






Foundations of a Framework for Peer-Reviewing the Research Flow

Alessia Bardi^(✉) , Vittore Casarosa , and Paolo Manghi 

Institute of Information Science and Technologies - CNR, Pisa, Italy
{alessia.bardi,vittore.casarosa,
paolo.manghi}@isti.cnr.it

Abstract. Traditionally, peer-review focuses on the evaluation of scientific publications, literature products that describe the research process and its final results in natural language. The adoption of ICT technologies in support of science introduces new opportunities to support transparent evaluation, thanks to the possibility of sharing research products, even inputs, intermediate and negative results, repetition and reproduction of the research activities conducted in a digital laboratory. Such innovative shift also sets the condition for novel peer review methodologies, as well as scientific reward policies, where scientific results can be transparently and objectively assessed via machine-assisted processes. This paper presents the foundations of a framework for the representation of a peer-reviewable research flow for a given discipline of science. Such a framework may become the scaffolding enabling the development of tools for supporting ongoing peer review of research flows. Such tools could be “hooked”, in real time, to the underlying digital laboratory, where scientists are carrying out their research flow, and they would abstract over the complexity of the research activity and offer user-friendly dashboards.

Keywords: Open peer review · Digital science · Open Science

1 Introduction

An increasing number of researchers conduct their research adopting ICT tools for the production and processing of research products. In the last decade, research infrastructures (organizational and technological facilities supporting research activities) are investing in “e-infrastructures” that leverage ICT tools, services, guidelines and policies to support the digital practices of their community of researchers.¹ To find an analogy with traditional science, where research is often done in a laboratory, e-infrastructures are the place where researchers can define their *digital laboratories*, i.e. the subset of assets and tools that they use to conduct their research. Researchers run their digital experiments (e.g. simulations, data analysis) taking advantage of the digital

¹ European Commission, <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/research-infrastructures-including-e-infrastructures>.

laboratory assets (e.g. RStudio², Jupyter Notebook³, Taverna workbench⁴) and generate new research data and computational products (e.g. software, R algorithms, computational workflows) that can be shared with other researchers of the same community, that can be discovered, accessed and reused and ultimately can become part of the e-infrastructure.

The role of digital laboratories is therefore twofold: on the one hand they support researchers in their advancement of science, offering the facilities needed for their daily activities; on the other hand, they foster the dissemination of research within the research community, supporting discovery, access to, sharing, and reuse of digital research products. In fact, their digital nature offers unprecedented opportunities for scientists, who can share not only scientific literature describing their findings, but also the digital results that they produced, together with the digital laboratory itself. Those features are fundamental for an effective implementation of the Open Science (OS) paradigm [1, 2]. OS is a set of practices of science, advocated by scientific/scholarly communication stakeholders (i.e., research funders⁵, research and academic organisations, and researchers), according to which a research activity (intended as an activity performed to answer a research question) and all the generated products should be freely available, under terms that enable their findability, accessibility, re-use, and re-distribution [3].

If supported with adequate degrees of openness, scientists would find the conditions to repeat (“same research activity, same laboratory”), replicate (“same research activity, different laboratory”), reproduce (“same research activity, different input parameters”), or re-use (“using a product into another research activity”) the results of research activities, thereby maximizing transparency and exploitation of scientific findings [4].

The ability to share research products, in combination with digital laboratories, opens the way to Open Science principles. According to these principles, science should be open not only once it is concluded, but also while it is being performed. In other words, scientists should, as much as possible, make their methodologies, thinking and findings available to enable/maximize collaboration and reuse by the community. The digital laboratory becomes therefore the core of this vision as it is the place providing the assets needed by the researchers to implement their research flow (i.e. the actual sequence of experiments required to prove the initial thesis) and at the same time the place providing the generated research products, for sharing and peer-reviewing. For example, scientists performing analysis of data using R scripts, may use a digital laboratory equipped with the software RStudio offered as-a-service by an online provider (e.g. BlueBridge e-infrastructure powered by D4Science⁶) and a repository where they can store/share their R scripts and their input and output datasets (e.g. Zenodo.org).

In the following we shall refer to the following concepts:

² RStudio, <https://www.rstudio.com/>.

³ Jupyter Notebook, <http://jupyter.org/>.

⁴ Taverna workbench, <https://taverna.incubator.apache.org/>.

⁵ Examples are research funders like the European Commission [2], Wellcome Trust and funders of the cOAlition-S (<https://www.scienceurope.org/coalition-s/>).

⁶ BlueBridge, <http://www.bluebridge-vres.eu/>.

