

# WG Data Versioning - RDA Plenary 11 Notes

**Session page:** <https://rd-alliance.org/wg-data-versioning-rda-11th-plenary-meeting>

**Group Page:** <https://www.rd-alliance.org/groups/data-versioning-wg>

Please sign up to group!!!

## Participants:

No.	Name	Organisation	email	Versioning issues you are concerned with?
1	Lesley Wyborn	NCI, ANDS,	<a href="mailto:lesley.wyborn@anu.edu.au">lesley.wyborn@anu.edu.au</a>	Big data sets that are too big to store multiple versions of, External Versioning
2	Mingfang Wu	ANDS	<a href="mailto:mingfang.wu@ands.org.au">mingfang.wu@ands.org.au</a>	
3	Martina Stockhause	WDCC/DKRZ	<a href="mailto:stockhause@dkrz.de">stockhause@dkrz.de</a>	ESGF versioning in connection with evolving data citation
4	Matt Lightfoot	CCDC	<a href="mailto:lightfoot@ccdc.cam.ac.uk">lightfoot@ccdc.cam.ac.uk</a>	
5	Meghan Tomlin	Ocean Networks Canada	<a href="mailto:mtomlin@oceannetworks.ca">mtomlin@oceannetworks.ca</a>	Dynamic data, how to assign DOI to data that are continually collected in near real-time, very large data sets, how to handle reprocessed data, changes to QA/QC flags, updates to data products etc....
6	Nicole Giliard	ZIB	<a href="mailto:giliard@zib.de">giliard@zib.de</a>	
7	Thilo Paul-Stueve	Kiel University	<a href="mailto:paul-stueve@rz.uni-kiel.de">paul-stueve@rz.uni-kiel.de</a>	
8	Toshihiro Ashino	TOYO Univ.	<a href="mailto:ashino@acm.org">ashino@acm.org</a>	Physical constants, material properties.
9	Lars Vilhuber	Cornell University	<a href="mailto:Lars.vilhuber@cornell.edu">Lars.vilhuber@cornell.edu</a>	
10	Barton Satchwill	Cybera	<a href="mailto:barton.satchwill@cybera.ca">barton.satchwill@cybera.ca</a>	open government data, versioning derivative data

11	Richard Ostler	Rothamsted Research	<a href="mailto:richard.ostler@rothamsted.ac.uk">richard.ostler@rothamsted.ac.uk</a>	Low throughput time-series data. Dataset aggregations. Re-presentation of data either in different data format or visualisations
12	Stefanie Schumacher	AWI PANGAEA	<a href="mailto:Stefanie.schumacher@awi.de">Stefanie.schumacher@awi.de</a>	
13	Nigel Rees	NCI	<a href="mailto:nigel.rees@anu.edu.au">nigel.rees@anu.edu.au</a>	
14	Carsten Schirnack	GEOMAR	cschirnack@geomar.de	Versioning of research data along with code (execution) and Docs for full(?) documentation of data provenance (à la Jupyter Notebooks); => Git for Research
15	Rolf Andersen	UiT The Arctic University of Norway	Rolf.andersen@uit.no	
16	Tobias Weber	Leibniz Supercomputing Centre	weber@lrz.de	
17	Chris Pankratz	University of Colorado/ LASP	pankratz@colorado.edu	
18	Roy Dragseth	UiT the Arctic University of Norway	roy.dragseth@uit.no	
19	Boris Radosavljevic	GFZ German Research Centre for Geosciences	Boris.radosavljevic@gmail.com	
20	Christian Loeschen	Technische Universität Dresden	Christian.Loeschen@tu-dresden.de	
21	Guillaume Brissebrat	CNRS / OMP	Guillaume.Brissebrat@obs-mip.fr	
22	Damien Boulanger	CNRS	Damien.Boulanger@obs-mip.fr	
23	Carsten Hoffmann	Leibniz Centre for Agricultural Landscape	hoffman@zalf.de	

		Research (ZALF)		
24	James A.J. Wilson	University College London	j.a.j.wilson@ucl.ac.uk	
25	Katharina Sack	CCCA - Climate Change Centre Austria	katharina.sack@ccca.ac.at	
26	Milan Daneček	CESNET	danecek@cesnet.cz	
27	Katarzyna Biernacka	Humboldt-Universität zu Berlin	biernacka@hu-berlin.de	
28	Tomas Torgrimsby	University of Oslo	<a href="mailto:Tomas.Torgrimsby@ub.no">Tomas.Torgrimsby@ub.no</a> (?)	
29	Ludwig Hulk	Reiner Lemoine Institut	ludwig.huelk@re-institut.de	
30	Andreas Rauber			
31	Ari Asmi			

### Meeting Slides:

[https://docs.google.com/presentation/d/1kzZFrao5B1MseLW4B\\_Geh00GasAwfvvsJb5cVfaAoqY/edit#slide=id.g3580aef542\\_0\\_84](https://docs.google.com/presentation/d/1kzZFrao5B1MseLW4B_Geh00GasAwfvvsJb5cVfaAoqY/edit#slide=id.g3580aef542_0_84)

### 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 12

### Meeting Notes:

Other relevant groups:

Use cases?

EU ICT board

Use Slack

Are the dimensions policy and/or technical -

How frequently do you version

What are the policies for versioning

Can we use GIT

Hard to separate policy vs technical - needs to be in one document

Data Versioning group has some policy work but do not have semantic versioning as part of the core concepts.

Semantic will help across discipline

Are we talking about the big data centres or the long tail

Large data sets are more complex - long tail is a derivative of the big data problem

Web services puts more pressure on the repositories to be able to provide users with details of when they change the version.

P2P versioning system - user has nothing to do with the versioning it is automated - more computer science based.

Disruptive technology?