From reproducible to reusable: Lessons for neuroscience from data science

Elizabeth DuPre & Jean-Baptiste Poline

We are at the beginning of a necessary evolution from reproducible to reusable analyses (Peng, 2011). This will enable neuroscientists to examine the generalizability of their findings (Whitaker, 2017) and to increase the overall efficiency of science.

Here, we provide **four key recommendations** for neuroscientists from data science best practices to increase the reusability of their analyses.

For a more comprehensive introduction to reproducible data science, we recommend **The Turing Way**.

① Licensing

Unlicensed research products are unusable. For common research outputs, we recommend:

For **code,** MIT License
For **data**, Public Domain Data License (PDDL)
For **manuscripts**, Creative Commons Attribution

For **everything else**, we recommend that researchers use **https://choosealicense.com** to select the most appropriate license.

② Modular code

Breaking code down into minimal units that do not depend on one another makes code easier to understand, test, refactor, and maintain (Wilson et al., 2017).

Modular code also means that the same code can be re-used across different analysis projects with minimal updating.
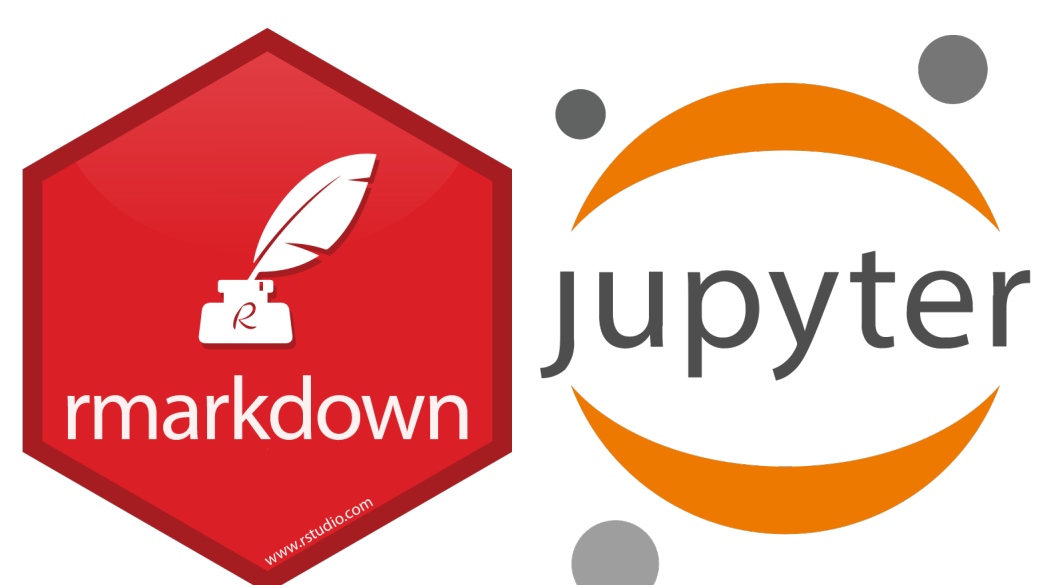
③ Documentation

Documentation should be an integral part of software, not supplementary material.

There are many options for generating and hosting documentation, including **Read the Docs** and **MkDocs**.

Manuscripts can also be considered a kind of documentation for software if the code and data are tightly linked. Tools like **RMarkdown** and **Jupyter Book** help to make this happen.

④ Community-driven standards

By adopting community-driven standards, we can move towards sustainable interoperablity.
We suggest that neuroscientists consider:

Teaching Integrity in Empirical Research (TIER) Protocol for their **project management**,

the **Brain Imaging Data Structure (BIDS)** for their **data management**, and

SciCrunch for their data annotation.