

00:40:45 Maggie Hellström: @Deb: don't you actually mean "machine-actionable" or "machine-interpretable", rather than just "machine-readable"!?

00:40:59 Shelley Stall: @Maggie - yes, that's right

00:42:44 Deb Agarwal: @Maggie - yes, sorry - I need to be better about being precise with language

00:43:28 Frederick Bingham: What is the difference between a crate and a reliquary?

00:44:26 Deb Agarwal: @Frederick - the reliquary is a concept. These tasks are describing actual implementations of things that might make up or be a reliquary

00:44:41 Deb Agarwal: tasks -> talks

00:45:58 Hans Pfeiffenberger: „Any URI-addressable content“ - so no persistent identifier required for elements of RO-Crate?

00:47:08 Stian Soiland-Reyes: Hans, right, not required, but recommended if it is following the FAIR Digital Object profile. L

00:47:31 Raul Palma: Indeed, totally agreed with Stian

00:48:36 Stian Soiland-Reyes: Starting point is “I have some file on my desktop” end point is “I gave DOIs to every element”. Usually truth is somewhere in between but RO-Crate want to support users at both ends.

00:52:28 Carole Goble: I think the questions were answered by Stian....I will look through

00:53:01 Stian Soiland-Reyes: Carole's slides: <https://drive.google.com/file/d/1-6VhnHpz13kMPgDA08Ss166NYi5OHvTY/view?usp=sharing>. (Linked from agenda)

00:55:40 Shelley Stall: Burton, Adrian, Fenner, Martin, Haak, Wouter, & Manghi, Paolo. (2017, November 21). Scholix Metadata Schema for Exchange of Scholarly Communication Links (Version v3). Zenodo. <http://doi.org/10.5281/zenodo.1120265>

00:56:07 Stian Soiland-Reyes: These link sets remind me of Signposting link sets (due to be RFC) <https://www.ietf.org/archive/id/draft-ietf-httpapi-linkset-02.txt>

00:56:07 Shelley Stall: <http://www.scholix.org>

00:57:20 Elisha Wood-Charlson: Is scholix planning to expand their relationship terms?

00:57:49 Stian Soiland-Reyes: Also <https://www.w3.org/TR/void/#linkset> (Void Linkset) but that is different in that they are mainly about identifier-scheme to identifier-scheme mappings, rather than these relational links.

00:58:53 Elisha Wood-Charlson: @Carole - is RO Crate similar to RAiD? (<https://www.raid.org.au/>) Learned about that recently as well.

00:59:52 Megan Force: For a RO-Crate containing datasets from several different repositories, what is the publishing entity of record for the RO-Crate for citation and indexing purposes?

01:00:13 Stian Soiland-Reyes: Elisha, I think RAID can be used by RO-Crate to set identifiers for things like projects - we have used ROR to identify organisations and ORCID for people, but other things are harder for “lay people” to get persistent identifiers for.

01:00:28 Jerry Carter: As a data provider I want to determine who is using my data in a publication, can I easily obtain a list of publications that reference my data through RO-Crate or Scholex?

01:01:09 Stian Soiland-Reyes: @Megan, you can set yourself as publisher on the overall RO-Crate and give different publishers for each of the existing datasets. The same metadata can be applied at every level - similar for mixed licenses.

01:01:11 Elisha Wood-Charlson: @Stian - thanks!

01:02:02 Shelley Stall: Paolo, can you please share your slides?

01:02:45 Shelley Stall: You can email them to me for sharing in the Zenodo deposit we'll create for the workshop.

01:02:58 Carole Goble: Thanks For fielding the questions Stian while I have a coughing fit!

01:03:07 Shelley Stall: Thank you so much for the presentation!

01:04:04 Hans Pfeiffenberger: +1 Jerry! This question has indeed not been addressed by the 2 previous talks! More „important“ for some: Will datasets (wrapped in a crate) contribute to one's h-index?

01:05:01 Paolo Manghi: slides available from here:

<https://drive.google.com/file/d/1t5YZvTlybCoHWQtJHq056LdaVi5U1iT8/view?usp=sharing>

01:05:45 Jerry Carter: @Hans, thanks for getting to the core of my question

01:05:45 Shelley Stall: Thank you

01:05:50 Stian Soiland-Reyes: Please let's bury the h-index in the bottom of the Pacific Ocean. But transitive citation credit I think is one motivation for citation reliquaries – and likewise RO-Crate's ability to give attribution on each sub-element is a way to give better nested credit even for “small” contributions that may not give traditional article authorship but are nevertheless important.

