Open Methodology

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato@scholar.social

Open Science (A.Y. 2024/2025)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

CC BY

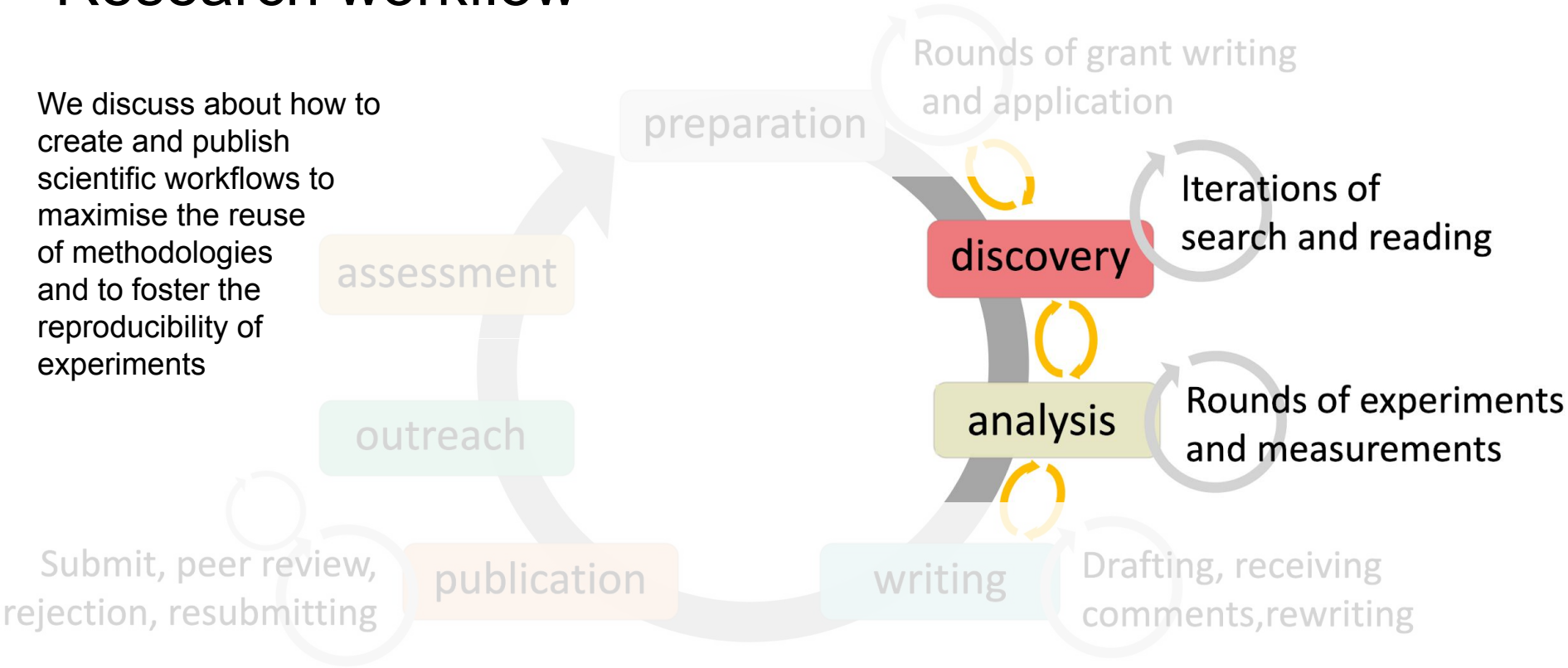
DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA

Discussion

Chigbu, U. E., Atiku, S. O., & Du Plessis, C. C. (2023). The Science of Literature Reviews: Searching, Identifying, Selecting, and Synthesising. Publications, 11(1), 2. https://doi.org/10.3390/publications11010002

Research workflow

We discuss about how to create and publish scientific workflows to maximise the reuse of methodologies and to foster the reproducibility of experiments



Kramer, B., & Bosman, J. (2015, June 18). The good, the efficient and the open—Changing research workflows and the need to move from Open Access to Open Science. CERN Workshop on Innovations in Scholarly Communication (OAI9), University of Geneva, Geneva, Switzerland. https://www.slideshare.net/bmkramer/the-good-the-efficient-and-the-open-oai9

What is a methodology

A methodology is the definition and description of the methods used in a study to collect and analyse the data necessary to answer to specific research questions

In research articles, usually it is introduced within the “Methods” section, which “should provide the readers with sufficient detail about the study methods to be able to reproduce the study if so desired”

Such a section “should be specific, concrete, technical, and fairly detailed” and should contain information about “[t]he study setting, the sampling strategy used, instruments, data collection methods, and analysis strategies” adopted to answer the research questions

Perneger, T. V. (2004). Writing a research article: Advice to beginners. International Journal for Quality in Health Care, 16(3), 191–192. https://doi.org/10.1093/intqhc/mzh053

