



## Hybrid Workshop

**Day 1:** Thursday 3 November 2022, 9-4 EDT

**Day 2:** Friday 4 November 2022, 9-3 EDT

**Meeting time** (Agenda times below in EDT - use [this tool](#) to see your time zone)

|       | PDT        | EDT               | BST/ART     | UTC         | CET         | AEST       |
|-------|------------|-------------------|-------------|-------------|-------------|------------|
| Day 1 | 6:00-13:00 | <b>9:00-16:00</b> | 10:00-17:00 | 13:00-20:00 | 14:00-21:00 | 23:00-6:00 |
| Day 2 | 6:00-12:00 | <b>9:00-15:00</b> | 10:00-16:00 | 13:00-19:00 | 14:00-20:00 | 23:00-5:00 |

Link to meeting folder: <https://bit.ly/NotebooksNOW>

## Day 1 - 3 November

| Time (EDT)                | Agenda Item  |
|---------------------------|--|
| <b>8:00am</b><br>60 mins  | <b>Breakfast</b><br>Breakfast will be served outside the meeting room starting at 8:00am   |
| <b>9:00am</b><br>30 mins  | <b>Welcome and Introduction</b><br><i>Shelley Stall</i> <ul style="list-style-type: none"><li>Review meeting <a href="#">code of conduct</a>, objectives, and logistics</li><li>Discuss the rules of engagement</li></ul>  |
| <b>9:30am</b><br>45 mins  | <b>Background and Perspectives: Implementing Computational Notebooks as a Format into the Publication Workflow (Part 1)</b><br>Steering committee members will present 15 minute talks: <ul style="list-style-type: none"><li><b>Teaching, Learning and researching with notebooks</b><br/><i>Yanina Bellini Saibene</i></li><li><b>Jupyter project, researcher use of notebooks, and community</b><br/><i>Fernando Perez</i></li><li><b>The happiest notebooks on Earth (R Markdown)</b><br/><i>Alison Presmanes Hill</i></li></ul> |
| <b>10:15am</b><br>15 mins | <b>Break</b>   |

|                                   |  |
|-----------------------------------|--|
| <p><b>10:30am</b><br/>35 mins</p> | <p><b>Background and Perspectives: Implementing Computational Notebooks as a Format into the Publication Workflow (Part 2)</b></p> <p>Steering committee members will present 15 minute talks:</p> <ul style="list-style-type: none"> <li>● <b>Introduction</b><br/><i>Brooks Hanson</i></li> <li>● <b>Peer review vendor workflow</b><br/><i>Joel Plotkin (eJournalPress)</i></li> <li>● <b>Notebooks in publishing workflows</b><br/><i>Alberto Pepe (Wiley)</i></li> </ul>  |
| <p><b>11:05am</b><br/>30 mins</p> | <p><b>Discussion on Background and Perspectives Presentations</b></p>  |
| <p><b>11:35am</b><br/>25 mins</p> | <p><b>Breakout Instructions and Workstream Discussion</b></p> <ul style="list-style-type: none"> <li>● Participants will receive instruction on breakout group work</li> <li>● Brief presentation on project charters</li> <li>● Discuss potential updates to workstreams</li> </ul>   |
| <p><b>12:00pm</b><br/>60 mins</p> | <p><b>Lunch</b></p>  |
| <p><b>1:00pm</b><br/>90 mins</p>  | <p><b>Breakout 1: Workstream Charters</b></p> <p>Please select the workstream that aligns closest with your expertise and interest:</p> <ul style="list-style-type: none"> <li>● <b>Pre-Submission</b> - Exploratory stage where early organizational approaches can help structure work in notebooks, to better prepare for publication submission.</li> <li>● <b>Submission &amp; Metadata</b> - Manuscript submission in structured format with associated data/software and where metadata are collected (e.g., author information, etc)</li> <li>● <b>Editorial &amp; Peer Review</b> - Conversion and extraction of manuscript (metadata) information. Facilitation and tracking of manuscript peer review.</li> <li>● <b>Production &amp; Post Production</b> - Dissemination, hosting, and preservation of final output (version of record)</li> </ul> |
| <p><b>2:30pm</b><br/>15 mins</p>  | <p><b>Break</b></p>  |
| <p><b>2:45pm</b><br/>75 mins</p>  | <p><b>Workstream Summaries and Discussion</b></p> <p>Participants will report-out on their workstream discussions</p>  |
| <p><b>4:00pm</b></p>              | <p><b>Adjourn / Reception</b></p> <p>Please join us for an optional reception with light refreshments</p>  |

