recently won the CODATA Connect award (see here for the essay)*
2. **Emerging or planned infrastructure and services projects for FAIR samples and collections (40 minutes)**
    - *Interoperable Enriched Specimen Information Models (Alex Hardisty)*
    - *iSamples (Neil Davies)*
    - *ESIP Samples Cluster (Val Stanley)*

    i. Background presentation: https://youtu.be/QpAReiPYHws?t=278 (Starts at 4:38, goes til 16:50)
  - *IGSN 2040 (Sarah Ramdeen)*
  - *AuScope Geochemistry Network (Alex Prent)*
    i. Presentation: https://drive.google.com/file/d/1uprJ5TVrPmpJLHNHsthYG_XPlqYUWzOs/view?usp=sharing
    ii. Related video: https://www.youtube.com/watch?v=ZDK7cqYwk44
  - *Questions & discussion (audience can contribute additional projects & initiatives that they are involved with)*
    i. What needs to happen in the future?
    ii. What is the priority right now to make that happen?
    iii. How can we get communities more involved?

3. **Harmonizing multi-disciplinary vocabularies for physical samples and collections (30 minutes)**
  - *Introduction to the concept of a vocab agnostic system for samples (Lesley)*
  - *Outcomes of the ESIP Plenary session "Proliferation of Vocabularies in Solid Earth, Space and Environmental sciences" (Sarah)*
  - *Reports from other metadata mapping efforts (TBD)*
  - *Audience participation to identify existing and emerging vocabularies*
  - *Propose 'low-hanging fruit' activity for harmonizing sample and collection vocabularies*

Who wants to participate in a group developing sufficient metadata for discovery and connection

- *Anousha*
- *Kirsten Elger*

4. **Synthesis and Next steps (10 minutes)**

MEETING NOTES:
Introductions

Emerging or planned infrastructure and services projects

Alex Hardisty is working on a large-scale European project “the distributed system of scientific collection (DiSSCo)”. DiSSCo is a collaboration of 120 natural history museums and research institutes that has the aim to mobilise the data from 1.5 million specimens that are held in research collections across Europe. In total there are ~3 billion specimens (preserved items, plants, insects, vertebrates, rocks, minerals, meteorites). These collections go back hundreds of years, access to them is difficult because the collections have to be visited or borrowed. It is now increasingly easier to make data available on the internet about these specimens and data mining and machine learning techniques can now be applied. In the collections there are multiple identification systems, we are not planning to unify this system. We can enrich specimen information with links to third party sources (GenBank, Nucleotide). Digital specimens are underrepresented compared to other digital sources. We see the same ideas emerging in other parts of the work. USA: work of the biological collection network, extended network concept?,

iSamples (Neil Davies)

Neil is the Executive Director University of California Gump South Pacific Research Station. We have been working on biological samples and how to track them from the field to biological collections and museums. How do we track and trace these? How do we sample the natural world? These samples are of very little value without the metadata around them. We have a lot of legacy work now, because we have historical samples, but we’re also generating more samples every day as they are collected now and in the future. We need to be able to trace the origin and downstream uses of biological specimens. Working with biological specimens primarily, sequences end up in banks all over the world and we need to figure out how to connect these streams of information. Sometimes the samples used in geological studies can be reused by other fields and domains. We need to be able to communicate between these domains and we need ways to map and exchange knowledge. Then we can see the value of reuse and specimens. To do that we need to understand what the process is. We need different metadata depending on the use of the specimen. One thing that we need to understand is the minimal amount of information is needed, what has to be collected straight away in the field: we only have limited resources and we need to improve reuse. These decisions have to be made along the chain. Ethical/legal social domain is critical and has a lot of metadata that is of use to other fields: what are we doing, who is it being used by and what is it used for. This information is often needed by funders. Often things are being used downstream but places where they

came from can't reconstitute the data and they are not always informed about the knowledge that is generated. Permits, CARE principles, how do we combine all this without burdening the researchers/field workers. We need the infrastructure with the researcher to make it easy to collect metadata.