01:06:46 Stian Soiland-Reyes: Level 0 is for authors to cite data where it already have DOIs but they just didn't bother looking it up or were afraid of page limits etc.

01:07:10 Paolo Manghi: @Hans: the datasets, as well as the links (citations), will be made available, so anybody can then define whatever Citation Index wants to build. Personally, I believe in multiple indicators, which may vary from community to community. Scientific value is a complex matter, citations are too simple to capture it alone.

01:08:19 Stian Soiland-Reyes: Q Ugis: Can usage of compact identifiers like RRID be detected and indexed in a similar way as references (which are submitted to crossref/datacite), or do we still have to text-mine PDFs to find their usage.?

01:08:38 Maggie Hellström: Getting the "credit archaeology" right (i.e. opening even recursive crate references to list all related creators) can be very important when older legacy datasets are referenced - such datasets may be associated with licenses requiring specific types of attribution for reuse etc.

01:09:32 Stian Soiland-Reyes: @Maggie +1 - I've got one of those datasets I'm trying to resurrect.. it's only from 2014 but already half the links are dead :/ – luckily all the metadata is captured.

01:09:36 Hans Pfeiffenberger: Agree with Stian and Paolo on h-index being very problematic (and should be abolished) But the reality is, unfortunately, different. (Most people asking for the h-index of a person have probably not read the article by Hirsch ...)

01:09:43 Deb Agarwal: @Maggie - great point!

01:10:39 Carole Goble: omicsdi.org/

01:11:22 Raul Palma: +1 @Stian, also given that an RO-crate is in fact the container including the links between different artefacts (e.g., links between datasets and articles), obtaining such information would be quite straightforward. The main point would be of course to collect or extract such links and make them as explicit metadata. Services from OpenAire are already doing something in this direction in their Knowledge Graph, but there are also other sources that can be used to collect such data. In Reliance project we will be exploiting some of these for the generation of RO-crates

01:11:41 Carole Goble: +1 Raul

01:12:09 Paolo Manghi: apologies, new version of the slides:
<https://drive.google.com/file/d/1aXZkixdtaLXOFDam8UCuB9osDzWzO-b5/view?usp=sharing>

01:12:16 Paolo Manghi: forget the previous one

01:12:51 Shelley Stall: Thank you Paolo. I'll update the link on the agenda.

01:12:58 Maggie Hellström: I really like the RO-Crate and Scholix approaches that "preserve" the metadata of all individual digital objects that are included. Previously, during some data mining and/or metadata harvesting operations by "aggregator services" could corrupt the creator info by erroneously replacing the original creator list with the name of the repository from which catalogue the metadata was obtained! (Still happens occasionally, unfortunately.)

01:13:13 Kiri Wagstaff: <https://www.gbif.org/>

01:13:39 Stian Soiland-Reyes: @Raul, do you think we need to "archive" metadata on usage, just like web.archive.org? Especially institutional repositories may change over the years. Lifting them into RO-Crate is one way to kind of snapshot and unify metadata from multiple repositories - but changing vocabularies may lose some details – perhaps a service (or an RO-Crate?) could also archive the original API responses.

01:14:10 Ugis Sarkans: @Stian - text mining is done by our colleagues @EuropePMC (europepmc.org) - and the results captured in BioStudies records.

01:15:18 Stian Soiland-Reyes: @Ugis, ah, that is great, so presumably all of EuropePMC is theoretically covered then. Is it possible to contribute BioStudies records in other ways..? Say from a conference.

01:16:15 Ugis Sarkans: @Stian - yes, there are in principle 2 ways how BioStudies acquires data - from EuropePMC post-publication, and via direct submissions pre-publication.

01:17:51 Stian Soiland-Reyes: Daniel, this use of derived dataset is great - it's providing provenance of data selection but also showing more details of their data origin – for transitive data citation.

01:17:56 Shelley Stall: Thank you Daniel!

01:18:08 Francisco Manuel Sánchez Cano: Thank you Daniel

01:19:27 Kiri Wagstaff: <https://www.pangaea.de/>

01:19:39 Raul Palma: @Stian, I think that's another of the great benefits of research objects, actually being able to track the evolution and lifecycle of the associated research. We can generate RO-crate snapshots which can also include in some form the information of the responses from services used at that particular time (in addition of the information of the service itself).

01:21:07 Paolo Manghi: to be fair, I had 10 minutes!!!! :)

Data Citation Community of Practice Workshop for Data Citation
Chat from 8 June 2021

01:21:28 Deb Agarwal: @Paolo - I did give you 10 min :)

01:21:48 Stian Soiland-Reyes: Like that GBIF service could do a metadata snapshot when a derived dataset is made. Perhaps store it as an RO-Crate. Hopefully you won't need it, but with so many different repos combined it is bound that some of them will fall over over time. – today Fastly went down and crashed half of the Web!