## Day 2 - 4 November

| Time (EDT)                | Agenda Item   |
|---------------------------|---|
| <b>8:00am</b><br>60 mins  | <b>Breakfast</b><br><b>Breakfast will be served outside the meeting room starting at 8:00am</b>   |
| <b>9:00am</b><br>15 mins  | <b>Day 1 Summary and Highlights (Shelley)</b>   |
| <b>9:15am</b><br>85 mins  | <b>Lightning Talks on Notebooks Approaches, Services, and Workflows</b><br><b>Community experts will present 10 minute lightning talks:</b> <ul style="list-style-type: none"> <li>● <b>Notebooks 4 Education <i>Lorena Barba</i></b></li> <li>● <b>JupyterBook <i>Chris Holdgraf</i> (<a href="#">zenodo record</a> of slides)</b></li> <li>● <b>rOpenSci <i>Yanina Saibene and Karthik Ram</i> (slides: <a href="#">rOpenSci - Notebooks Now</a>)</b></li> <li>● <b>Quarto <i>JJ Allaire</i></b></li> <li>● <b>Curvenote <i>Rowan Cockett</i></b></li> <li>● <b>Neurolibre/Notebooks (Peer Review) <i>Jean-Baptiste Poline</i></b></li> </ul> <p><b>There is time reserved at the end for questions and group discussion.</b></p> |
| <b>10:40am</b><br>20 mins | <b>Break</b>  |
| <b>11:00am</b><br>75 mins | <b>Breakout 2: Workstream Charters (Shelley)</b><br>Please select another workstream that aligns closest with your expertise and interest. You are welcome to join the same workstream as Breakout 1: <ul style="list-style-type: none"> <li>● <b>Pre-Submission</b></li> <li>● <b>Submission &amp; Metadata</b></li> <li>● <b>Editorial &amp; Peer Review</b></li> <li>● <b>Production &amp; Post Production</b></li> </ul>  |
| <b>12:15pm</b><br>60 mins | <b>Lunch</b>  |
| <b>1:23pm</b><br>75 mins  | <b>Group Discussion on Workstream Charters</b> <ul style="list-style-type: none"> <li>● Participants will report-out on their workstream discussions</li> <li>● The group will develop a schedule for coming months and identify relevant stakeholders</li> </ul>   |
| <b>2:30pm</b><br>30 mins  | <b>Next Steps and Wrap-up</b>   |

## Rules of Engagement

- **Bring grace and understanding to your conversations with other attendees.**
  - We're all here because we care about the end goal!
- **Avoid preconceptions; listen to the voices of those around you.**
  - *For instance:*
    - “Existing publishing systems are too inflexible to incorporate new technologies and workstreams”
    - “If it's not completely open all the time, it's not worth doing”
- **Anything else? Please tell us!**

## Community Notes

**\*\* You can add your notes in the section \*\***

Slack invite [link](#)

#workshop-discussion channel [link](#)

