

# **Earth and Space Science**

#### **EDITORIAL**

10.1029/2023EA003458

#### **Key Points:**

- Notebooks Now! is a workflow for the submission and publication of computational notebooks as the primary element of the scientific record
- The model fulfills the findable, accessible, interoperable, and reusable principles of scientific publication
- This project commits to Open Science, and is part of AGU's contribution to openly accessible sustainable development

**Correspondence to:** 

B. Sedora, BSedora@agu.org

#### **Citation:**

Caprarelli, G., Sedora, B., Ricci, M., Stall, S., & Giampoala, M. (2023). *Notebooks Now!* The future of reproducible research. *Earth and Space Science*, *10*, e2023EA003458. https://doi. org/10.1029/2023EA003458

Received 29 NOV 2023 Accepted 30 NOV 2023 Corrected 20 FEB 2024

This article was corrected on 20 FEB 2024. See the end of the full text for details.

#### **Author Contributions:**

Conceptualization: Graziella Caprarelli, Brian Sedora Supervision: Brian Sedora Visualization: Graziella Caprarelli Writing – original draft: Graziella Caprarelli, Brian Sedora, Mia Ricci Writing – review & editing: Graziella Caprarelli, Brian Sedora, Mia Ricci, Shelley Stall, Matthew Giampoala

© 2023. The Authors. Earth and Space Science published by Wiley Periodicals LLC on behalf of American Geophysical Union.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

# Notebooks Now! The Future of Reproducible Research

Graziella Caprarelli<sup>1</sup> <sup>(D)</sup>, Brian Sedora<sup>2</sup> <sup>(D)</sup>, Mia Ricci<sup>2</sup> <sup>(D)</sup>, Shelley Stall<sup>2</sup> <sup>(D)</sup>, and Matthew Giampoala<sup>2</sup> <sup>(D)</sup>

<sup>1</sup>Centre for Astrophysics, University of Southern Queensland, Toowoomba, QLD, Australia, <sup>2</sup>American Geophysical Union, Washington, DC, USA

**Abstract** In this paper, we introduce a new ambitious project led by AGU in collaboration with Wiley and Curvenote and a team of geoscientists, computer and data scientists. The project, aptly called Notebooks Now!, aims to develop an end-to-end workflow for the preparation, submission, peer-review and publication of Earth and space science papers, based on computational notebooks, specifically Jupyter Notebooks or R Markdown. The idea, first proposed by Dr. Peter Fox, Editor in Chief of *Earth and Space Science* from 2019 to early 2021, is to revolutionize the way data- and computation-rich scientific research is performed and published. The vision is to establish the computational notebook as the centerpiece of the scientific work, around which the manuscript is then written. In this model, the notebook is the primary publication, accompanied by a static "traditional" paper as the enduring archive. Advantages of this model include: ease of editing of data and code during manuscript writing and revision; open access to the same by reviewers and readers; prompt reusability and repurposing of research outputs by other researchers; access to sophisticated computational tools by the global community; "multiplier effect" of the research. The model thus fulfills the findable, accessible, interoperable, and reusable principles of scientific publishing, and commits to Open Science. This is part of AGU's strategic initiative to accelerate openly accessible scientific contributions to sustainable development.

**Plain Language Summary** Transparency and reproducibility are the fundamental principles of publication of scientific research. In compliance with these principles, data and methods described in papers published in AGU's journals must be findable, accessible, interoperable, and reusable (FAIR). This is especially important for data-rich and computationally-intensive studies. *Notebooks Now!* is a project based on computational notebooks as the primary publication output, that will fulfill the requirement for FAIR. The project will also advance AGU's commitment to Open Science, the global effort to increase participation and access to scientific research. Here we report on the current progress of the project.

#### 1. Introduction

In 2022, AGU was awarded a grant from the Alfred P. Sloan Foundation (Grant ID 10070) to develop an end-to-end workflow for the submission, peer review, and publication of computational notebooks as the primary element of the scientific record. This ongoing effort is now known as Notebooks Now! (NN!) and it involves AGU, Wiley, Curvenote (https://curvenote.com/), and a large group of community stakeholders, geoscientists, computer and data scientists. The NN! project aims to smash existing barriers to the use and reuse of big data and code, thus fulfilling the findable, accessible, interoperable, and reuseable (FAIR) principles to which all AGU journals abide, and AGU's strategic commitment to Open Science and sustainable development. The effort stems directly from the vision and work of Dr. Peter Fox, Editor in Chief of *Earth and Space Science* from 2019 to early 2021. The following sections highlight some of the salient principles underpinning the *Notebooks Now!* project, as well as its history. We believe this is an inspirational project, which will break new ground in the way peer-reviewed scientific research is published.

#### 2. Background and Progress to Date

All AGU journals adhere to the principles of FAIR: "Findable," "Accessible", "Interoperable", "Reusable" (Wilkinson et al., 2016). These definitions represent guiding principles toward achievement of digital reproducibility and reusability of data and code linked to a published paper. But what does it mean for a journal to operate in full compliance with FAIR? The principal aim of *Earth and Space Science* is to publish new data and novel methods directly linked to examples of application to scientific questions (https://agupubs.onlinelibrary.wiley. com/hub/journal/23335084/aims-and-scope.html). Thus, FAIR principles are at the very core of its mission:



data, metadata and code must be easily findable and openly accessible through a unique digital identifier, the software used to process and analyze data should be non-proprietary (as much as possible) to facilitate reuse by the scientific community in a non-exclusive manner, the code and algorithms should be adequately documented and explained and ready to be applied to similar projects. From a scientific perspective, the combination of these qualities fulfills the requirement of transparency and reproducibility of published research (Gil et al., 2016). From an ethical perspective, they advance us one step closer to Open Science (https://www.agu.org/learn-about-agu/about-agu/open-science), the global effort to increase participation and access to scientific research.

Achieving FAIR and Open Science is not without challenges, however: what online repository can be "trusted" to permanently host data and methods (script, workflows, etc.) openly and accessibly? What restrictions are imposed on the dissemination of data and methods by the authors' organizations? How can we recognize the additional time and resources required by the authors to provide all the information needed to properly archive data and methods in an online repository? And overall, what are the costs associated with these efforts? There is no easy answer to these questions, and the model may not be unique, strongly depending on the type of research published.

A new way of thinking about FAIR and Open Science was promoted and actively researched by Dr. Peter Fox, Editor in Chief of *Earth and Space Science* from 2019 to early 2021. Throughout his career, Dr. Fox worked to challenge and advance the way scientists acquire, process, analyze and interpret data. He collaborated with NASA, NOAA, USGS, and other organizations, to explore ways in which published papers could be enhanced by direct links between the papers and the data and algorithms underpinning results and interpretations. He was therefore inspired to shift the publication paradigm, specifically to move toward a model whereby computational tools could feed directly into the preparation, submission, peer-review, publication and archiving of papers (Fox, *pers. comm.*).

A demonstration of this new publication paradigm is found in two seminal papers by Iacovino et al. (2021) and Elkins and Spiegelman (2021). The first paper presents Volatile Equilibria and Saturation Identification calculator, a comprehensive tool to model the solubility of water and  $CO_2$  in magmas. Here, the authors explain the properties of the program, which is provided as python code, ready to be incorporated in other scripts. The second paper presents the theoretical underpinnings and computational tools for the forward melting modeling of uranium-series disequilibria in basalts. The static versions of both papers are published in the journal, but they also contain the link to their Jupyter notebook versions, which readers can use to run the code for their research without the need to compile a python script. Importantly, the notebooks were reviewed as part of the publication process. From these beginnings, stems the inspiration for *Notebooks Now!* 

The current project flips the script to deliver the notebook as the version of record, while providing a static pdf as a secondary product. AGU invited experts from the computational notebook and publishing communities to join our efforts, first to identify the challenges involved in building and implementing this workflow at a November 2022 workshop, and then to begin to answer those questions and plot a path to a minimum viable product that can be pilot-tested, by joining working groups that looked at specific parts of a potential workflow: (a) Pre-submission; (b) Metadata; (c) Editorial and Peer-review; (d) Production and Post-production; and (e) Content Consumer. These efforts led to the development of a workflow that relies on a Manuscript Exchange Common Approach (MECA) bundle to package the contents of a notebook to usher it through peer-review and production systems. Currently, the project is progressing on schedule and at the stage of testing the technical workflow for peer review and publication of papers. The final stage of the project will involve soliciting new manuscripts through this journal, to run through the entire process of manuscript preparation, submission, peer-review and publication.

