

Creating a Categorisation Schema of Digital Research Tools Mapped to the Research Data Lifecycle

Plenary session link:

<https://www.rd-alliance.org/plenaries/creating-a-categorisation-schema-of-digital-research-tools-mapped-to-the-research-data-lifecycle/>

Group(s) name(s) organising the session: RDA-OfR Mapping the landscape of digital research tools WG

Session summary (for Group co-chairs)

Please complete the table below by **7th June, close of business**. The information provided will be used to include your outcomes to the RDA community as a report organised by the Technical Advisory Board.

Summarise the session's key points and discussions in three sentences:
The session provided an overview of three deliverables drafted by the WG: 1) A Harmonised Research Data Lifecycle Model and Crosswalk to Existing Models , 2) A Categorisation Schema of Digital Research Tools ; and, 3) A Prototype of a Visual map of digital research tools landscape . Positive and constructive feedback was solicited from the community about the deliverables and suggestions for next steps were discussed.
Key outcomes/actions/takeaways
1. Unique value of harmonised, curated RDL and tool categories acknowledged. 2. The visual map of the digital research tools landscape (final model) needs to be open-source and dynamic so that it can be further developed and made sustainable for use in the future. 3. There is interest in creating a follow-up Working Group to progress the development of the visual map of the digital research tools landscape prototype. This will be established in the coming month. 4. There is merit in collaborating with the RDA's Global Open Research Commons (GORC) groups to supplement their work on mapping Services and Tools.
Synergies and/or possible collaborations identified with RDA groups and other groups:
• GORC International Model WG

Get involved in [RDA Community](#)

Check out [VP22 programme sessions](#)

This meeting will take place according to the [RDA Code of Conduct](#)

Session 1: Tuesday 14 May, 07:00 - 08:30 UTC

Attendee Check-in from the first session

Please complete this table to indicate your attendance (add rows as needed):

Name	Affiliation	Location	Email/social media
Mingfang Wu	ARDC	Melbourne, Australia	mingfang.wu@ardc.edu.au
Tom Honeyman	ARDC	Sydney, Australia	tom.honeyman@ardc.edu.au
Ville Tenhunen	EGI Foundation	Savonlinna, Finland	ville.tenhunen@egi.eu
Mari Kleemola	Tampere University/FSD	Tampere, Finland	mari.kleemola@tuni.fi
Holly Ranger	The University of Sheffield	UK	h.ranger@sheffield.ac.uk
Sandra Boerman	DeiC	Copenhagen, Denmark	sandra.boerman@deic.dk
Roxanne Wyns	KU Leuven - LIBIS	Leuven, Belgium	Roxanne.Wyns@kuleuven.be
Christian Berger	Philipps-Universität Marburg	Marburg, Germany	berger@hrz.uni-marburg.de
Emmanuel Delage	OPGC - CNRS	Clermont-Fd, France	emmanuel.delage@uca.fr
Andrew Treloar	ARDC	Melbourne, Australia	andrew.treloar@ardc.edu.au
Giulia Caldoni	University of Bologna	Bologna, Italy	giulia.caldoni2@unibo.it
Vashti Galpin	University of Edinburgh	Edinburgh, Scotland	Vashti.Galpin@ed.ac.uk
Heike Görzig	Helmholtz-Zentrum Berlin	Berlin, Germany	heike.goerzig@helmholtz-berlin.de
Jenny O'Neill	HEAnet	Dublin, Ireland	jenny.oneill@heanet.ie

Louise Bezuidenhout	Leiden University	Leiden, Netherlands	l.m.bezuidenhout@cwts.leidenuniv.nl
Roman Gerlach	Friedrich Schiller University	Jena Germany	roman.gerlach@uni-jena.de
Michelle Barker	Research Software Alliance	Australia	michelle@researchsoft.org

Total: 31 participants

Collaborative Session Notes *from the first Session (To be used by participants and chairs during the session).*

Quick links:

- [About the WG](#)
- [Collaborative notes](#)
- [Presentation slides](#)
- [Deliverable 1](#) - Harmonised Research Data Lifecycle Model and Crosswalk to existing models
- [Deliverable 2](#) - Categorisation schema of digital research tools
- Deliverable 3 - Visual map of digital research tools landscape (prototype)

Session Agenda:

Time (min)	Agenda item	Lead
5	Welcome and introduction	Facilitator
5	Presentation: Introduction to the RDA-OfR Mapping the Digital Research Data Infrastructure WG Background and description Aims and objectives	Marcelo Garcia
15	Presentation: Deliverable 1 & Methodology Presentation: Deliverable 2 & Methodology Presentation: Deliverable 3 & Methodology	Adam Vials Moore, Co-chair Rory Macneil, Co-chair
20	Q&A	All
5	Proposed next steps	Rory Macneil
35	General discussion	Facilitator
5	Closure of session	Facilitator

Presentation slides:

P22 Plenary.pptx

Presentation notes

They found 26 research data lifecycles (only 26??)