ESIP Samples Cluster (Val Stanley)
Val Stanley, Antarctic Core Curator, Marine & Geology Repository. Val works at a marine sediment core repository, we have over 23 miles of sediments stored and host terrestrial rocks (collected by hand and hammers) from around the world. ESIP samples cluster was brainstormed initially over a year ago out of a curator meeting. This is a meeting where curators of repositories get together to discuss strategy of curating samples, exchange information. We needed a way to continue these conversations between these meetings. We wanted to discuss vocabularies and standards but this is more broadly applicable than just the curators. In collaboration with Sarah Ramdeen they discussed what the conversations are that need to happen. ESIP cluster was set up to discuss the curation of physical samples: identifiers, FAIR, infrastructure needs of researchers. The discussion about curation is not just related to the geoscience community but also other communities: we need to get people talking to each other. We had an information gathering session in the summer to brainstorm and come up with a proposal of ideas that needed to be discussed more broadly in a cluster. Currently thinking about working groups, projects that need to be set up and a webinar series to connect people.

IGSN 2040 (Sarah Ramdeen)
Sarah is a data curator for System for Earth Sample Registration (SESAR), allocating agent for the IGSN. The IGSN is the global sample number for physical samples. Sarah reviews sample metadata and approves the issue of IGSNs to the user. There is this need for a globally unique persistent identifier to track samples. Even when a physical sample is destroyed you still need the data, and the IGSN can link some of these different materials. It can also help to find a particular sample at a repository. THE IGSN 2040 is founded by the Sloan foundation to redesign the current IGSN infrasture. As part of this the IGSN 2040 has organised a series of workshops, discussing the business model, technical infrastructure. They redesigned the core services and redesigned these processes. For the PIDs to be FAIR and trustworthy the business that is offering the PIDs is sustainable. We want to use these services and know that it will still be there in the next five years. In addition IGSN 2040 also looks at partnerships, going back to the services that they offer and see which ones can be outsourced. How can they reach a broader, morediversie audience? And what services need to stay with IGSN because it fits better with their expertise. As IGSN grows (including archaeological, forensic samples), the needs of these different domains need to be voiced and articulated in the services that are offered. The final outcomes (reports) of IGSN 2040 will become available in spring 2021. They still organise workshops and the roadmap.

AuScope Geochemistry Network (Alex Prent)
Alex is the coordinator of the AuScope Geochemistry Network, a network of geochemistry laboratories. This will facilitate, for example, the creation of an isotopic atlas of Australia. It is part of NCRIS (national research infrastructure of Australia). AuScope's purpose is to create

wide and open access to earth science research infrastructure to resolve challenges. The AGN project team is situated at the Macquarie University, University of Melbourne and Curtin University. The AGNs data platform connects the labs. The AUSGeochem connects the laboratory to the repository. In order to do this sample metadata is needed. They make use of minimum metadata fields (5 are mandatory and 33 are highly recommended). In Geochem platform this metadata is then linked with services. You give the laboratory the IGSN number which is linked to the data which is stored on the institute repository which is shared with a cloud storage repository and linked to the Geochem database. We can visit that data, for which a placeholder is saved for minting, can be viewed through an access control layer (collaborator, private, public). Connecting the analytical data to the sample and the IGSN will make it future proof and will make it possible to trace where the sample was collected.

Discussion
The goal is convergence: There are a set of powerful projects and infrastructures for physical samples. We should not fragment the landscape too much, despite different requirements and communities. There are sample differences: samples that may not be worth a lot and destroyed and samples that are curated in museums. However, if we want to come to more consistent procedures, standards, metadata around samples we need to communicate.

What needs to happen for your initiative:

Alex Hardisty: for the biological community/collections. We need to arrive at a standard means to represent the physical data that is derived from specimens. We need to curate physical samples in an accessible manner. There is a distinct difference from the physical data and the digital data that is derived from this.

Neil Davies: The importance of tracking information from downstream and across domains. We need data management plans and embed the information there so that we can plan before the collections. Developing machine actionable data management plans is crucial here. We are working on the FAIR island project, an utopian idea where you can build the best data management policy and the community would follow the instructions. What would this be and how would you implement it so that it works? Can we demonstrate the benefits? What are you now able to do? The crucial thing is the case studies and examples so that the benefits are clear. Credit should flow back upstream to those that worked on it.

Val Stanley: Agrees with Neil and thinks that there are a lot of things that need work. Everyone needs to be in the room, we need to streamline workflows, sharing standards/protocols, making them official and getting the community to know how to work with physical samples. What do researchers and curators need to know?