What is an (open) methodology

An open methodology is a methodology described in sufficient detail to allow other researchers to repeat the work and apply it elsewhere

It is not always possible, of course, to access the same unique resource of the original study, e.g. a specific computer hardware

However, it is important to clearly describe the full methodology because it is the only way to enable reproducibility and to allow others to learn from what scientists have done

Crucial aspect: to reveal how a scientist carried out an experiment is at the heart of any study

Watson, M. (2015). When will ‘open science’ become simply ‘science’? Genome Biology, 16(1), 101. https://doi.org/10.1186/s13059-015-0669-2

Reusing existing research

The “discovery” step is an important aspect related to the definition of the methodology one wants to define to address research questions

It is possible that prior works have addressed either the same or similar research questions, or that the methodology proposed in a particular research can be reused to address, in principle, different research questions

Thus, it is crucial to perform a literature review to understand whether prior work can be of any help – following the principle of standing on the shoulders of Giants

A literature review should provide a concise examination and discussion of evidence in a particular area and is usually present in several written research outputs such as proposals for funding, proposals for academic degrees, research articles, guidelines for professional practice, and reports

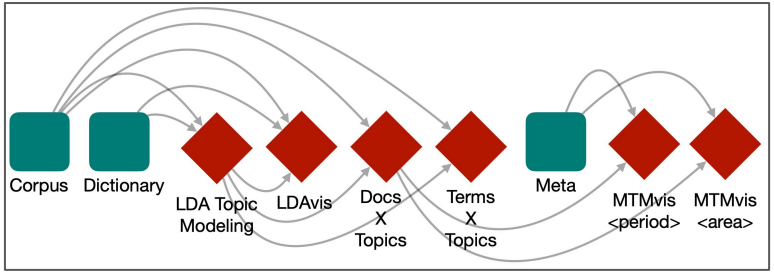
Bolderston, A. (2008). Writing an Effective Literature Review. Journal of Medical Imaging and Radiation Sciences, 39(2), 86–92. https://doi.org/10.1016/j.jmir.2008.04.009

Workflow

A workflow is a way to define an experiment (and the methods characterising it) as a directed graph where the nodes represent operations and edges specify dependencies between the operations

Once specified, a workflow can be reused by other scientists and enables the understanding an experimental process, the replication a previous experimental result, or its repurposing as a building-block in the design of new workflow-based experiments

Issue: workflows may decay, i.e. could not be understood, or executed when downloaded



Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., Mina, E., Corcho, O., Gómez-Pérez, J. M., Bechhofer, S., Klyne, G., & Goble, C. (2015). Using a suite of ontologies for preserving workflow-centric research objects. Journal of Web Semantics, 32, 16–42. https://doi.org/10.1016/j.websem.2015.01.003

Crucial aspects for defining workflows

Example data inputs: make available exemplar input data accompanied by their description that enable the successful execution of the workflow, otherwise both experiment reproducibility and the ability to understand the function the workflow is inhibited

Long-term archivability: preserve the workflow in appropriate systems together with provenance traces of its data results, to track how results were produced by the workflow and to enable repairing broken workflows

Appropriate documentation: describe, in natural language, all the steps of the analysis, including its inputs, intermediate steps, and outputs, and make them annotable, since insufficient documentation impairs the runnability and understandability of workflows

Trackability of changes: use tools and approaches to identify changes of a workflow (including in the data and the environments used to run it), which is of particular importance when third party resources, that are not under the author's control, are used – the goal is to allow a user

- to retrieve the original version of input data and environment setting in order to reproduce/verify the original results
- to retrieve the different parameter configurations used to generate the different versions of outputs
- to identify the changes made to the workflow specification

Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., Mina, E., Corcho, O., Gómez-Pérez, J. M., Bechhofer, S., Klyne, G., & Goble, C. (2015). Using a suite of ontologies for preserving workflow-centric research objects. Journal of Web Semantics, 32, 16–42. https://doi.org/10.1016/j.websem.2015.01.003

Computational workflows

In computational workflows each task represents the execution of a computational process, such as running a piece of code, using a service, calling a tool, accessing to a database, submitting a job to a compute cloud, executing data processing script, etc.