1. Presentation by Yani bellini Saibene:
  - a. [Bellini - Notebook Now](#)
2. Presentation by Fernando
  - a. Link to the LIGO/Black Hole sonification notebook: <https://bit.ly/black-holes-woop>
  - b. Repo that estimates the number of public jupyter notebooks on Github: <https://github.com/parente/nbestimate>
  - c. “Why I think ending article-processing charges will save open access”: <https://www.nature.com/articles/d41586-022-03201-w>
  - d. Open Source and Open Science in Latin America <https://cziscience.medium.com/open-source-and-open-science-in-latin-america-e3a1f5cc6744>
  - e. Código Abierto y Ciencia Abierta en América Latina <https://www.metadocencia.org/post/ciencia-abierta-americalatina/>
3. Presentation by Allison
  - a. Link to the slide: <https://apreshill.github.io/happiest-notebooks-on-earth/>
  - b. Relevant links:
    - i. Selected complaints about notebooks from 2018: <https://yihui.org/en/2018/09/notebook-war/>
  - c. Discussion topics surfaced:
    - i. What is the line between notebook as a demonstration/publication v.s. Software package development?
      1. JOSS (Journal of Open Source Software) is a place where it can publish softwares
    - ii. Here's how we (neurolibre) iterated for a full fledged reproducible paper <https://docs.neurolibre.org/en/latest/> I don't think many groups have the technical know-how to produce this type of submissions. But many readers will see the value of the outcome.

### Publishing Workflows

4. Presentation by Joel Plotkin (*eJournalPress*)
  - a. Publication process has a scalability consideration
  - b. MECA (from NISO) to govern the publishing workflow (<https://www.niso.org/standards-committees/meca>)
  - c. Journal Article Tag Suite (JATS): <https://www.niso.org/standards-committees/jats>
5. Presentation by Alberto Pepe (*Wiley*)
  - a. Current publishing workflows “flatten” rich research content
  - b. Discussions:

- i. Issue of discoverability
  1. As long as rich metadata are available, a publication solution could/should allow discoverability and downstream services (DOI registration, etc.)
  2. JATS follows community closely; if we can come up with a recommendation, JATS may be able to include the recommendation to help aid the discoverability

#### Lightning Talks from Day 2

6. **Notebooks 4 Education** *Lorena Barba*
  - a. [paper] Barba, L.A., 2020. "Engineers Code: reusable open learning modules for engineering computations." *Computing in Science & Engineering*, 22(4), pp.26-35. doi:10.1109/MCSE.2020.2976002 Preprint arXiv::2001.00228
  - b. [repository with learning modules] <https://github.com/engineersCode/EngComp>
  - c. **Open book:** – Barba, L.A., Barker, L.J., Blank, D.S., Brown, J., Downey, A.B., George, T., Heagy, L.J., Mandli, K.T., Moore, J.K., Lippert, D. and Niemeyer, K.E., 2019. "Teaching and learning with Jupyter." Read online: <http://go.gwu.edu/jupyter4edu>  
Citable archive (PDF): <https://doi.org/10.6084/m9.figshare.19608801.v1>  
Blog post: <https://blog.jupyter.org/teaching-and-learning-with-jupyter-c1d965f7b93a>
  - d. [course platform] <http://openedx.seas.gwu.edu>
  - e. The Journal of Open Source Education <http://jose.theoj.org>
7. **JupyterBook** *Chris Holdgraf* ([zenodo record](#) of slides)
  - a. The Principles of Open Scholarly Infrastructure: [openscholarlyinfrastructure.org](https://openscholarlyinfrastructure.org)
  - b. The Executable Books Project: [executablebooks.org](https://executablebooks.org)
  - c. MyST: Markedly Structured Text ([myst.tools](https://myst.tools))
  - d. JupyterBook: [jupyterbook.org](https://jupyterbook.org)
  - e. Open Global Glacier Moder: <https://docs.oggm.org>
8. **rOpenSci** *Yanina Saibene and Karthik Ram*
  - a. Slides: [rOpenSci - Notebooks Now](https://ropensci.org/notebooksnow)
  - b. Bringing the good part of the academic peer-review to research software
  - c. rOpenSci: [ropensci.org](https://ropensci.org)
  - d. rOpenSci Packages Development, Maintenance, and Peer-Review: [devguide.ropensci.org](https://devguide.ropensci.org)
9. **Introduction to Quarto** *JJ Allaire*
  - a. Slides: <https://jjallaire.quarto.pub/notebooksnow-quarto>
  - b. Single-source publishing: <https://coko.foundation/articles/single-source-publishing.html>
10. **Curvenote** *Rowan Cockett*
  - a. Curvenote: <https://curvenote.com>

