

Goals:

- Publish in the 'proper way', e.g. notebooks
- How to make notebooks more sexy so they will be reused.
- Sharing materials data, JSON that describes the notebook. The separability of the notebook from the kernel. How do you version control and describe in a certain way.
- QUILT - take data and wrap with notebook into a container
- How to create metadata in a standardized way, esp. with domain-specific metadata standards.
- Can notebooks be metadata in themselves?
- Supporting people who use notebooks, e.g. repository, sharing.
- FAIR workflows esp. workflows in Jupyter, at least 10 separate ontologies for workflows
 - <https://github.com/fair-workflows>
 - <https://github.com/Sheeba-Samuel/ProvBook>
- Provbook - plug in for the Jupyter notebook, can store info for each cell
- Notebooks to document from going from raw to controlled data, also data viz.

What are the gaps and opportunities?

Chopportunities (Challenge + Opportunity)

- Notebooks as a way to embed provenance and data transformation.
 - But what about when the plug-in stops working.
- No clear definitions for making notebooks FAIR, e.g. no metrics or maturity indicators.
- A notebook is and is not software. It contains cells with codes, but is also a document.
- Scope - what is in a notebook and what is included, differs between notebooks.
- What about JupyterLab? Replacing Jupyter but when to begin.
- Notebooks as a way to document FAIR for a workflow.
- Licensing for data in the notebooks not the same as license for the notebook.
- FAIR is inherently about being machine testable/readable.
- Not everyone uses notebooks, e.g. linguistics people put code on Zenodo. Don't concentrate so much on notebooks, we forget to serve other scientists who don't use them.
- There are more steps than are in the notebook. Manual steps.
- How to work with big data sets. Being able to reproduce even getting the data, e.g. big climate datasets.

What could this group (or the RDA community) do here that would be done better than elsewhere?

- Establish a working group:
 - Connect up with other WG to understand the implications of data and notebook licensing differ, esp. for sharing notebooks. (WG: Gail Clement, on Data Policies and research compendia).

- SOFTWARE PROVIDER: What would it mean for a notebook provider to be FAIR compliant? Have opportunity to create FAIR notebooks. And to give recommendations on how to make them FAIR.
 - Lock your notebook as a version. How to make it findable. Need metadata about the notebook and then about what the notebook's 'point' is.
- IT/RESEARCH COMPUTING: Best practices for institutional notebook hosting. Course support/hosting, but also for research. JupyterHub or ?
- USERS/RESEARCHERS:
 - Best practices for including manual steps not in notebook.
 - Making workflows FAIR.
 - Working with large datasets.
 - [Batch vs. interactive job launching from notebooks. From EGI-led group]
- How to machine test notebooks for FAIR.
- Scope - what is included in a notebook? No: the code that drives the notebook. How much should be embedded in the notebook vs. called.

What are the next steps? (More time to read and discuss? Reports to be written? Code to be developed?)

- Socialize FAIR Software to RDA community (about to be published)
- Working group with output on
 - FAIR definitions/metrics/maturity indicators.
 - Making notebooks FAIR and behaviors that engender FAIR.
 - Best practices (or 'known good config'/this works) Institutional notebook hosting for courses and research