This is their work plan to date



Deliverable 1: The creation of a harmonised research data lifecycle model and crosswalk to existing models

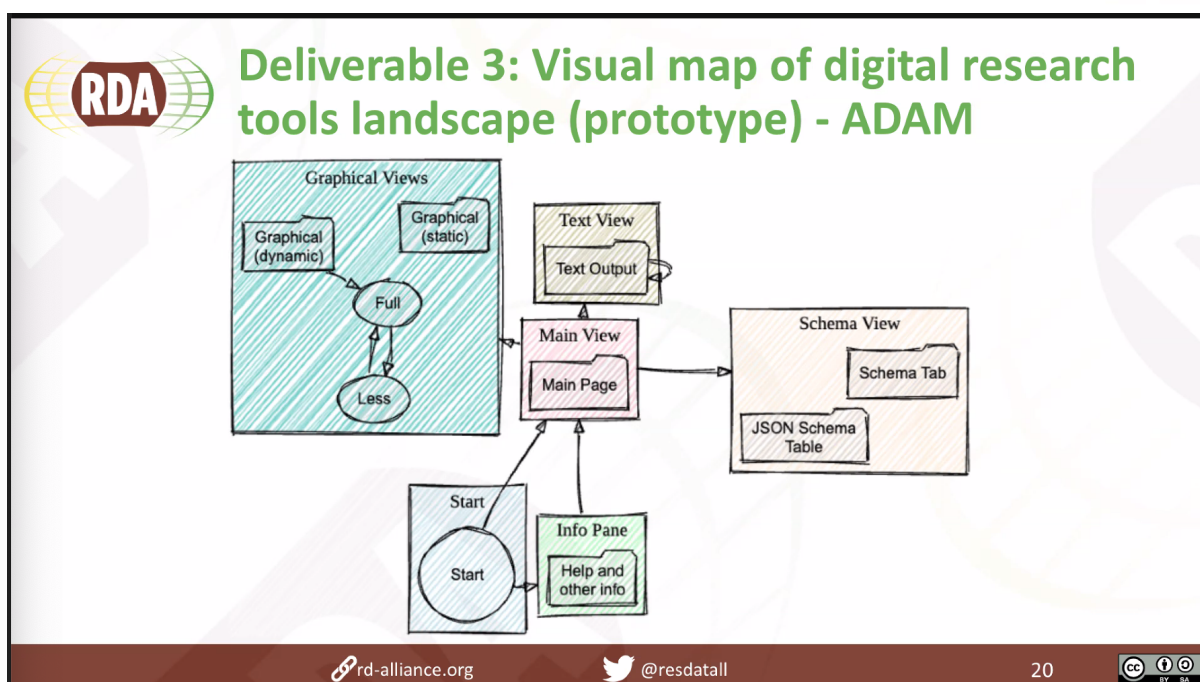
- First step: Set up a Google Sheet! (of course)
- Nice concentric mapping of the top 5 lifecycles they identified onto Maldreth uber-model

Deliverable 2: Categorisation schema of digital research tools

- Nice use of active verbs for each stage of lifecycle
- Just having a list of tools for each stage wasn't enough - needed to categorise tools as well

Deliverable 3: Visual map of digital research tools landscape (prototype)

- "Rather than producing a Google Sheet that we serialise into a report..."
- MaLDReTH prototype: <https://adamvialsmoore-jisc.shinyapps.io/r-proto/>



- rd-alliance.org
- Draft report by end of May 2024

Q&A Notes

Question (Q), Comment (C)