Sarah Ramdeen: Getting engagement. We can develop the infrastructure but we need to engage the community to make use of it. We will need to educate people on what options are available. We need to come up with messages for researchers to communicate the benefits. Lesley Wyborn: The AuScope project starts with you having an IGSN for your sample, otherwise you cannot proceed. It is impressive how this is streamlined.

3. **Harmonizing multi-disciplinary vocabularies for physical samples and collections (30 minutes)**

Lesley: You cannot have a vocabulary for all around the world that describes all samples. For example, beer measures are described in different terms in Australia depending on the state you're in.

The solution is a "ring" vocabulary with a common kernel. See Lesley's RDA Poster.

The Fair Implementation Profiles (FIP), a collection of implementation choices made by a community of practise around the FAIR principles which can be captured as a dataset. What is the minimum data needed to describe the sample? We need agreement on this registration process to describe the metadata that is needed to make it reusable.

Outcomes of the ESIP Plenary session (Sarah Ramdeen):

See notes here:
https://2020esipsummermeeting.sched.com/event/clvR/and-samples-a-proposed-esip-cluster-for-the-physical-samples-community

During the ESIP plenary meeting six break out sessions discussed the vocabularies, ontologies, knowledge graphs within the community space. One breakout session focused on physical samples. One thing in particular was the need for common core metadata that can help in search in discovery. How can we make something discoverable so that it can support interdisciplinary research? We need interdisciplinary vocabulary, machine readable, some level of standardisation, but there needs to be user buy in for it to succeed. The recommendations for a kick off activity to establish the space: community, education, technical sprints related to vocabulary to work together on activities related to the vocabulary. How to involve the ESIP cluster in these discussions? Join the ESIP IG session (break out session 2) to discuss this further!

Audience participation to identify existing and emerging vocabularies

Propose 'low-hanging fruit' activity for harmonizing sample and collection vocabularies

- *Alex Hardisty*: You need different quantities of information for different purposes. What is digitisation and what are you collecting? Collecting contextual information from the field. How much information do we need? Before we can talk about shared vocabularies we need to know the scope and the purpose and criteria that it has to meet. Otherwise you will talk across purposes because there is a different understanding of what the objectives are.

- *Claudia Bauzer Medeiros*: has worked with groups of researchers from all fields to define a common vocabulary/metadata. Everyone comes to the conclusion that there is a limit. Findability (location, is not applicable to all samples, but is very important for others). The scope is very important. We need to collaborate across institutes, rather than trying to resolve everything ourselves.

Does anyone want to work on a minimum profile (sufficient metadata for discovery)?

- *Anusha, Kirsten Elger*

I (Maggie) would be happy to help out with a "FAQ flyer", or the like, if there's interest.

Next steps?

Start organising focused discussions on metadata and the need to create more awareness of what is already available and how physical samples are relevant and part of open science/FAIR/reproducible science. How can the IG contribute to growing awareness? How can we use the RDA and reach other stakeholders in the field?

- Kirsten PIDapalooza21 could be a place (January)
- Kerstin: how can we reach out to institutes to maximise impact? Can we reach out to administration and make them aware about the possibilities for their samples.
    - *Maggie Hellström*: One needs very clear arguments to present to the people and illustrate the benefits. Can we set up a flyer with the benefits/why? (like the 23 things from RDA).
    - How can we collect these stories? *Maggie*: We also didn't discuss ethical aspects. *Neil:* see the CARE principles and BioCultural Labels and Notices (and Traditional Knowledge Labels)
    - *Val Stanley*: training for early career/graduate students
    - *Esther*: Reaching out through Research Support offices at libraries?
- *Anne Cambon-Thomsen*: There are a lot of restrictions in using samples from human beings. We would need metadata to know what you are allowed to do with these samples (consent form information).

Conclusions: We should have another meeting in between the next interest group, such as focus meetings to discuss things more in depth.

Claudia: what is the main difference in metadata between digital and physical samples?

*Ted Habermann***:** I had two important take-aways from the discussion:

1. If we can change sentences that we use from "minimum metadata" to "minimum metadata for use case" it would be a huge step forward. Even minimum metadata for fundability, accessibility, interoperability, or reuse would be huge.

