

The Role of Libraries in the Age of Computational Reproducibility

Vicky Steeves (@VickySteeves) & Gabriele Hayden (@GabrieleHayden)

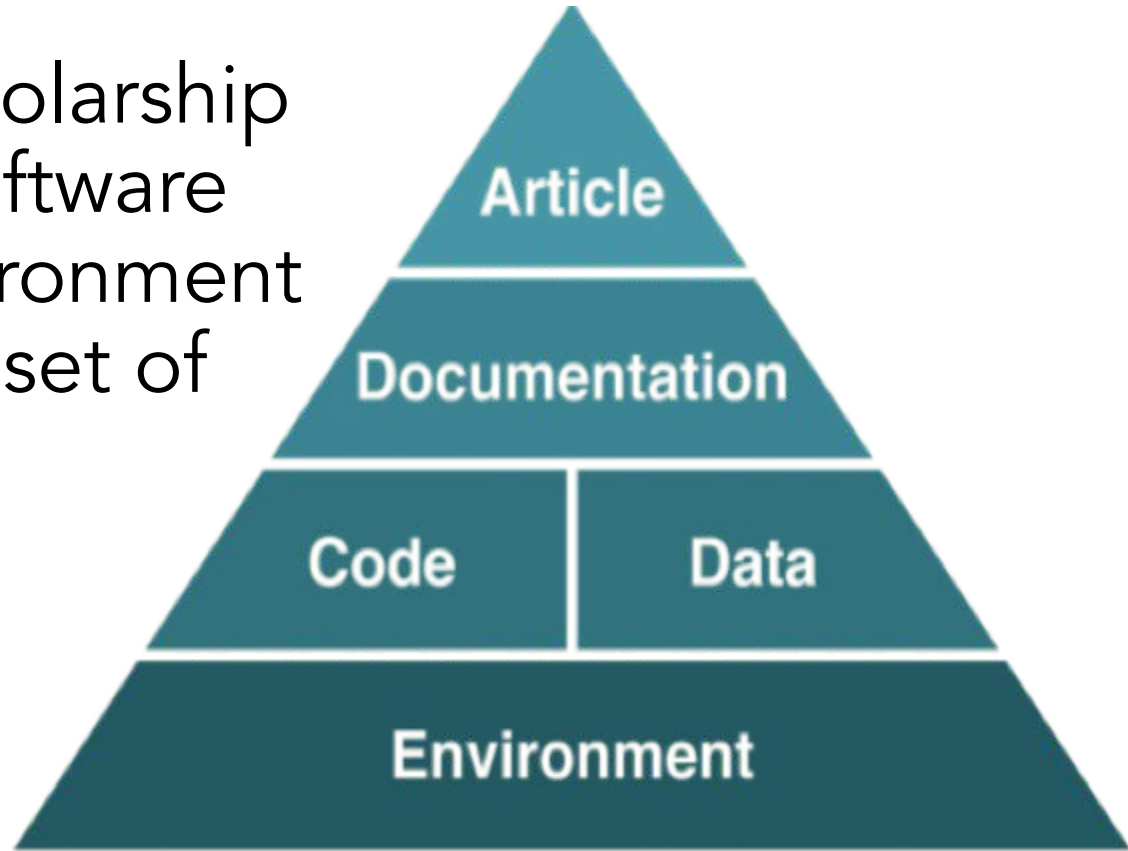
The Librarians for Research Data Management and Reproducibility
(!!!)

"even the most perfect reproduction of a work of art is lacking in one element: Its presence in time and space" - Walter Benjamin

Aura < - > Computational Environment

“... The actual scholarship is the complete software development environment and the complete set of instructions which generated the Figures”

(drink!)



Reproducibility Pyramid. Image courtesy of Andrew Rarig (NYU)

Data management is needed for reproducibility, but **doesn't guarantee it**

Data Management best practices:

- Documentation/metadata
 - Literate programming
- Open source code
- Open file formats
- Version control
- Storage + backups (3-2-1 rule!)
- Publishing research materials

Reproducibility requires:

- Everything in data mgmt best practices
- PLUS:
 - A description or capture of *workflow* and *provenance* of research practices
 - A description or capture of computational *environment*

Data Management + Reproducibility + Libraries

Field of data librarianship has sparked changes in what librarians are offering:

- Data management/data services
 - Data/information literacy, data collection for reuse
- Grants support
 - Data management plan review
- Repository services, especially data repositories
 - Publishing and licensing support
- Digital preservation
 - This means not only keeping bits, but enabling re-use!)

We are interweaving these into holistic research reproducibility services!

Reproducibility Services in the Library

- Infrastructure
 - Providing funds, labor, and use of FLOSS reproducibility tools for comm
- Education
 - Teaching classes/workshops in library as well as embedded in classes
 - Consultations with community to help them make work reproducible
- Collection of Research Materials
 - *area of improvement* -- making sure the things we collect are reproducible (potential: ReproZip-ing stuff before inclusion in collection)
- Digital Preservation
 - **Example:** Emulation as a Service as a means of preserving software + OS, also provides these computational environments to researchers to work with older programs (e.g. WordPerfect)

Call to action!

1. If you want to see more reproducibility-themed events or services on your campus, **reach out to your librarians**
 - Chances are they are thinking about what you all want, so just tell them -- it's our job + goal to help you with your work!
2. **Advocate for your librarians** to be in conversations around reproducibility, either in policy or IT or T&P (let us be your info hub)
3. **Cite librarians!** Especially if you're talking about provenance (started in libraries), re-using materials in a repository (20+ years of lit on this)

Some light reading --

- Steeves, Vicky. [Reproducibility Librarianship](#)
- Sayre, Franklin and Riegelman, Amy. [The Reproducibility Crisis and Academic Libraries](#)
- Vitale, Cynthia R.H. [Is Research Reproducibility the New Data Management for Libraries?](#)
- Barba-group reproducibility syllabus:
<https://hackernoon.com/barba-group-reproducibility-syllabus-e3757ee635cf>
- Ten Simple Rules for Reproducible Computational Research:
<http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003285>
- Five selfish reasons to work reproducibly:
<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0850-7>
- How scientists fool themselves – and how they can stop:
<https://www.nature.com/news/how-scientists-fool-themselves-and-how-they-can-stop-1.18517>