- A *research activity* is performed to answer a “research question”, usually formulated as one or more hypotheses to be proved true;
- A *research flow* is made of a number of experiments, realized as sequence of steps in the context of a digital laboratory, executed by scientists driven by the ultimate intent of proving an initial scientific thesis;
- An *experiment* is a goal-driven sequence of steps set to verify a thesis, and whose result may inspire further experiments to address the target of the overarching research activity. One experiment can be constituted of several series of sequential steps executed in parallel.
- A *digital laboratory* is a pool of digital assets (e.g. on-line tools, desktop tools, methodologies, standards) used by scientists to perform the steps of an experiment and generate research products.
- A *research product* is any digital object generated during the research flow that was relevant to complete the research activity and (possibly) relevant for its interpretation once the research activity has been completed. Products are digital objects, whose human consumption depends on computer programs; they are concrete items that can be discovered, accessed, and possibly re-used under given access rights. Examples are datasets in a data repository (e.g. sea observations in the PANGAEA repository⁷), but also entries in domain databases (e.g. proteins in UNIPROT⁸), software (e.g. models implemented as R algorithms in GitHub), and of course the scientific article, reporting about the findings of a research activity.

A research activity may therefore generate a number of research products that enable scientists to draw their conclusions. Indeed, several “intermediate” products are generated at different stages, e.g. input and outputs of unsuccessful experiments, versions of the final products to be refined. A research activity can therefore be described by a research flow, i.e. a sequence of steps $S_1 \dots S_n$, possibly grouped into experiments, carried out in the frame of a digital laboratory (see Fig. 1 left). More specifically, each step S_i of a research flow is in turn a sequence of actions enacted by humans, possibly by means of digital laboratory assets, that may require or produce (intermediate) research products. Clearly, some (or all) of the research products generated during the research flow may become, at some point in time, new assets of a digital laboratory. An example related to the field of geothermal energy science is shown on Fig. 1 (right). A researcher gets data from a GIS database and provide those data as input to the 3D GeoModeller application: those are the assets of the digital laboratory. The experiment is composed of two steps: the researcher selects one of the equations available from the GeoModeller and provide the needed input parameter for the generation of the model. The researcher then interprets the model and produces the scientific article to be published. In this (simplified) example both the generated model and the article are research products that can be shared and peer-reviewed. The configuration of the application used for generating a model could also be made available, to increase transparency and replicability. In a more complete scenario, the research flow could

⁷ PANGAEA: <https://www.pangaea.de/>.

⁸ UNIPROT: <https://www.ebi.ac.uk/uniprot>.

include several repetitions of the model generation, with different input parameters or with different equations, until the interpretation of the generated model would satisfy the researcher.

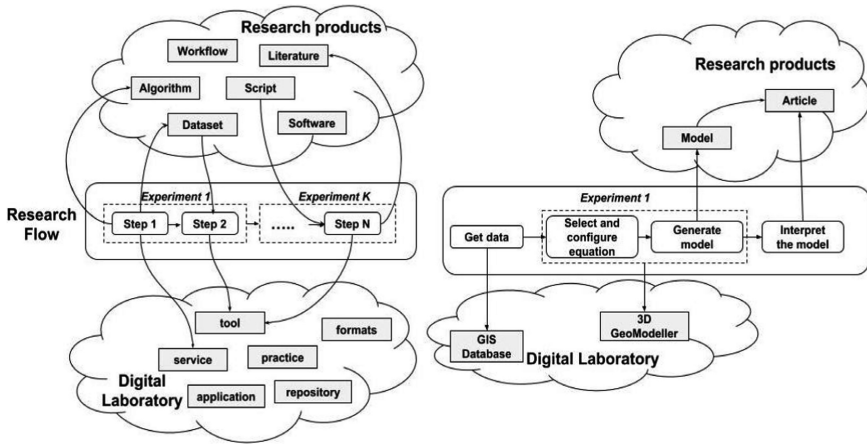


Fig. 1. The research flow in the digital era and an example on geothermal science

This article presents the foundations of a framework for the representation of research flows in support of peer review for a given discipline of science. The aim of the framework is to enable research communities to formally define research flow patterns that define which are the steps that should be peer-reviewed. Such a framework may become the scaffolding enabling the development of tools for supporting ongoing peer review of research flows. Such tools could be “hooked”, in real time, to the underlying digital laboratory, where scientists are carrying out their research flow, and they would abstract over the complexity of the research activity and offer user-friendly dashboards to examine the adopted scientific process, explore the ongoing research flow, and evaluate its intermediate experiments and products.