2. I agree stories are really important. Let's make sure we include user stories as well as data provider stories and keep in mind that what makes it easier for one of these groups typically makes it harder for the other.

Comments from the chat

Claudia M. Bauzer Medeiros: I believe the difference is that digital data can be copied, shared, reused, destroyed etc. Also, we must remember that one physical object may have several digital renderings depending on which aspect one wants to reuse. THis is why linking digital "incarnations" to the original physical object is important.

Maggie Hellström: @Esther: "all information available in one location": is that really necessary, as long as we can link different info sources (via PIDs)

Esther Plomp: @Maggie: If we have at least one location where we could find all that information, I think we could make the research a lot more findable! It doesn't necessarily indeed be located at that single location, as long as we have the references/pids gathered there.

Claudia M. Bauzer Medeiros: Question - what is a FAIR physical sample, most of them cannot be reused, and have to be kept and archived somewhere. THus, I would call them FAIr (small R) - reuse implies refactoring. If I reuse one of Esther's teeth in something else (say, a collage) no one else will have access to it.

Anne Cambon-Thomsen: lot of work done on that in BBMRI (the European infrastructure consortium for biobanking and biomolecular resources) concentrating on human biological collections, population of patients); the legislation on human body parts and on health data are distinct in many countries

Wouter Addink: @Claudia that is why we aim to create digital twins of physical specimens that can be FAIR

Kerstin Lehnert: Many samples can be re-used for new studies. A prominent example are ice cores or marine sediment cores that are made available via core repositories. Researchers take new samples from these and analyze new compositional aspects or use new techniques.

Claudia M. Bauzer Medeiros: @Kerstin, I agree, but physical samples are often too precious for you to take samples from. Tks @Wouter

Alex Hardisty: open digital specimens (openDS) aim to curate data from the moment of gathering.

Lesley Wyborn: This is why in some collections, there is a governance system where you have to apply to use a sample, particularly if your analysis method is destructive.

Alex Hardisty: open digital specimens also aim to curate data in cases where specimens don't (or no longer) exist.

Lesley Wyborn: The unique identification also means that you can link any analytical work done on a sample regardless as to where it is published

Val Stanley: Hi all, I work at a marine sediment repository. It is standard practice to maintain an archive volume which cannot be sampled destructively, but can be used for non-destructive analyses.

Alex Hardisty: @Esther, @Maggie: This is what open digital specimens aim to do - i.e., provide links to data derived from (or related to) specimens but not to collect all that data together into one place.

Claudia M. Bauzer Medeiros: @Alex I am confused. SUppose I generate 10 different digital files from one tooth - e.g., genetics, photos, x rays, 3D description for subsequent printing, xls file, textual comments. Is each of them a digital twin, or is the set of files the digital twin? They are digital renderings, no? Not twin.

Alex Hardisty : @Claudia - imagine the digital twin as a very large 'specimen label' that contains pointers to all that is known/exists about/from a specimen -i.e., includes pointers to all the renderings/images you mention - together with characteristics of those.

Neil Davies: @Alex and others. Digital Twin is a whole interesting topic! Also, important from some of the earlier discussion is machine actionable DMPs. I don’t think a digital representation of a specimen counts as a Digital Twin in the way that term is often used. (Requires feedback loop from real world entity to the twin and back again)

Alex Hardisty @Neil: An important attribute of a Digital Specimen (twin) is that you can perform operations on it, remotely over the Internet.
@Neil: twin with a small 't' :-)

Neil Davies: @alex — yes i like the small “t” and digital T(t)win concept is still emerging (e.g., we have a version we prefer to call an ecosystem avatar “Island Digital Ecosystem Avatar” which is a “twin” of a social ecological system)

Maggie Hellström: Many of my colleagues who are doing the actual collection & subsequent analyses of (atmospheric, oceanic & ecosystem) samples see no need to assign globally unique IDs - they say that having lab-internal ID systems is enough. They also claim that it would be too costly for them to adopt their existing databases etc with e.g. IGSNs. What arguments should I use to convince them to change their current practices?