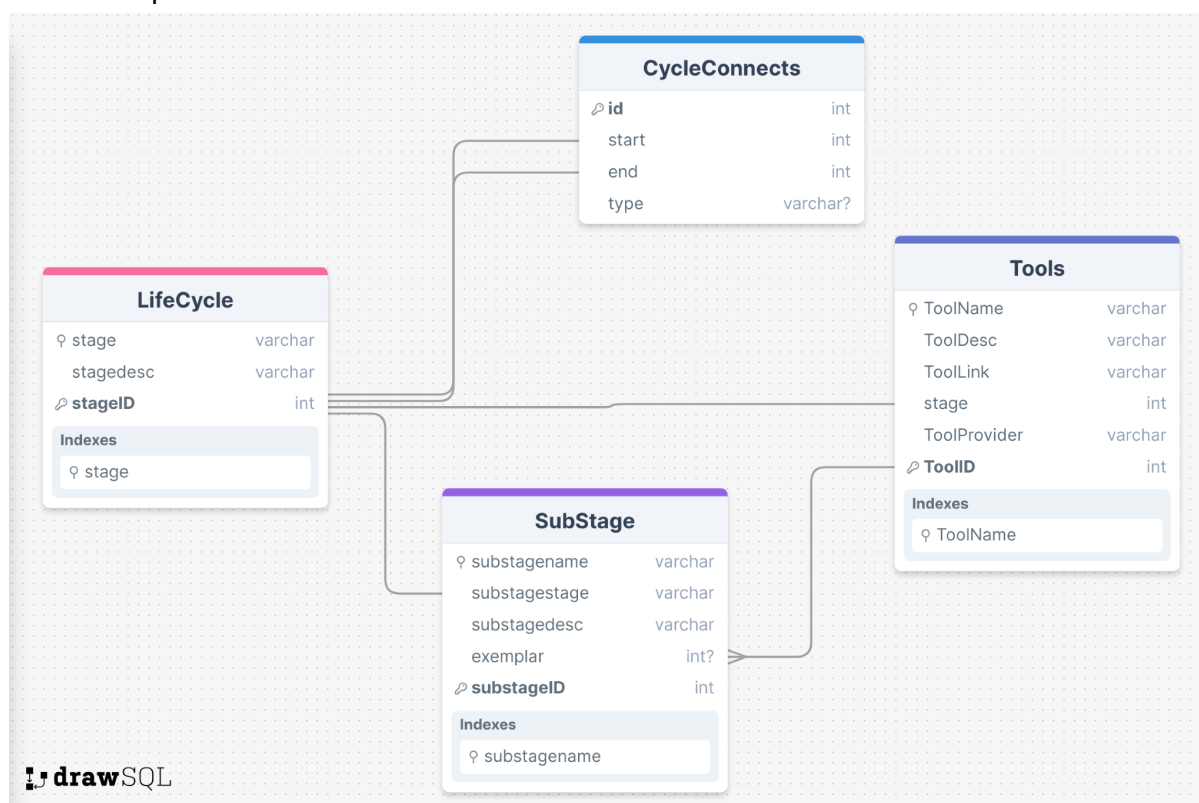
- They found 26 research data lifecycles (only 26??)
 - Literature review and knowledge of group members used
 - Had to move on in interest of time
- Q (Roxanne Wyns): Universities often offer specific tools and dedicated support for that. Not that researchers need to limit themselves to those tools, but they get institutional hosting, subscription, training and advice etc. Would it be possible to take this visualisation and apply it to an institutional context?
 - The discussion is how malleable and how adaptable this is for institutional purposes. A good aspect of the WG is how to facilitate institutions tailoring the work to their own needs.
- Q (Mingfang Wu): Is the categorisation schema intended to help people find the right tool?
 - Some information about adopters here: <https://www.rd-alliance.org/plenaries/creating-a-categorisation-schema-of-digital-research-tools-mapped-to-the-research-data-lifecycle/>
 - GORC Model overview at https://andrew.treloar.net/research/diagrams/gorc_elements.html
- C (Tom Honeyman): There's also a body of work looking at tools used in research by extracting mentions of software and classing them according to a much coarser system than done here (use, creation, sharing). Lots of branches in this but, at least software-mentions, SoftCite and SoMeSci etc etc. These sources of data would have given you a much larger list of tools, but the challenge is classifying them. And then researchers describe only a subset of the tools they actually use.
 - This is the right time to explore this. Work on Software PIDs and complex citations will help with tracking tool usage.