Computability is an added value to have for fostering the reproducibility of workflows

Python notebooks can be used to create reproducible computational workflows that can be run within the Jupyter environment

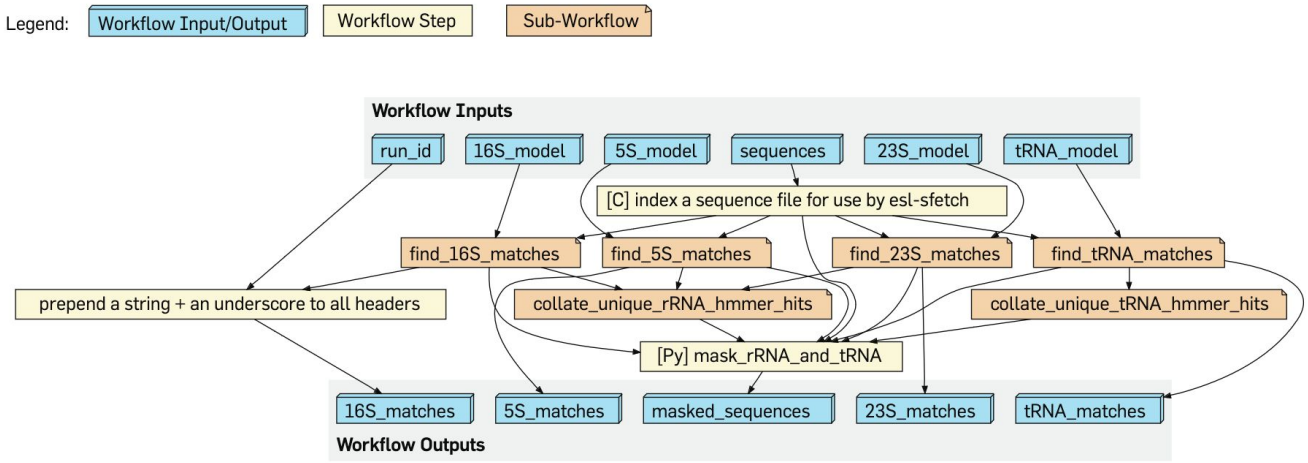
There are several tutorials that demonstrate how to use Jupyter for this activity, such as the one made available by Justin Kitzes, Shane Grigsby, Mark Wilber of the University of California and Santa Barbara

Beg, M., Taka, J., Kluyver, T., Konovalov, A., Ragan-Kelley, M., Thiery, N. M., & Fangohr, H. (2021). Using Jupyter for Reproducible Scientific Workflows. Computing in Science & Engineering, 23(2), 36–46. https://doi.org/10.1109/MCSE.2021.3052101, freely available at https://arxiv.org/abs/2102.09562

Common Workflow Language

The Common Workflow Language (CWL) is a set of open standards for describing and sharing computational workflows

It support the specification of critical concepts such as automation, scalability, abstraction, provenance, portability, and reusability, and it has been developed around core principles of community and shared decision making, reuse, and zero cost for participants



Crusoe, M. R., Abeln, S., Iosup, A., Amstutz, P., Chilton, J., Tijanić, N., Ménager, H., Soiland-Reyes, S., Gavrilović, B., Goble, C., & Community, T. C. (2022). Methods included: Standardizing computational reuse and portability with the Common Workflow Language. Communications of the ACM, 65(6), 54–63. https://doi.org/10.1145/3486897

Registered reports

Recently, a new approach to journal publishing has been introduced in order to foster the definition of better methodologies in research environments, i.e registered reports

1. The authors submit a registered report to a journal
2. The report is evaluated via the usual peer-review process, to obtain constructive criticisms and recommendations by peers (i.e. the reviewers) to improve the protocol proposed
3. The revised report is provisionally accepted and published in the journal as an article
4. The authors perform the experiment and expand the journal article to include both the results and discussion
5. The new part of the article is again peer-reviewed but, providing that the authors have implemented the agreed protocol, publication should be guaranteed

Benefits: minimising the potential for publication bias when publishing negative vs positive results and checking the soundness of the methodology before running the experiment

Clarke, P., Buckell, J., & Barnett, A. (2020). Registered Reports: Time to Radically Rethink Peer Review in Health Economics. PharmacoEconomics - Open, 4(1), 1–4. https://doi.org/10.1007/s41669-019-00190-x

Pure workflow publications

In order to support increased sharing of open research methodologies, some journals (e.g. PLOS One and MethodsX) have started to publish directly methodology articles as part of their portfolio without having them explicitly linked with the results of an experiment

Some justification for this choice:

- methods development and sharing deserve increased recognition, via a peer-reviewed publication
- understanding the exquisite details that make methods work is essential for reproducibility and for accelerating science
- researchers who develop methods deserve more recognition

Hrynaszkiewicz, I. (2020, December 7). Show your work. Peer-Reviewed Protocols. The Official PLOS Blog. https://theplosblog.plos.org/2020/12/show-your-work-peer-reviewed-protocols/

Mandatory text to read for the next lecture

Papanas, N., & Mikhailidis, D. P. (2024). Alice through the Looking-glass: Can We Improve Peer Review? The International Journal of Lower Extremity Wounds, 23(3), 356–359. https://doi.org/10.1177/15347346221084784

End

Open Methodology

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato@scholar.social

Open Science (A.Y. 2024/2025)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

CC BY

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA