Jupyter Book-based Supplemental Material: a FAIR Practice to Connect Research Articles with Scientific Data

Whyjay Zheng[1], Fernando Perez[1], Chris Holdgraf[2], Tasha Snow[3], Shane Grigsby[4], Facundo Sapienza[1], Erik Sundell[5], Matthew Siegfried[3], Jonathan Taylor[6], and Executable Books Community[7]

[1]University of California, Berkeley
[2]2i2c.org
[3]Colorado School of Mines
[4]NASA Goddard Space Flight Center
[5]Sundell Open Source Consulting
[6]Stanford University
[7]Affiliation not available

December 27, 2023

Abstract

Supplemental material (SM; also known as supplementary information) comes with its associated research article and provides study details such as metadata, additional figures and text, multimedia, and code. Well-designed SM helps readers fully understand the underlying scientific analysis, reproduce the work, and even reuse the workflows for exploratory ideas. Thus, the concept of FAIR (Findable, Accessible, Interoperable, and Reusable), which is originally designed for data sharing guidelines, also matches these core qualities for SM.

We evaluate different SM-preparation practices that are commonly found in Earth Science journal articles. These practices are classified into five tiers based on the FAIR principles and the narrative structure. We show that Jupyter Book-based SM belongs to the top tier and outperforms the other practices, despite being not as popular as the other SM-preparation practices as of 2022.

We identify the advantages of the Jupyter Book-based SM as follows. Jupyter Book uses a narrative structure to combine different elements of SM into a single scholarly object, increasing readability. Jupyter Book's direct support of HTML publishing allows users to web host the SM using services such as Github Pages, improving the web indexing ranks and resulting in higher exposure of both the research article and the SM. The entire SM is also eligible to be archived in a data repository and receive a Digital Object Identifier (DOI) that can be used for citations. In addition, Jupyter Book-based SM lowers the threshold of reproducing and reusing the work by accessing an interactive cloud computing service (e.g., MyBinder.org) with all data and code imported if the content is available on a code-hosting platform (e.g., Github).

These features summarize the core values of SM from the perspective of open science. We encourage researchers to use these good practices and urge journal publishers to be open to receiving such supplements for maximum effectiveness.

content goes here

Use Jupyter Book for your supplemental material!

Traditional supplemental material	jupyter {book} -based supplemental material
❌ Have to access the journal article first	✔ Independent and citable DOI
❌ Need to download files	✔ Online HTML; easy to share
❌ Invisible by direct online searching	✔ Can be indexed online (better SEO)
❌ Hard to find relevant data & info [1]	✔ All info interconnected in one place
❌ **Low Reproducibility**	✔ Higher presentation quality
	✔ Multimedia compatible
	✔ Executable on the cloud (optional)
	✔ **Better Reproducibility**

What is Jupyter Book?

Jupyter Book (jupyterbook.org) is a software tool that compiles Jupyter Notebooks and MyST markdown files (myst.tools) into fully structured written work in various printable and digital formats, such as PDF, ePub, and HTML.

Check out this QR code for an example showing real supplementary material built with Jupyter Book. The accompanying paper is published in *The Cryosphere*. [2]

[1] Greenbaum et al., 2017. 10.1186/s13059-017-1205-3
[2] Zheng, 2022. 10.5194/tc-16-1431-2022
[3] According to Stodden et al. (2018, 10.1073/pnas.1708290115), only 26% of the articles in *Science* are fully reproducible.
[4] According to Gabelica et al. (2022, 10.1016/j.jclinepi.2022.05.019), the chance to get the requested data is likely < 10%.

How does the new method work?

If the source files are hosted on GitHub, we can use GitHub Actions to make Jupyter Book automatically build and publish the written work to GitHub Pages. With a proper data source setup (e.g., Zenodo), we can further use a cloud computing environment (e.g., mybinder.org) to execute and reproduce all the results presented in the written work.

What are the current practices?

The current practices for preparing supplemental information can be classified into five categories based on how much effort is needed to reproduce the presented work. The results resemble a pyramid shape with ~⅓ of papers sticking with "data and code available upon request." (also [3])

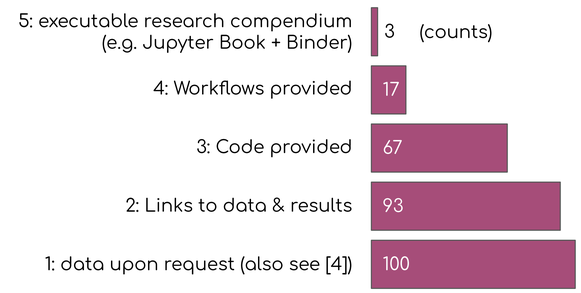

Figure 1. "Level of reproducibility" for all articles published in *The Cryosphere* between June 2021 and May 2022 (N=280).

Jupyter Book-based Supplemental Material: a FAIR Practice to Connect Research Articles with Scientific Data

Whyjay Zheng[1], Fernando Pérez[1], Chris Holdgraf[1,2], Erik Sundell[3], Matthew Siegfried[4], Tasha Snow[4], Shane Grigsby[5], Facundo Sapienza[1], Jonathan Taylor[6], Executable Books Community

[1]UC Berkeley
[2]2i2c.org
[3]Sundell Open Source Consulting
[4]Colorado School of Mines
[5]NASA Goddard
[6]Stanford University

This work is part of the Jupyter meets the Earth project, supported by the NSF EarthCube program (awards 1928406 & 1928374).

#AGU22
ED12C-0369