Proposed next steps Notes

- Q (Andrew Treloar): This work is great, but is it just a snapshot at a point in time? Are there plans to keep it updated as the landscape inevitably changes?
 - The model has been made as open and transparent as possible so that the work can be advanced
 - Automation could be used with manual curation also needed
 - It needs to be dynamic and needs to change with the evolving landscape
- Q (Andrew Treloar): I like the visualisation work but this could be another way of visualisation and there are many mappings. How to harmonise the mapping exercises that have been done? How can we converge on a smaller number of representations?
 - Balance between inspectability and automation (human-readable representation vs machine-readable)
- Lou Bezuidenhout and Andrew have been involved in many mapping exercises and there are many groups attempting to redo it which means that we may not have satisfied the need. How can we make it applicable to funders? So they can decide how to fund tool development. Needs to lead to convergence of effort
 - Part of the follow up could be to bring people and efforts in space together to think about the past 10 years of spreadsheets and think about where to take them forward.
 - Something becomes sustainable if the market or community likes it as it is used and supported by the community, or if there is some funding to keep the model alive. None have been captured as a standard but there are many reasons for this - the GORC model is gaining traction and could eventually become a standard.
- C (Andrew Treloar): GORC commons models into a commons representation - similar to the work this group has done on tools - applauds the group for their work.
- C (Andrew Treloar): Building on the legacy base of spreadsheets is a non-trivial problem.
- C (Andrew Treloar): Looking for maps that are not too general and not too specific.
- C (Tom Honeyman): this is a large landscape to map, you have to think on the roadmap about the methods that can be applied to scale otherwise it becomes a model for its own sake. The next phase would be to identify alternate models to scale up, it might need to be scaled down or elaborated upon to scale up, and have to accept a certain amount of noise. If it takes a person manual effort to scale up then this is an enormous effort and automated processes would be an option.
 - Classifiers and synchronising with other work.
- TH: Gathering of use cases could be a useful next step.
 - Great suggestion. Now we have a model where we could collect implementation use cases.
- They are planning on a follow up WG that will provide more detail on the tools (plus more tools)
- TH: see also a report that the ARDC commissioned looking at suitable infrastructure for greater visibility of software. In particular the “Software Use Infrascopes” in the recommendations.

<https://ardc.edu.au/resource/research-software-visibility-infrastructure-priorities-report/>

- Opportunity to interface with GORC groups: strong synergies and complementarities
- Simple schema



Expression of interest to continue this work

Please add your name and contact information if you would like to further this work via a follow-up RDA group.

Name	Email

--	--

Second, Repeat Session (if applicable): Tuesday 21 May, 14:00 - 15:30 UTC

Attendee Check-in from the second, repeat session

Please complete this table to indicate your attendance (add rows as needed):

Name	Affiliation	Location	Email/social media
CJ Woodford	Digital Research Alliance of Canada	Ontario, Canada	c.joseph.woodford@gmail.com
Veronika Stoka	J. Stefan Institute	Ljubljana, Slovenia	veronika.stoka@ijs.si
Andrea Budac	World Data System	Canada	abudac@oceannetworks.ca
Rory Macneil	Research Space	Scotland	rmacneil@researchspace.com
Marcelo Garcia	KAUST	Saudi Arabia	marcelo.garcia@kaust.edu.sa
Halle Burns	Princeton University	USA	halle.burns@princeton.edu
Graeme Diack	Atlantic Salmon Trust/Missing Salmon Alliance	Scotland	graeme@atlanticsalmontrust.org
Nick Juty	The University of Manchester; ELIXIR	UK	nsjuty@gmail.com
Andrea Medina-Smith	National Institute of Standards and Technology	USA	
Tilo Mathes	Research Space	Berlin, Germany	tilo.mathes@researchspace.com / LI
Sophie Fortuno	CIRAD	Montpellier, France	sophie.fortuno@cirad.fr
Adam Vials Moore	Jisc	UK	adam.vialsmoore@jisc.ac.uk

Total: 20 participants

Collaborative Session Notes *from the second, repeat Session (To be used by participants and chairs during the session).*

Quick links:

- [About the WG](#)
- [Collaborative notes](#)
- [Presentation slides](#)
- [Deliverable 1](#) - Harmonised Research Data Lifecycle Model and Crosswalk to existing models
- [Deliverable 2](#) - Categorisation schema of digital research tools
- Deliverable 3 - Visual map of digital research tools landscape (prototype)

Session Agenda:

Time (min)	Agenda item	Lead
5	Welcome and introduction	Facilitator
5	Presentation: Introduction to the RDA-OfR Mapping the Digital Research Data Infrastructure WG Background and description Aims and objectives	Marcelo Garcia
15	Presentation: Deliverable 1 & Methodology Presentation: Deliverable 2 & Methodology Presentation: Deliverable 3 & Methodology	Adam Vials Moore, Co-chair Rory Macneil, Co-chair
10	CJ Woodford presentation on Mapping - GORC	CJ Woodford
20	Q&A	All
5	Proposed next steps	Rory Macneil
25	General discussion	Facilitator
5	Closure of session	Facilitator

Presentation slides:

 P22 Plenary.pptx

Presentation notes

Deliverable 1: The creation of a harmonised research data lifecycle model and crosswalk to existing models

- ...
- ...
- ...

