International landscape on research software

Tom Honeyman, Mingfang Wu
Australia Research Data Commons

27 November, 2020
RDA 16th Virtual Regional Event

research data sharing without barriers
rd-alliance.org
In opening this session, I would like to acknowledge and celebrate the Bunurong people whose traditional lands I’m meeting with you from today. And I would like to pay my respects to their Elders both past, present and emerging. I extend that respect to Aboriginal and Torres Strait Islander peoples and Elders from other communities who may be joining us today.
Agenda Today:

1. Welcome, Icebreaker and Introductions

2. About RDA and How to become a RDA member and join groups

3. IG - Software Source Code: Existing Efforts and Practices Related to Software Source Code in Academia

4. WG: RDA/FORCE11 Software Source Code Identification
   a. WG Supporting Output: Use Cases and Identifier Schemes for Persistent Software Source Code Identification (V1.1)

5. Force11 activities


7. WG - FAIR 4 Research Software (FAIR4RS): Progress on FAIR 4 Research Software WG

8. BOF - Building a FAIR Roadmap for Research Software

9. WG - CURE-FAIR: Current State of Curation Practices that Support Computational Reproducibility Other Groups

10. BOF - Computational Notebooks

11. Posters
Session notes:
Muted during presentations
Unmute if you’d like to talk

For questions & comments
use the Chat box

Turn on your video
if you have the bandwidth

#RDAPlenary
We welcome little people & pets

Come prepared to contribute and have fun!
Enable the annotations when a screen is shared by navigating to “View Options” at the top of the window, and choosing “Annotate”.

- add text, draw lines or shapes, add a stamp
- use the eraser to delete one of your annotations or undo the last step
- use clear to clear all annotations, or just yours
- if you are leading a room in a breakout, choose “save” before the breakout closes so you can share the annotations back to the whole group
Where are you joining us from today?
### What is your involvement in the Research Data Alliance?

<table>
<thead>
<tr>
<th>None</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I know of it, but am not a member</td>
<td></td>
</tr>
<tr>
<td>I am a member</td>
<td></td>
</tr>
<tr>
<td>I am member of at least one Working or Interest Group</td>
<td></td>
</tr>
<tr>
<td>I am co-chair of at least one Working or Interest Group</td>
<td></td>
</tr>
<tr>
<td>I am a member of an RDA governance body</td>
<td></td>
</tr>
</tbody>
</table>
Did you attend any Research Data Alliance Virtual Plenary 16 sessions?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
What brings you to today’s session
Research Data Alliance

Vision
Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

Mission
RDA builds the social and technical bridges that enable open sharing of data.

The RDA in a nutshell (June 2020)
https://www.slideshare.net/ResearchDataAlliance
How to become a RDA member

https://www.rd-alliance.org/ -> Registration
VP16 Regional Event – Daily Sessions

• Monday, 23 November: COVID-19: Role of data in responding to COVID-19 pandemic. What can we do to respond better in the future?

• Tuesday, 24 November: Data versioning and data life cycle

• Wednesday, 25 November: FAIR

• Thursday, 26 November: Skills

• Friday, 27 November: Research Software

Slides will be made available at https://tinyurl.com/yy2fz4vh
THANK YOU VERY MUCH!

RDA
Email - enquiries@rd-alliance.org
Web - www.rd-alliance.org
Twitter - @resdatall
LinkedIn -
www.linkedin.com/in/ResearchDataAlliance
Slideshare -
http://www.slideshare.net/ResearchDataAlliance

Stefanie Kethers
Director of Operations, Research Data Alliance
Email – Stefanie.Kethers@ardc.edu.au
Software Source Code (SSC)

Is a digital object, the only representation of software that contains human readable knowledge

Objectives:

A forum for discussing research software inside RDA

- issues on management, sharing, discovery, archival and provenance of software source code.
- It will pay special attention to source code that generates research data and plays an important role in scientific publications.

Note: Slides 7 - 24 are adapted from the RDA session: [Existing efforts and practices related to software source code in academia”](#) (Prepared by Morane Gruenpeter, Inria, Software Heritage, France)
RDA Software Source Code Activities

BOF RDA P9, Barcelona April 2017
motivations => 60 participants

RDA P10, Montreal September 2017
motivations, survey of ontologies, metadata
use cases

RDA P11, Berlin March 2018  started the
idea for a dedicated identification WG

RDA P13, Philadelphia April 2019  FAIR for
Software Source Code and launch of the SCID
WG

RDA VP15, Australia March 2020  Open
discussion about the creation of a new group,
the FAIR4RS WG (which was launched in June
2020)

RDA VP16, Costa Rica November 2020
Existing efforts and practices in Academia
Joint RDA & FORCE11 (software citation implementation WG) effort

Activity: Capture and analyse the software identification state-of-the-art in the scholarly ecosystem