11. **Neurolibre/Notebooks (Peer Review)** *Jean-Baptiste Poline*
- a. NeuroLibre: <https://www.neurolibre.com/>
  - b. Builds on Jupyter Book, Open Journals, and BinderHub

## Day 1 Group Discussion on Workstream Charters

### Workstream Summaries & Discussion

Presubmission /Yuhan Douglas Rao (NCICS) and Rachel Kurchin (Carnegie Mellon)

- Broke into questions to “circumscribe” the effort;
  - **1st** broad question: Why would authors want to publish notebooks?
    - Broad reasons -
    -
  - **2nd**: What is the minimum viable product for submission?
    - Reusability vs replicability vs reproducibility
    - Specifying dependencies (package versions)
    - Draft of an Authors’ Checklist
      - ie required sections, citation formatting, graphic limits, etc
  - Note that not all communities are ready to use computation notebooks - instead, think of this as guidance to get them there
  - A good place to start communicating to: People who are ready or already use notebooks
  - Outcomes of those initial outputs will influence people to use them too
  - **3rd**: What elements does a notebook have to support to be part of the academic record?
    - Citations
    - Figure numbering
    - Navigation; table of contents for example
    - Keywords
    - Equations
    - Envision a chart; what supports what?
  - Authors should have a test or preview environment to ensure that their output works before publishing
  - Consider review process: what do authors provide? What do reviewers evaluate?
  - Question: Are we aiming for a product attached to a journal article, or an independent product?
    - It’s independent
    - for scoping, the goal is to submit a Computational Notebook into an editorial workflow: it is the paper with code & other projects embedded
      - Value in citing the notebook already in process; we are instead looking at *submitting the notebook as a paper*
    - citing a notebook is not difficult, but - “is the notebook actually going to run?”. That’s the tricky part



## Submission / Kenton

- Provide guidelines to submitters, not just infrastructure
  - Templates. Latex, Doc
  - General for all notebooks, or narrow down
- What can we leverage that is already there?
  - “CookieCutter” open source package
- Overlap w review: ability to run notebook, or does the reviewer DL and run?
- Metadata:
  - Reuse as much as possible
  - Software metadata?
- Citations are not handled well by current notebook systems
- “Why would researchers publish notebooks?”
- What are we making? A guide for publishers? Something to cite ourselves? “The Turing way” (?)
- crucial to work cross-functionally
  - Facilitate collaboration
- meeting cadence: once per month is sufficient

## Editorial process / Shelley Stall and Lorena Barba

- Ensure notebook is well-documented
  - Configurations
  - Dependencies
  - Must come as full package to the review cycle
- Review cycle deliverables:
  - Reviewers are asked to look at a product they have never seen before
  - How to involve existing reviewers?
    - Skillsets
    - Guidelines
  - What is the infrastructure for hosting a Computational Notebook?
    - What tools will the reviewer need?
    - Are they able to edit while interacting with it?
    - Protections for identity management
  - Ethical considerations
    - Protecting editors, reviewers
- Defining feature of Computational Notebooks:
  - **Interactivity**
  - Different from a static paper, especially during editorial and peer-review
  - Can execute code
  - Error checks must happen before review
- Security issues:
  - If editors/reviewers are running code, system needs protection
  - Reviewers must pledge to keep the code confidential
  - But other considerations: NFTs, mining bitcoin...