Deliverable 2: Categorisation schema of digital research tools

- ...
- ...
- ...

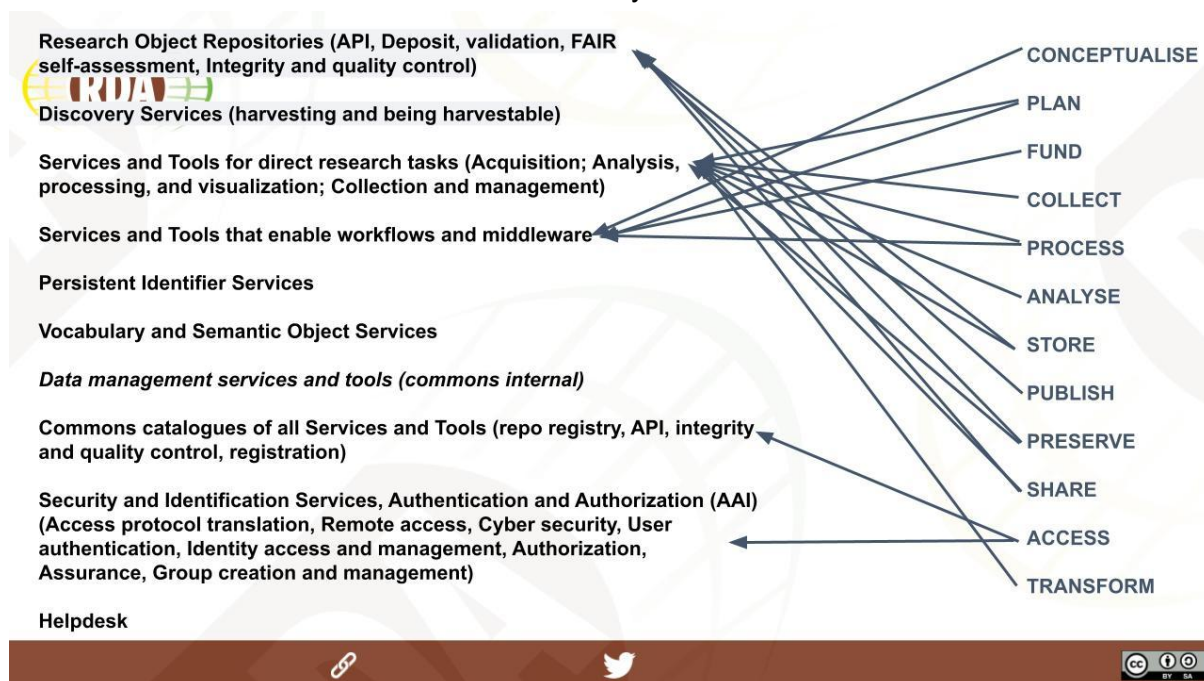
Deliverable 3: Visual map of digital research tools landscape (prototype)

- ...
- ...
- ...

GORC Notes

- International connection between research commons - encapsulating all research infrastructure effort on a shared platform to enable research
- Clouds (national, regional, discipline level) - make them interoperable so that research objects that are needed are available
- GORC IG Roadmap & GORC International Model WG - a grocery list of things to include in your commons
- Essential elements (IG) - Human and Technical elements (9). 400 items in the model. Each commons would have to have something in each element.
- Human - interoperability standards, Governance structures, rules of participation and access, engagement, sustainability and human capacity.
- Technical - Compute storage network AAI, Services and tools and research objects
- KPIs and metrics included
- Services and tools - complementary work being done between GORC and MaLDReTH

- Mapping between GORC services and tools and the MaLDReTH work - mapping from MaLDReTH to GORC and not the other way round



Q&A Notes

Question (Q), Comment (C)

- (Q) CJ: Question for Adam and Rory: how did you determine whether a tool was interoperable or not? It looks like a binary value in the visualisation, if it really is True/False, what are your current criteria to determine the state?
 - This needs to be worked on in subsequent iterations
- (Q) CJ: Who is this for? - the market determines who the product is for.
- (Q) Andrea: What would success look like for these outputs? - Usage, for example UCL could use as a benchmark, GORC - those who build research commons, ELIXIR - domain infrastructures,
- (Q) Connie to CJ: Automation for mapping, and read and revise is good but not many AI models can achieve this. Mapping has to be done manually as there is a lot of conceptual work behind it.

Proposed next steps Notes

- ...
- ...
- ...

Expression of interest to continue this work

Please add your name and contact information if you would like to further this work via a follow-up RDA group.

Name	Email
------	-------

Halle Burns	halle.burns@princeton.edu