Output: Use cases and identifier schemas for persistent software source code identification  DOI: 10.15497/RDA00053
Who cares?
citation manager, collaborative development platforms, creator/librarian/digital archivists/funder, indexer, institution/research center/university, institutional/national/domain repository, library, package manager, policy maker, publisher/publication venue, registry, researcher as a software user, research as a software author, software engineer

What is at stake?
Archival, Reference, Description, Credit

DOI: 10.15497/RDA00053
## Example of 29 use cases

<table>
<thead>
<tr>
<th>Actor</th>
<th>Use case description</th>
<th>Action</th>
<th>Identification target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive</td>
<td>Identify all the software artifacts I hold</td>
<td>Archiving, referencing</td>
<td>Release and smaller artifacts</td>
</tr>
<tr>
<td>Citation manager</td>
<td>Curate the software citation entries</td>
<td>Credit</td>
<td>Project, release</td>
</tr>
<tr>
<td>Curator / librarian / digital archivist</td>
<td>Catalog and browse the development history of legacy software source code for preservation purposes (The Apollo mission source code is a good scenario on how making code available on GitHub isn’t enough for persistence purposes)</td>
<td>Archiving</td>
<td>Project, release and smaller artifacts depending on the reference</td>
</tr>
<tr>
<td>Publisher</td>
<td>Create/retrieve identifiers quickly for use in the paper for all software including commercial packages.</td>
<td>Referencing, describing</td>
<td>Any item (all granularity levels)</td>
</tr>
<tr>
<td>Registry</td>
<td>Identify and curate the software entries I hold</td>
<td>Archiving, referencing, describing, credit</td>
<td>Project</td>
</tr>
<tr>
<td>Researcher as a software user (RSU)</td>
<td>Access and use SSC no longer available on a collaborative platform</td>
<td>Archiving</td>
<td>Snapshot, release, revision, directory</td>
</tr>
</tbody>
</table>
## SSCI Working Group - Output (3)

<table>
<thead>
<tr>
<th>Granularity level (GL)</th>
<th>ID target</th>
<th>Extrinsic identifiers</th>
<th>Intrinsic Identifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ASCL</td>
<td>ARK</td>
</tr>
<tr>
<td>GL1</td>
<td>project</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GL2</td>
<td>project version</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>GL3</td>
<td>module</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>GL4</td>
<td>repository</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GL5</td>
<td>repository snapshot</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>GL6</td>
<td>release</td>
<td>X</td>
<td>X**</td>
</tr>
<tr>
<td>GL7</td>
<td>commit</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>GL8</td>
<td>directory</td>
<td>X</td>
<td>X**</td>
</tr>
<tr>
<td>GL9</td>
<td>file</td>
<td>X</td>
<td>X**</td>
</tr>
<tr>
<td>GL10</td>
<td>Code fragment</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

* The HAL-ID when combined with a SWHID can identify also the directory of the source code in its metadata  
** depends on the deposited artifact

DOI: [10.15497/RDA00053](https://doi.org/10.15497/RDA00053)

---

[RD Alliance](https://rd-alliance.org)
Next step:

- The working group has now **completed** its work in its current form
- Maintenance of the working group output transfers to the **SSC IG**

“The next step would be to produce a set of recommendations based on these findings.”
FORCE11 (The Future of Research Communication and e-Scholarship) is an international coalition of researchers, librarians, publishers and research funders working to reform or enhance the research publishing and communication system.

FORCE11 Software Citation Working Group (April 2015 - February 2016)
- Output: 6 software citation principles
  - Importance
  - Credit and Attribution
  - Unique Identification
  - Persistence
  - Accessibility
  - Specificity

Objective: Produce concrete guidelines for software citation, and implement them within the scholarly research community (software developers, repositories and registries, journals and conference and publishers, indexers, institutions)

Activity: A community with monthly calls to discuss challenges and progress in implementing software citation, with task forces for

- **Guidance** - developing documents for developers, authors, and reviewer
- **Journals** - coordinating editors and publishers to simplify and implement guidance
- **Repositories** - developing best practices document for handling software
- **CodeMeta** - standardizing metadata for software, moving towards merging into schema.org
Objective: Develop guidance for different stakeholders to help implement software citation, principally authors of research articles seeking to cite software correctly and developers of software looking to make their software easier to cite.

