



Data Versioning WG

RDA Plenary 10 | 21 September 2017

Agenda

1. Introduction
 2. Recap of Why, How and What of Data Versioning
 3. Review of use cases, including the W3C Dataset Exchange Use Cases and Requirements (<https://docs.google.com/document/d/1TfBPIfjTVg0YcFxuW0UzAXPYrRmyZ6PCxtxKx8-uGg>)
 4. Discussion of case statement
(<https://docs.google.com/document/d/1jZoON7biETH46lvoXcyxt0is4qyt0CWGJrVpHTyfyk>)
 5. Workplan for RDA Data Versioning WG
- 

RDA Guidelines: Review Criteria for a Working Group

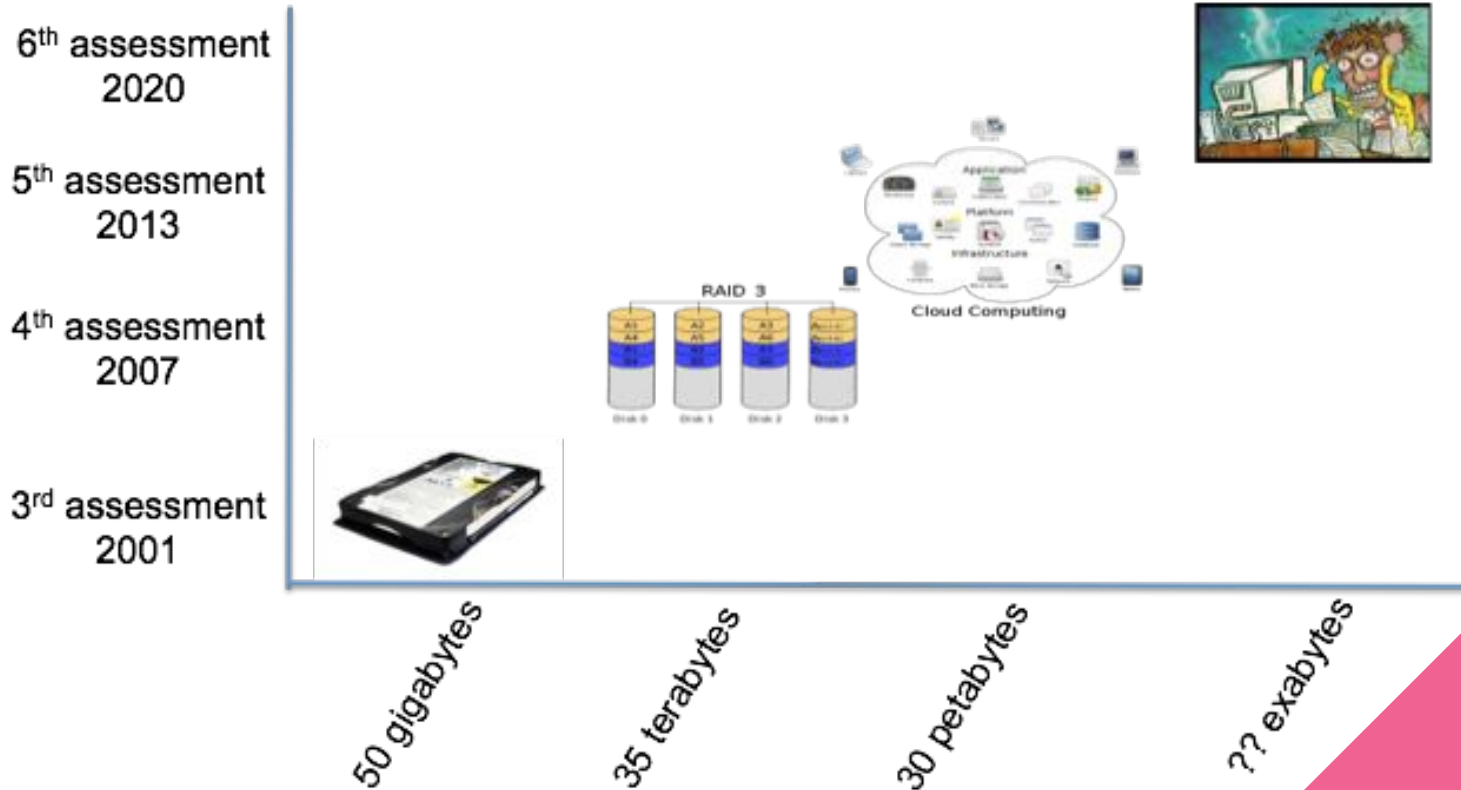
- **Focus:** Are there measurable outcomes?
 - **Impact and Engagement:** Will the outcome(s) of the Working Group be taken up by the intended community? Will the outcome(s) of the Working Group foster data sharing and/or exchange?
 - **Timeframe:** Can the proposed work, outcomes /deliverables, and Action Plan described in the Case Statement be accomplished in 12-18 months?
 - **Scope/Fit:** Is the scope too large for effective progress, too small for an RDA effort, or not appropriate for the RDA?
- 

Recap: The Why, How and What

- Datasets published on the Web may change over time.
- Some datasets are updated on a scheduled basis, and other datasets are changed as improvements in collecting the data make updates worthwhile.
- Others are updated because errors are found
- In order to deal with these changes, new versions of a dataset may be created.
 - What is the significance of the change?
 - Is the new version compatible with the previous version?