Sarah Ramdeen: @Maggie, there are always going to be costs involved with any change, but investing the time, money, etc. now will make it easier to adopt new and developing technologies in the future (what is the saying, the best day to plant a tree was 20 years ago, the second best is now?). While their identifiers may be unique internally, having that global unique aspect is valuable, as is the resolvablity - which can be built into systems in the future as links and ways of aggregating connections... How frequently was my identifier listed in publication, reused by the community, etc.

Maggie Hellström: @Sarah: thanks. One of their main arguments is that up to now, no one has ever come back to them "post-publication" (of data and/or resulting papers) to request sample-related data & metadata. The "reusability and reproducibility" aspect is sort of uninteresting to them. But I'll give the tree planting analogy a try! :>)

Wouter Addink: @maggie you could tell them that without a globally unique and resolvable identifier, when you publish about the samples used then another researcher can probably not find back the samples to analyse them again or even find proof that they existed. Also new findings from other researchers based on the same samples cannot be linked

Sarah Ramdeen: Right, we need to show the value added benefit for taking the time to do this work instead of what they see as their core role of conducting science. I think Esther is addressing this well now :)

Anusha: Yes, global unique identifiers also ensure the work's legit and accountable by all

Erin Robinson & Ted Habermann: In many cases, making it easy for researchers means making it harder for users. The goal has to be balancing the hardship across both groups.

Maggie Hellström: THANKS to all who contributed to address my question! I'll report back at the next plenary... ;-)

Lesley Wyborn: (link to RDA poster on 'Ring' of vocabularies)
https://poster-rda2020.streampoint.com/index.aspx#&&8zxvzCKU9BKlUA0PaBVptsrQ1Lfdng9slsgOwGF0P46nxoF2oD52vKBuTq+sJSlO2lJN7MZWFHYMJx4dDijdm49cqRW6OvQ+wbwkgcKzMz+GDHWW2I+gBRAe/3/Vn8W+0NL2akgGNuvG1WA2nm4IDh4klBj4zaXU8ErQPGnHIVUw739Q

Catherine Patterson: +1 to @Erin&Ted.... I'm coming at this from the 'developing a system for the user' side, and am super interested in what the minimum metadata etc. for physical samples is to help balance out that burden.

Ted Habermann: I generally try avoid using 'minimum' and 'metadata' in the same sentence...

Experience clearly indicates that if you specify minimum metadata that is all you get.makes more sense to describe these metadata using the use case they support. Minimum metadata is almost always discovery metadata. Let's just call it that.

Alex Hardisty: @Erin and Ted: Sufficient is probably a better term than minimum. What is the smallest set of metadata that is sufficient for the purpose?

Maggie Hellström: @Ted: of course quantity is just one aspect; making sure that the metadata is of high quality is even more difficult. In my experience, "cut & paste" should be banned in the context of (manual) metadata entry! (Also goes for DMPs...)

Lesley Wyborn: This is a link to that ESIP samples session: https://2020esipsummermeeting.sched.com/event/cIvR/and-samples-a-proposed-esip-cluster-for-the-physical-samples-community If you scroll down you can access the notes from this ESIP session

Ted Habermann @Alex - yes, sufficient for discovery in this case.
@Maggie +1

Anusha: +1 @Maggie
Alex Hardisty: @Erin and Ted: which is different from what is sufficient for A, I, R !

Ted Habermann: @Alex +2!

Sarah Ramdeen: Vocabularies and ontologies list https://docs.google.com/spreadsheets/d/1IJ3nXherJWW5wmpyHYkxQ_UHUJaSeney1fGHT0SasM0/edit?usp=sharing

Claudia M. Bauzer Medeiros: I know that Humanities and social sciences have more than 1 thousand ontologies. Why not the physical world?

Ted Habermann: Sufficient metadata for discovery. or, including PID's: Sufficient metadata for discovery and connection. Yes! Thanks Leslie

Kirsten Elger: @Lesley: I am interested in joining

Anusha: +1 , I am interested in joining

Sarah Ramdeen: @Claudia I think it is OK to have many, just understanding how they relate, which you are using when you use one, and when you are not , etc. is important.

Claudia M. Bauzer Medeiros: The practical approach is - the sufficient is insufficient for specialized knowledge domains. So, the most that can be done is to specialize the sufficient into

a few additional fields. And I suspect we will end up with 8 or 9 metadata fields and the rest will be specialized.

