

2019-12-16 Technical WG Agenda and Notes

Date

16 Dec 2019

Attendees

- John Kunze
- Sheila Morrissey
- Karen Hanson
- Greg Janée
- Bertrand Caron

Discussion items

Time	Item	Who	Notes
	updated timeline		timeline approved
	proposed ARK spec cleanup		no objections to broad cleanup of spec, even if they generate noisy diffs
	ongoing ?info inflection proposal discussion <ul style="list-style-type: none">• json vs json-ld (rdf)<ul style="list-style-type: none">◦ eg, semantic web issues• accommodating descriptive hierarchy<ul style="list-style-type: none">◦ finding aid >◦ box >◦ folder >◦ letter◦ dataset >◦ row >◦ cell◦ article landing page >◦ pdf file		KH: article seems to still endorse RDF a bit; there's value in aligning with DataCite BC: going from XML to JSONLD logic isn't just a question of applying an xslt; we spent several years to come up with xml / mets data model; Gautier Poupeau (author of blob post) designed the dig repo at BnF BC: agree that for people who have XML-based metadata, it's easy to express it as json; but a json-ld expression requires revising your whole data model and so is much harder and very complicated GJ: could you explain why it's not just a simple mapping? BC: because you have to express everything in triples with entity-relationships JK: why isn't it just a set of key/value pairs? JK: should we invite Justin L to future subgroup meeting to help us understand? Bertrand, do you want to join as well? KH: yes, that would find that helpful BC: yes, I would GJ: see https://blog.datacite.org/schema-org-register-dois/ BC: for persistence, we'll be defining our own vocabulary BC: if we went with Dublin Core entity-relationship model, that's simple enough; but if they have to invent their own model, that's much harder; by using linked data, one says that all my data is expressed as triples KH: see also https://github.com/rmap-project/rmap-documentation/blob/master/guides/useful-ontologies.md BC: nice to have a position about how we express metadata, but we have to go further and say, at least for basic metadata, what the recommended model is JK: *model* is important – we need to understand this better as a group KH: fortunately, vocabulary choice is independent of syntax BC: at BnF we chose XML and snippets; we could recommend a set of optional return formats; users at many French archives will have a hard time putting out something different from what they already do; this is an argument for giving people options for returned data KH: make it lightweight, but steer people in certain directions KH: See the Portland Common Data model SM: Portico would like a conceptual model for a landing page as an archival unit; it would be an enormous simplification to be able to have a landing page model for this

Notes

In follow-up email from KH:

Reflecting on the discussion, I think it may have gotten lost that when we were discussing JSON-LD a few weeks ago, I believe it was imagined as part of a series of recommendations, not as requirement. If that is still the case, then perhaps it's not necessary to add a meeting to discuss the pros and cons of linked data further, since implementation would be looser and JSON-LD would be encouraged but not required. I think it should be perfectly fine for Portico to produce JSON-LD and BnF to produce XML.

I'm wondering if the guidelines might go something along these lines:

1. At minimum, [?info](#) must resolve to a human readable landing page, and should provide a gateway to machine-readable metadata
2. It is strongly recommended that meta tags with [something like] DC are implemented (since they are simple html, and all orgs should be able to do something with those).
3. Secondary to this, we encourage but don't require JSON-LD with [schema.org](#) (and/or DC) on the [?info](#) page (in alignment with the [JDDCP recommendations in the Scientific Data article](#)).
4. Finally, regardless of whether JSON-LD is implemented, we encourage organizations to use whatever data format[s] is appropriate in their context as the machine-readable data version of [?info](#), but encourage that:
 - a. Organizations include DC metadata in this where possible
 - b. Organizations utilize either content negotiation or add “&format=[json|xml|etc]” property to deal with alternative formats.

This is just a rough example, but maybe something like this approach might work to give a little structure but plenty of flexibility.

Action items

- make sure ?info leaves options for people, but steers them towards json
- add Bertrand to subgroup