Milestones and activity:

- Developed and published [Software Citation Checklist for Authors](#) and [Software Citation Checklist for Developers](#)
- Developed Software Citation Primer which was used as basis for paper by the Journals Task Force
- Guidance Task Force on hiatus while the Journals Task Force provides feedback on guidance
Objective: Work with organizations that publish journals, proceedings, monographs to improve how software is cited in their works and the scholarly processing ecosystem

Activity:
- Published paper (in peer-review) - about 18 of authors representing publishers
- Working on comms plan
- Communities and institutions will produce versions of the document with software examples and citation styles that are appropriate for their intended audience
- Next steps:
  - Work on what happens after article is submitted – how citations are processed and indexed – to ensure they are correctly registered and tracked
  - You: tell your communities about this, encourage publishers to support it, encourage authors & reviewers to follow it
Objective: Bring together representatives of Research Software Registries and Repositories to discuss and improve practices

Milestones and activity:

- Drafted Quick Start Guides for Research Software Registries and Repositories Best Practices document
  - Aiming to release by end of 2020
- Working on guidance document for users/community best practice
- Working on code that takes information from a Python setup.cfg file and generates a codemeta.json file
- Held [workshop for discipline-specific software registries and repositories](#) in Nov 2019
  - Produced drafts of documents such as "A good-enough workflow for software citation"
- Task force was somewhat stalled during COVID-19, but is restarting
Objective: express all codemeta properties using schema.org

Milestones and Activity:
- An academic community discussing software metadata
- A crosswalk table - mapping the metadata landscape (including 23 schemas or ontologies)
- CodeMeta properties
  - 67 in total, 57 of them are properties from schema.org
  - 10 proposed properties (e.g. readme, issueTracker, buildInstuction, et al.) that are new to schema.org

Vision: Research software: recognised and valued as a fundamental and vital component of research worldwide

Mission: To bring research software communities together to collaborate on the advancement of research software.

Task Forces:
- Software landscape analysis
- Evidence for the importance of research software
- Register of research software funding opportunities

More information: https://www.researchsoft.org/

To subscribe to the monthly ReSA newsletter send a blank email to research-software-alliance+subscribe@googlegroups.com
• FAIR principles for research software
• Guidelines for applying them
• Definitions, implementation guidelines and adoption examples.
• Endorsed in late April
• Subgroup activities underway
• Subsequent activities very likely
• Middle of next year for FAIR principles
FAIR4RS WG

- A fresh look at FAIR for Research Software
- FAIR work in other contexts
- Definition of research software
- Review of new research related to FAIR Software
(1) Almost finished on a proposed set of principles
(2) Compared to FAIR notebooks, training, workflows, other research objects
(3) Collected quotes, exemplar software, currently sorting
(4) Running a survey (still open)
Building a FAIR Roadmap for RS (BoF)

● how and when to plan a broad integration across the software landscape of the FAIR principles for software

● Another session at IFCS next Tuesday (10pm AEDT)

● (mailing list)
“Curation for Reproducibility”

- establish standards-based guidelines for curating for reproducible and FAIR data and code
● Endorsed in July

● Snapshot of current practices
● Synthesis of practices
● Guidelines
4 Subgroups:

- Definitions
- Practices
- Challenges
- Alignment
“Current State of Curation Practices that Support Computational Reproducibility”

- Report back and activity
CURE-FAIR WG: Update

Role

- Researcher / RA / Head of Research group
- Research Software / Research Science /...
- Data professional (Archivist / Curator / Steward)
- Reproducibility Advocate / Chair / Reviewer
- Committee at AGILE Conference
- Data Scientist
- Journal Editor
- Student
- Teacher
CURE-FAIR WG: Update

- **Findable**
  - Difficult to find data
  - Author did not make available
  - Environment no longer available
  - Proprietary software
  - High cost of archiving big data

- **Accessible**
  - Author did not make available
  - Software version no longer available
  - Environment no longer available
  - Proprietary software

- **Interoperable**
  - Does not work in another environment

- **Reusable**
  - Code did not run/not executable
  - Code did not work as intended
  - Code could not be fully automated
  - Code written in various languages/different formats
  - Code/libs/packages deprecated and obsolete
  - Differences between OS may impact results
  - Uncertain if pipeline will still work in the future
  - Results not reproduced
  - Very hard to do
CURE-FAIR WG: Update

Other: Journal review process slow to change
Other: journal would not publish a correction
Increase requirements for authors
Lack of training
Lack of training about documentation
Lack of incentives/rewards

Recommendation: Journal standards for reporting
Recommendation: Journal standards for documentation
Recommendation: Open software
Recommendation: Make source code freely available
Recommendation: Include libraries and dependencies

Recommendation: More carrots to encourage reproducibility
Recommendation: More training

Solution: Better documentation
Solution: Standardized template for transparent reporting
Solution: Publish a correction
Solution: Write my own code
Solution: Update the code
Solution: Run on same computing environment as author
Solution: Create own tool
CURE-FAIR WG: Update

Survey still open
forms.gle/R1MSBL5jQgkgXPTn7

Activity: gather challenges by role
cure-fair_challenges@googlegroups.com
Computational Notebooks (BoF)

- **Second BoF**
- *(first in Helsinki)*

- Seeking interest in forming an IG
Computational Notebooks (BoF)

- DataCite experiences
- Archiving and preserving notebooks
- Big data and compute
- Earthcube meeting


Thank you ...