WG Data Versioning - RDA Plenary 12 Notes

Session page: https://www.rd-alliance.org/wg-data-versioning-rda-12th-plenary-meeting

Group Page: https://www.rd-alliance.org/groups/data-versioning-wg
Please sign up to group!!!

Participants:

No.	Name	Organisation	Country	email
1	Lesley Wyborn	National Computational Infrastructure, Australian Research Data Commons	Australia	lesley.wyborn@anu.edu.au
2	Afe Idowu			idowu.afe@fedcs.gov.ng
3	Jens Klump	CSIRO	Australia	jens.klump@csiro.au
4	Robert Downs	CIESIN, Columbia University	USA	rdowns@ciesin.columbia.edu
5	Grace Bowno(?)	Botswana National Library Service	Botswana	tshidg@gmail.com
6	Hildah Bawani	Botswana National Library Service	Botswana	hibawani72@gmail.com
7	Keith Russell	Australian Research Data Commons	Australia	keith.russell@ardc.edu.au
8	David Lescinsky	Geoscience Australia	Australia	david.lescinsky@ga.gov.au
9	Ralph Muller-Pfefferkorn	Technical University Dresden	Germany	Ralph Muller-Pfefferkorn@tu.dresden.de (?)
10	Martina Stockhause	DKRZ	Germany	stockhause@dkrz.de
11	Lady O. Tausane	Botswana Police Service	Botswana	ltausane@yahoo.com

12	Joao Rocha da Silva (?)	INESC TECI FEUP	Portugal	joaorosilva@gmail.com (?)
13	Scott Edmunds	GigaScience/BGI Hong Kong	Hong Kong	scott@gigasceicejournal.com (?)
14	Tukka Nappila (?)	University of Tampere	Finland	tukka.nappila@ut.fi (?)
15	Anca Hienola	Finnish Meteorological Institute	Finland	anca.hienola@fmi.fi (?)
16	Mable Lettaditswe (?)	Botswana Police Service	Botswana	mablet@gmail.com (?)
17	Lynette Mmoi (?)	Botswana Police Service	Botswana	lynettemmoi@yahoo.com
18	Martina Trognitz	ACDH, OAW	Austria	martina.trognitz@oeaw.ac.at
19	Norbert Kihuha	Kemri-Wellcome Trust Research Programme	Kenya	nkihuha@gmail.com
20	Thu-Mai Christian	Odum Institute	USA	thumai@email.unc.edu
21	James Savage	University of Cambridge	UK	jlss9@cam.ac.uk (?)
22	Andreas Rauber	TU Wien	Austria	rauber@ifs.tuwien.ac.at
23	Tomasz Miksa	TU Wien & SBA Research	Austria	Miusa@ifs.tuwien.ac.at
24	Roman Gerlach	Friedrich Schiller University Jena	Germany	???
25	Rosie Higman	University of Manchester/University of Sheffield	UK	
26	Reyna Jenkyns	Ocean Networks Canada	Canada	reyna@uvic.ca
27	Moses T Mafatlhe	Botswana National Archives and Records Services	Botswana	mmafatlhe@gmail.com (??) mMafatlhe@gov.bw (??)
28	Dave Connell	Australian Antarctic Division	Australia	dave.connell@aad.gov.au
29	Neo Moweu (??)	Botswana National Library Service	Botswana	neomosweu@gov.bw (??)

Meeting Slides:
https://docs.google.com/presentation/d/1e818TKNvjXuU354K0jQIUhWhulOv8xg_w57dFnSefQY/edit#slide=id.p

Agenda:

- Introduction
- Recap of Why, How and What of Data Versioning
- Review of use cases, including the W3C Dataset Exchange Use Cases and Requirements
- Work plan for RDA Data Versioning WG
- Engagement with other RDA and external groups
- Outline of white paper on data versioning practices
- Scheduling of online meetings up to Plenary 13

Meeting Notes:

After presenting the agenda and starting with some summarising slides (see slides above), Lesley Wyborn presented some issues that may occur in data versioning (see slides above).

Simon Cox gave an update on the activities at W3C, where this WG provided use cases to:
- There is a project on a dataset catalogue vocabulary (dcat), which aims at describing metadata
- Some of the use cases collected by RDA were provided to the Dataset Exchange Working Group of W3C, and thus could contribute some unconsidered cases and present new details that had not been covered yet; esp. regarding relationships between original and derivative or new versions of data
- It should be noted that the W3C WG is a busy group, where the RDA WG is not on top of the stack (contact and exchange thus not as active as one may wish)
- W3C in general tries to apply W3C technologies for their solutions (-> provenance ontology; used for RDF)
- Lesley pointed out, that RDA is not a standards body. This RDA WG feeding its outputs up through the W3C is a first. That is that the RDA versioning group is effectively feeding some of its use cases into a standards body as a way of influencing standards development.
- Jens points out that feeding into W3C can provide a large community adoption

Main topic of the meeting was: "What should be in the WG Data Versioning White Paper?"
Suggestions by the audience:

- A general introduction
- Practical guidance
- Versioning patterns are relevant, but should only be included as a section with a description
- Elaborate on difference between version and derivation
- For publishers: when should the catalogue record be changed? When should a snapshot record be produced?
- Reproducibility
- DOI with different version numbers (?)
- What changes should be recorded when changing version; How to document changes?
- How to communicate changes; -> Versioning policy
- Versioning policy should state when a new version is set vs. when "manual" versioning is done
- Recommendations for continuous measurements (e.g., if algorithms for derivatives change) / open time series identification
- Time stamping; window parameters
- Versioning patterns and the data life cycle
- When can a dataset seen as final in the archival process?
- A section summarising the use cases
- A section explaining on the terminology would be very helpful

Update on the work plan:
- At Plenary 13 (April 2019, Philadelphia) the white paper should be presented for discussion
- Still new use cases should be collected and comments made on the existing ones
- Patterns should be applied to use cases (?)

Action: Lesley to chase additional case study from Germany