# 3. Concluding Statement

We are excited about the NN! project, which represents a revolutionary approach to the publication of scientific papers: primary research outputs are shared within the *Notebooks Now!* platform, and are directly linked to publication outputs. The principles of FAIR are applied, transparency and reproducibility are integral to the publication from the beginning of the research, data and code can be tested by peer-reviewers in the computational notebook sandbox, and, once the paper is published, data and code can be reused and repurposed directly from within the publication. Thus, the publication itself becomes a live research laboratory, from which new scientific

insights and new research stems. The positive externalities of this new publication paradigm are clear: scientists can promptly validate research outputs, test its boundaries, reuse data, and apply or adapt methods to similar problems without the need to transfer "chunky" data sets and software across servers. Complex computational tools are open to the global scientific community, thus enabling diverse research. Collaborations across groups and countries are facilitated, cross-fertilization of ideas drives the next steps of scientific investigation, and inclusivity becomes the operational default in the practice of peer-reviewed scientific publication. In turn, this feeds directly into a "multiplier effect" of the published research, which because of its transparency, reproducibility, and accessibility, is open to global communities outside of academia and research center environments. Ultimately, the NN! project will make the practice of scientific publication central to the fulfillment of many of the United Nations Sustainable Development Goals.

## **Data Availability Statement**

Data were not used nor created for this article.

## References

- Elkins, L. J., & Spiegelman, M. (2021). pyUserCalc: A revised Jupyter Notebook calculator for uranium-series disequilibria in basalts. *Earth and Space Science*, 8(12), e2020EA001619. https://doi.org/10.1029/2020EA001619
- Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., et al. (2016). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science*, *3*(10), 388–415. https://doi.org/10.1002/2015EA000136
- Iacovino, K., Matthews, S., Wieser, P. E., Moore, G. M., & Bégué, F. (2021). VESIcal Part I: An open-source thermodynamic model engine for mixed volatile (H<sub>2</sub>O-CO<sub>2</sub>) solubility in silicate melts. *Earth and Space Science*, 8(11), e2020EA001584. https://doi.org/10.1029/2020EA001584
  Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific

data management and stewardship. Scientific Data, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

# **Erratum**

In the originally published version of this article we inadvertently omitted Chris Erdmann and Laura Lyon from the Acknowledgments. Additionally, we have listed the members of the Notebooks Now! Steering Committee and the Notebooks Now! Working Group Chairs directly in the acknowledgments. The Acknowledgments have been corrected, and this may be considered the authoritative version of record.

#### Acknowledgments

The authors thank the Notebooks Now! Steering Committee members: Lorena Barba, Yanina Bellini, Alison Presmanes Hill, Chris Holdgraf, Kenton McHenry, Alberto Pepe, Fernando Perez, and Karthik Ram; the Notebooks Now! Working Group Chairs not yet mentioned: Jeffrey Beck, Tony Castronova, Rowan Cockett, Rae De Guzman, Rachel Kurchin, James Munroe, Steve Purves, Doug Schuster, Sam Teplitzky, and Julia Wagemann; the participants of the workshop and working groups; the Curvenote team not yet mentioned: Mike Morrison; the AGU team: Brooks Hanson, Jeanette Panning, Erin Syring, Susan Townsend, Nick Violette, Kristina Vrouwenvelder, Chris Erdmann, and Laura Lyon; the Wiley and EJournalPress team not previously mentioned: Jenn Satten, Al Troyano, Jamie McReickel, Emily Otterson, and Joel Plotkin; the Posit team: J.J. Allaire and Charles Teague; and most especially the AGU authors Kayla Iacovino and Lynne Elkins for supporting this work by being the first showcased articles in the Notebooks Now! format hosted on Curvenote's platform. We are forever grateful to Dr. Peter Fox for having the vision to publish computational notebooks as a new peer-reviewed journal format in AGU journals.