- AGU trying to push engagement with reviewers across globe
  - Incl regions that don't have resources to perform these computations
  - How do we ensure equity?
    - explore what the reviewer is getting from their service
    - Review may or may not take longer than a regular article
    - Questions
      - What happens when code in the Notebook depends on...
        - Private data?
        - High performance computing?
      - How do we let reviewers vet those notebooks?
        - same problems with a normal manuscript. Obscured data for instance
        - Each journal will have solutions particular to their field
        - Begg the question: if you can't run the notebook, why would you create - or submit - a notebook?
          - "It's not that you can't run it, it's that the reviewer can't."
          - if the notebook cannot be executed, why would you want it reviewed?
          - your colleagues can review/run it - but not *everyone* can
      - I frame this to my students like "reproducibility is nonbinary and not one-dimensional", a complex space - not a checkbox
        - Facilitate collaboration. Allow for people to build upon the output
        - If you can't re-run the model, you can include a reduced version of the dataset that users can start with at a later point in the process
        - If you can't share data, you can share synthetic data to evaluate the algorithm

#### Production / Alberto Pepe and Rowan Crockett

- A post acceptance world: we just need to publish this thing, not worry about reviewing.
  - Previous workstream produces a format that is already good to go
- Limitation: current publishing workflows are built to a different purpose
  - How do we use them for our purpose?
  - How do we think out of the box?
  - What other systems - not publishing - can we use?
    - Ie, copyediting tools will never take an XML file with embedded notebook/code
- Deliverables:

- Simplest: Publish as plain text
  - Better than nothing
  - Like a snippet
  - Easy to add to the scholarly record
- Or, incrementally, enable workflow
  - Compile parts of the workbook appropriately:
    - Text
    - Executable code
- Focus questions, primarily around archivability:
  - How do we preserve structured data and convert it to XML using standards?
  - What types of metadata do we need to include?
    - Environment requirements
    - Runtimes
    - Pointers to existing data
  - How does what we capture degrade over time?
    - Over time, a notebook is archived and returned to a static resource
    - “Graceful degradation”
- Question: how do we keep the size/scope/depth of the notebook in check?
  - Ex, a paper is 20,000 words
  - What is the version of this called?
- We still need the interactivity to usable for readers
  - Alberto: we need to think out-of-the-box
- the ideal solution is to take the “richest” piece of content (the Notebook) and structure it in a way that one can derive products from it to support services, ie
  - DOI registration
  - Readability
  - Computational environment
- so, what is the structure to enable scalability/tractability?
- we want to be sure that the output can be cited
- There are ways to upgrade the existing formats, ie Jupyter Notebooks and markdown
  - This would break the markdown into much more structured data; how to integrate between the two formats
- Jupyter is a richer format than the notebook; can for instance ingest Bibtext
  - We should consider above a notebook, but like a Jupyter notebook; handle refs
- envisions a Jupyter Notebook assembled out a repository, entry point of a narrative; but the Notebook should not have a *lot* of code but should instead summarize the results

- A scientific “story” should ask a question
- Key point: pulling in results, pulls in other factors but those other factors have their own literary form
- “Two types of narrative being woven”
- EarthCube has two kinds of notebooks;
  - Narrative that asks the question and shows how to demonstrate
  - Demonstrate a technology independent of a particular question
- What does our scope cover?
- archivability with lots of dependencies?
  - How do we ensure they (dependencies) are preserved?
  - the community will maintain the most important things that need to be kept up-to-date for reference in the far future
  - What does it mean to be as archivable as possible?
- Are these complimentary to journal articles, or do they stand on their own? What will they become?
- suggests analogizing a variety of outputs to different existing manuscript types (to navigate this brand new ms type)

## PARKING LOT

1. Issue of accessibility
  - a. How can we ensure the notebook publication is accessible to people with special assistance need?
2. Issue with long term archiving per federal/regional policy requirement
  - a. What file formats do those archives currently support?
3. Notebook articles: towards a transformative publishing experience in nonlinear science <https://arxiv.org/pdf/2103.05770.pdf>

## Workstream Summaries & Discussion

### Day 2 Group Discussion on Workstream Charters

- Requirements for large computational resources
- Pros & Cons of current legacy infrastructure
  - Disruptive model that was lighter (didn't rely on current infrastructure).
    - Still need to explore workflow, copy editing, references, etc.