01:23:14 Paolo Manghi: damn...

01:23:24 Paolo Manghi: I wanted that price

01:23:43 Deb Agarwal: :)

01:24:54 Paolo Manghi: @Uwe: the “data collection” object (“partOf” relationship), as well as a “biostudy”, are examples of reliquaries in the research data domain

01:27:34 Stian Soiland-Reyes: Q Uwe: Can a bundled publication be made by someone who does not “own” the children? Like a review selects a bunch of data.

01:28:53 Stian Soiland-Reyes: ... who becomes the “authors” in that case, are they promoted from their content, or is the author who did the selection?

01:30:15 Shelley Stall: Thank you Uwe!

01:32:08 Shelley Stall: LOVE the Reliquary!!

01:32:33 Stian Soiland-Reyes: Pingbacks are like putting those little candles

01:33:39 Uwe Schindler: @stian: No the bundled publication has a string relationship parent -> child and vice versa. If you would put datasets of somebody else into your own parent, the child would refer to you and you would take "ownership"

01:33:58 Paolo Manghi: need to go! thanks for the discussion, extremely interesting

01:34:00 Uwe Schindler: so for that case (other work put together), you would use an editorial

01:34:19 Deb Agarwal: @Paolo, thank you for the excellent talk!

01:34:30 Paolo Manghi: long :)

01:34:47 Paolo Manghi: talk soon!

01:35:46 Stian Soiland-Reyes: do we need to specify the reliquary as a special data cite type, or would Collection suffice? I guess it's hard enough already to get journals to provide this subtypes as they are just sending DOIs blindly to crossref.

01:36:10 Stian Soiland-Reyes: (Easier to achieve with data-to-data citations made in platforms)

01:39:23 Stian Soiland-Reyes: Very important point there about versioning - Reliquaries may be “live” objects or generated dynamically (e.g. “All GBIF repository entries from New Zealand”) and need snapshot identifiers.

01:40:58 Matthew Cannon: Hi Stian - what would you want journals to do here? Is this about the metadata to be fed to CrossRef?

01:43:27 Maggie Hellström: why share data? well, don't discount the importance of that "warm and fuzzy feeling";-)

01:43:38 Shelley Stall: @Matt and @Stian - We're trying to navigate around asking journals to do something new. Citing a reliquary as a collection is the goal.

01:44:41 Deb Agarwal: Not all data sharing is in support of a science result but instead is data collected to enable other research.

01:44:58 Carole Goble: @Maggie - ah, the "love" incentive. as opposed to "fame" and "money"

01:45:22 Justin Buck: Trying to link all this back to the original motivation, which is how do we handle the citation of subsets of the data unpinning DOIs or datasets where there are 100,000s of sub entities in papers. DOI may not be feasible for all entities we need to be able to cite handles and ePIC with DOIs. The DO crate seems to give us the citation ability, am struggling to align this with the other approaches presented though, does this make it a fundamentally new community requirement?

01:45:29 Deb Agarwal: @Maggie - and for the good of your community.

01:45:40 Matthew Cannon: Thanks @Shelley - would we need to be able to tag a citation as being to a collection, rather than a single data set?

01:45:41 Kirsten Elger: "Salami slicing" absolutely, good Point, Martin

01:47:01 Carole Goble: moving towards living papers that can have appropriate credit measures may take the wind out of Salami slicing

01:47:03 Stian Soiland-Reyes: @Shelley, I think we need to go ahead of the traditional journals. There are some that may join early because there are enthusiasts - e.g. <https://jcheminf.biomedcentral.com/> has editor Egon Willighagen which is strong on FAIR - but still they may be trapped in publication workflows mandated by Springer Nature. So start first on our own (e.g. Zenodo let you provide any type of data cite relationship) and then perhaps invite in some ahead-of-the-curve journals.

01:47:15 Uwe Schindler: Yes!

01:47:30 Carole Goble: Great wrap up Martin!

01:48:32 Elisha Wood-Charlson: +1 Martin on people still only citing journal articles. We need the data landing pages to be more dynamic, similar to what GBIF is maybe doing with derived datasets, so we can adapt and make sure follow in citations are tracked.

01:49:58 Maggie Hellström: @Martin (and others): is MakeDataCount also considering (in the long term) to capture statistics like the number of times that a DOI is contained in published workflows and similar processing-related records (like provenance tracing)? I think that with increased machine actionability and/or (semi-)automated workflows, a lot of referencing information becomes buried inside of metadata...