Neil Davies: +1 (we have an RCN proposal pending with NSF that I think would help here too - linked to iSamples project)

Maggie Hellström: What Lesley just said about some metadata (like exact locations) being sensitive, I think it is important to recognize that the "as open as possible, as closed as necessary" statement applies equally to both data and metadata, and that technological implementations of e.g. catalogues must be supported by adequate AAI solutions.

Anusha: +1 @Maggie … absolutely , data should be open as possible and necessary for the field

Claudia M. Bauzer Medeiros: recent work on analysis of hundreds of metadata standards find out that the only commonalities are 8 to 10 DC fields. Because we cannot generalize specialized fields.

Ted Habermann: similar work reported on at Metadata2020:
http://www.metadata2020.org/blog/2018-02-02-can-we-agree/

Anusha: Maybe some areas would require more disclosure as opposed to others , so might not warrant for the same

Ted Habermann: Document early - document often!

Maggie Hellström: @Anusha: yes, absolutely - but of course "all" metadata must be sustainably & securely stored somewhere, so that it can be revealed to e.g. selected reviewers, should there arise questions about the quality, scientific relevance, ethics etc.

Markus Kubin: @Claudia: that sounds interesting. Could you please refer to a publication of this review work?

Anusha: True @Maggie , the data , when placed securely would also ensure the data doesn't fall into any unwanted viewers

Claudia M. Bauzer Medeiros: Hi Markus, send me your email cmbm@unicamp.br and I will forward what I found, unfortunately some is in Portuguese. I work a lot in helping researchers from many fields, and if we don't want to go crazy we have to keep simple and short.

Lesley Wyborn: PIDApalooza also enables us to get IGSN into the global PID Ecosystem

Neil Davies: sufficient vs. minimum debate. In DMPs you need to do sufficient for the scope of your project (F) and do it well (A,I,R) … The latter enables potential expansion of uses and

reuse, but this also sometimes goes beyond ‘minimum/sufficient’. there should be advice (automated hopefully) on how you can maximize downstream (re)use for questions you have not imagined in fields you are not a specialist. That means how you can integrate metadata that keeps options open across as many fields as possible recognizing there is probably some cost so weighing up cost/benefit is part of it.

Val Stanley: Another avenue - training for early career/grad student researchers.

Neil Davies: ethical aspects. See CARE principles. see also BioCultural Labels and Notices (and Traditional Knowledge Labels)

Wouter Addink: CARE principles: https://www.gida-global.org/care

Anusha: +1 @val stanley … I think such kind of training would ensure that the future generation of researchers is well versed from the beginning.

Maggie Hellström: @Wouter: exactly!

Neil Davies: Biocultural labels see https://www.enrich-hub.org/bc-labels

Val Stanley: Thanks @anusha! I agree, make it a part of research workflows from the start.

Maggie Hellström: @Neil: thanks for the link, very interesting!

Anusha: Thanks @Neil @val , exactly. This would enable more organised flow of data , across the different fields.

Anne Cambon-Thomsen: citing the physical collection of samples you used can be standardised, making a difference (automatically traceable) between sample collections you cite but not use and samples you actually use; see: 15.3. BRAVO E., CALZOLARI A, DE CASTRO P, MABILE L, NAPOLITANI F, ROSSI A M, CAMBON-THOMSEN A, Developing a guideline for a standardized citation of bioresources in journal articles (CoBRA). BMC Medicine 2015, 13:33 (17 February 2015)

Val Stanley: I'm envisioning a workflow/decision tree diagram for researchers. Visuals help

Lesley Wyborn: At least we know now that this RDA conference system does not cut you off right at the end of the session!

Anusha: They definitely do help , as they say better to start easy

Kirsten Elger: good to know, Lesley :-)

Sarah Ramdeen: Please join the group on our RDA page, https://www.rd-alliance.org/groups/physical-samples-and-collections-research-data-ecosystem-ig In order to get emails/communications

Anusha: @lesley :) good one

Neil Davies: workflow/decision tree diagram — this was something came up in IOC-UNESCO Ocean Best Practices System workshop

Anne Cambon-Thomsen: the sHARC interest group is both about data and physical samples for recognition of the work that leads to sharing