The issue will not go away



Use Cases

So far the Versioing IG has compiled a list of use cases on

<https://docs.google.com/document/d/1TfBPIfjTVg0YcFuw0UzAXPYrRmyZ6PCctxKx8-uGg>

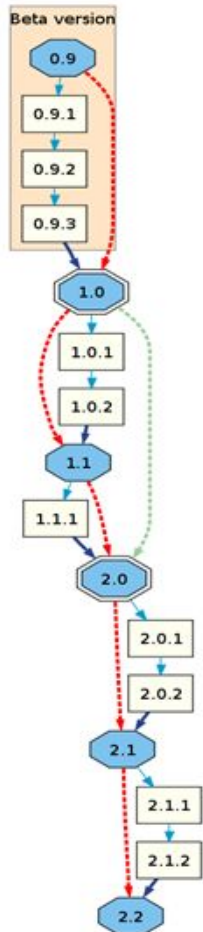
Simon Cox will speak on the W3C Exchange.



Versions vs Builds

- In version control systems, version numbers are incremented sequentially.
- Many software publication systems distinguish between “builds” (sequentially numbered) and “versions” (releases with semantic labelling).
- The number of changes is not a useful indicator of change as even small changes can have significant consequences.





Semantic Versioning

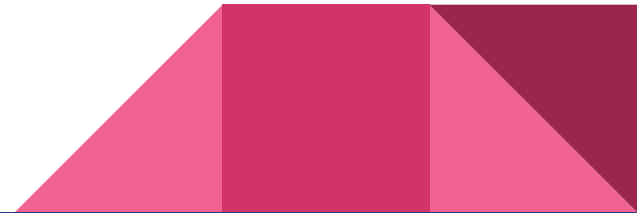
Versioning can carry meaning (Semantic Versioning)

- Sequence-based identifiers
 - Change significance
 - Degree of compatibility
 - Designating development stage

Version n.n.n : (major release).(minor release).(patch)

RDA Case Statement

<https://docs.google.com/document/d/1jZoON7biETH46lvoXcyxt0is4qyt0CWGJrVpHTyfpk>




Versioning and Identifiers

- It is currently being debated in FORCE11 and DataCite about persistently identifying versions:
 - Should versions be reflected in the formatting of DOIs?
 - How should other versions be referred to in DOI metadata?
- This is the next step once we have a common understanding of data versioning.



RDA Guidelines: Case Statement (general)

Working Groups are expected to:

- Develop clear outcomes and put them into action to create tangible progress (see also [Working Group Goals and Outcomes](#) and the [RDA Outputs and Intellectual Property Policy](#)).
 - Work openly and transparently with respect to the community.
 - Document their efforts as they operate.
 - Meet regularly with the RDA to facilitate coordination and communication.
- 

RDA Guidelines: Case Statement (general)

- What is the research case (will the WG produce something useful)?
- What is the business case (will people use it)?
- Is there capacity (are the right people involved to adopt and implement)?

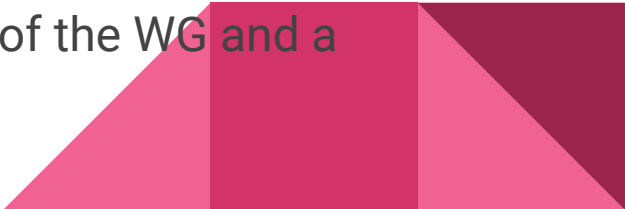


RDA Guidelines: Case Statement (format)

- WG Charter: A concise articulation of what issues the WG will address within a 12-18 month time frame and what its “deliverables” or outcomes will be.
- Value Proposition: A specific description of who will benefit from the adoption or implementation of the WG outcomes and what tangible impacts should result.
- Engagement with existing work in the area: A brief review of related work and plan for engagement with any other activities in the area.



RDA Guidelines: Case Statement (format)

- Work Plan: A specific and detailed description of how the WG will operate including:
 - Adoption Plan: A specific plan for adoption or implementation of the WG outcomes within the organizations and institutions represented by WG members, as well as plans for adoption more broadly within the community. Such adoption or implementation should start within the 12-18 month timeframe before the WG is complete.
 - Initial Membership: A specific list of initial members of the WG and a description of initial leadership of the WG.
- 

Work Plan

Objective of this session is to draft a workplan and a case statement.





Thank you!

See you at RDA P11

RDA Data Versioning WG

Jens Klump, Lesley Wyborn, Ari
Asmi, Robert Downs

[https://www.rd-alliance.org/groups
/data-versioning-wg](https://www.rd-alliance.org/groups/data-versioning-wg)