01:51:31 Martin Fenner: @Maggie we MDC are currently focusing on data citation and data usage stats, but are of course interested in other metrics down the road.

01:52:42 Stian Soiland-Reyes: @Matthew, see for instance the journal articles in <https://doi.pangaea.de/10.1594/PANGAEA.887388> - most of them are not open access so I can't even access them to see what data references they may have. Only if it's submitted to CrossRef/DataRef as DOI references will they become of the DataCite knowledge graph and programmatically accessible. For instance in <https://api.crossref.org/v1/works/10.1016/j.geoderma.2011.09.012> which I found from <https://doi.pangaea.de/10.1594/PANGAEA.887388> there is not a single citation back to PANGAEA, only journal references. Often there will be a duality - a journal article about a dataset for instance - but journal articles tend to cite other journal articles as preference.

01:52:49 Maggie Hellström: @Martin: great, hope you get the possibility (funding etc) to pursue also other sources of usage statistics!

Data Citation Community of Practice Workshop for Data Citation
Chat from 8 June 2021

01:52:59 Elisha Wood-Charlson: Have we agreed upon a "relation type" for citing data set components of a reliquary? If we were consistent in relation terms, it could help with metrics across DOI trackers.

01:55:37 Martin Fenner: @Maggie the most obvious metric to add (as the tooling is already in place) is social media metrics. Which is of course a very different beast, about attention rather than impact. But it could address other relevant use cases.

01:58:56 Maggie Hellström: @Martin: yes, in some disciplines social media mentions are very common, and there altmetrics are important. What I was thinking of specifically was trying to crawl/harvest e.g. provenance-related metadata fields in dataset cataloguing information. This is of course challenging as this probably won't be stored at e.g. DataCite, but at repository level - necessitating more calls back and forth + a fair amount of interpretation power to properly unlock the information !

01:59:49 Stian Soiland-Reyes: Who is building the reliquary? Will it be a set of recommendations to follow (e.g. Datacite+RO-Crate), or do we need a new type of index/repository to develop/maintain/host/forget ?

02:00:09 Stian Soiland-Reyes: +100 use simplest path

02:00:18 Carole Goble: Simplest Path!!!! yes!!

02:00:36 Maggie Hellström: Sounds almost Zen!

02:01:23 Carole Goble: And a pilot that works in practice

02:01:24 Stian Soiland-Reyes: Like <https://signposting.org/FAIR/> is just 7 existing link relations put into system. No database or anything, the Web can already do it.

02:01:35 Shelley Stall: @Uwe, for the four types of collections at Pangaea - when you register the collection - are you using the DOI Collection schema?

02:01:57 Shelley Stall: The DataCite DOI Collection?

02:02:15 Uwe Schindler: yes it is the type "collection" in datacite

02:02:21 Shelley Stall: ok great.

02:02:22 Justin Buck: Facilitating the use cases is central to all of this, filling the gaps in citation that we cannot achieve with current approaches

02:03:08 Howard Ratner: Minimal Viable Product = Keep it simple

02:03:14 Uwe Schindler: we refer from childs to parent for dependent stuff "IN: XXXX"

02:03:43 Martin Fenner: One example of keeping it simple is build on existing citation practices. So that we not first have to redefine how reference lists look like.

02:04:00 Carole Goble: right - its the fourth incentive - NUDGE

02:04:00 Uwe Schindler: recently we also started to list all childs as "hasPart" inside the datacite collection

02:04:04 Stian Soiland-Reyes: Use - you have already made reliquary then! Assuming the hasPart is also propagated to Datacite, with it looks like doing already..?

02:04:06 Elisha Wood-Charlson: Thank you!

02:04:08 Howard Ratner: Thanks. Good talks today

02:04:13 Joerg Heber: Thank you!

02:04:14 Elisha Wood-Charlson: @uwe - good to know, thanks

02:04:17 Kiri Wagstaff: Thanks all!

Data Citation Community of Practice Workshop for Data Citation
Chat from 8 June 2021

02:04:20 Martin Fenner: Thank you!
02:04:22 Reyna Jenkyns: Thanks all!
02:04:23 James Ayliffe: thank you
02:04:29 Francisco Manuel Sánchez Cano: Thank you!!
02:04:31 Oscar Corcho: Thanks, bye
02:04:33 Julie Dionne (SLGO): Thank you all !!
02:04:38 Uwe Schindler: yes we have hasPart/isPartOf in Datacite metadata
02:04:45 Oscar Corcho: Get better, Carole!!