- All tools will be explored eventually, should choose what fits this specific community and can support/interact with other features.
  - Want to be as inclusive as possible.
  - Notebooks could lead to a more robust publishing ecosystem if designed correctly.
  - Accessibility will be complex, but needs to stay a concern.
- No Supplementary Material with Notebook submission
  - Need to explore pros & cons, and decide where data goes.
  - Any required data should have a repository with DOI that can be interacted with.
  - Software ideally submitted/reviewed with DOI.
  - Length/size limits will pose a challenge.
- Peer-review process
  - How will reviewers interact/compute?
  - Ethical/policy concerns possible.
- Milestones
  - Short-term
    - Review of platform and tools
  - Medium-term
    - Metadata
      - Paper (authors, institutions, etc.)
      - Software
      - Environment (to run)
      - Data
  - Reproducibility
    - At least for now, possibly long-term.
- [Presubmission](#)
  - Define necessary components for support
  - Engage with early adopter communities
  - Minimum viable product
    - Is true computability necessary? (/ how much \_work\_ for computability)
    - Guide authors by highlighting what is acceptable
    - “Author’s checklist”
  - Materials for researchers on why/how to share computational notebooks

## Mission Statements:

### Presubmission:

The purpose of this working group is to describe the technical and cultural/ behavioral / educational challenges and complications relating to pre-submission of notebooks, considering key stakeholders (authors, reviewers, publishers, funders, readers, and infrastructure providers including datastewards/RSEs); defining how to enable authors

to prepare manuscripts composed of computational notebooks for submission and review.

### **Submission & Metadata**

The purpose of this working group is to develop and provide recommendations on how notebooks can be submitted as a manuscript including a reflection on existing challenges. The goal is to guide authors to prepare notebooks for a submission. This includes defining a set of minimum requirements for a submission and will reflect on how to enrich a submission with for example metadata, software requirements or data. This working group will closely collaborate with group 1 and group 3.

### **Editorial & Peer Review**

The purpose of this working group is to define the requirements to enable a journal to support interactive and executable use of a notebook format as a manuscript submission, for the purpose of effective and equitable peer review. This includes identifying technologies, requirements, and best practices for facilitating the review process (e.g. annotations and track changes) during the notebook publication process, as well as computing requirements to facilitate computational notebook review. This workstream will consider key stakeholders; describe any existing solutions that need to be scaled / adopted / adapted; and identify any collaboration needed with other working groups.

### **Post Production**

The purpose of this working group is to define a concrete roadmap for including, hosting, preserving, and executing non-article research outputs such as computational notebooks in post-acceptance publishing workflows. Identify the publishing, metadata, hosting, and archiving requirements necessary to enable replicability and ultimately scientific reuse of scientific publications.

### **Content Consumption and Accessibility - User experience?**

The purpose of this working group is to:

- Reading & Accessibility
- Executing in federated “hubs” that are close to research infrastructure
- Librarians
- “Reading Rooms”
- Capability to “fork” a published notebook and build upon that work.

- 
- LOCKSS/CLOCKSS <https://clockss.org/wp-content/uploads/2020/02/Comparison-CLOCKSS-GLN-Portico-30July2019.pdf>

- Dark archive; not intended for interaction
- For instance, library canceling a subscription and receiving a backfile
- However: these preservation services are in case a journal fails
  - Ie Portico
  - A time capsule
- Partner with services to provide “submit to journal” service
  - Lightweight solution - link back to the sources
  - Is this what others are envisioning? Or something more elemental/raw - uploading a zip file

## Day 2 (Nov 4) Next steps, wrap-up

- Steering Committee <https://data.agu.org/notebooks-now/2022/08/02/meet-steering-committee.html>
- Plus working group co-chairs will meet in the next couple of weeks to discuss next steps
- Notional time for the working groups is 6-8 months