Outline. The state of the art on current practices for the peer review of research flows is presented in Sect. 2. Section 3 describes the framework in support of peer-review of research flows. Conclusion and future work are addressed in Sect. 4.

2 Current Practices for the Peer-Review of the Research Flow

Researchers usually tend to make a clear distinction between the phase of research activities and the phase of research publishing. Research publishing is generally intended as the moment in which researchers share their findings with the broader community of all researchers, hence also the moment at which the peer review of the research flow starts, assuming that the published material somehow “represents” the whole research flow.

Traditionally, the peer review of the research flow has been delegated to scientific literature (e.g., articles, books, technical reports, PhD theses) which is still regarded as the common omni-comprehensive unit of scientific dissemination. The published material provides the means for the review of the research flow (and possibly reproducibility) by explaining and describing the different steps, the digital laboratory where they were conducted (i.e., methodology, tools, standards, etc.), and describing any product used or yielded by the research activity, thus facilitating reproducibility by a detailed, theoretically unambiguous, description of the experiments. However, a natural language description of a methodology can have different interpretations and typically does not include all the details that are needed in order to replicate the experiment or reproduce the results. In addition, it has been found [5–7] that “methodology” sections of many papers often include generic sentences, and lack the details that would be necessary to attempt the reproduction of the results. To overcome this issue, the Centre for Open Science, in collaboration with more than 3,000 journals, is testing the approach of “pre-registered reports”, i.e. documents that describe the research flow in a structured and detailed form and that are submitted to the journal before the research starts [8].

To overcome the drawbacks of publishing only scientific literature, a common approach adopted today across several disciplines is that of publishing articles together with links to other digital products of the research, deposited in dedicated repositories. In the majority of cases the papers provide links to datasets, although some cutting-edge research communities are experimenting with links to computational products (e.g. software, scientific workflows), experiments and methodologies.

A growing number of data repositories and archives assign unique, persistent identifiers to the deposited datasets and apply the FAIR principles [9, 10] (data should be Findable, Accessible, Interoperable and Re-usable). Relevant examples are Zenodo and figshare (cross-discipline and allowing deposition of products of any type), DRYAD (mostly for life science), PANGAEA (earth & environmental science), Archeology Data Service (archaeology), DANS (multi-discipline, mostly humanities and social sciences). Data repositories typically implement data review processes that focus on checking the technical details of the datasets, validating their metadata and, possibly, their descriptions, without addressing the scientific value of the dataset itself. In fact, this type of data review usually is not performed by “peers”, but by editors and curators of the repository, who are not necessarily researchers in the same field of the depositors, but are instead expert of data management, archiving, and data preservation.

A different approach to data peer review is adopted by data journals [11], which publish data papers, i.e. papers describing datasets in terms of content, provenance, and foreseen usage. Data journals inherited the peer-review process from traditional journals of scientific literature and apply it, with slight changes, to the data papers. The peer review of data papers is mostly focused on the review of metadata, whose completeness and clarity are considered fundamental to facilitate data re-use [12–14]. With the existing approaches, the reproducibility of a dataset (when applicable, as some datasets cannot be reproduced, such as those generated by devices for atmospheric measurements) is not considered an important aspect of data (peer) review, although reproducibility is crucial to demonstrate the correctness of data and its analysis, upon which researchers’ conclusions are based.

In addition to the publishing of research data, researchers started to publish also their computational products, such as software, R algorithms, computational workflows. The publishing is typically performed by means of tools and services that are not meant for scholarly communication but that implement general patterns for collaboration and sharing of computational products. Examples are software repositories (or Version Control Systems (VCSs)) with their hosting services like Github, and language-specific repositories like CRAN (The Comprehensive R Archive Network), the Python Package Index, and CPAN. Github is currently the most popular online software repository and, thanks to the collaboration with Zenodo, DOIs can be assigned to software releases.

In order to proceed on the road of Open Science, research in information science has started to explore and conceive solutions that focus on generating research products whose purpose is sharing “experiments” rather than providing “final results”. Such products are digital encodings, executable by machines to reproduce the steps of an experiment or an entire research flow. They extract from the scientific article the concepts of experiment and research flow, making them tangible, machine-processable and shareable products of science. Research in the area of scholarly communication has focused mainly on data (information) models for the representation of digital products encoding experiments, and on tools for generating (and later executing) experiment products that should include all the details of the digital laboratory used to run the experiment and needed for repeatability and reproducibility. Relevant examples of information systems for experiment publishing and reproducibility are protocols.io [15], ArrayExpress [16] and myExperiment [17].

In conclusion, literature is the most common way to make a research flow sharable, since other scientists can discover and read about somebody else’s methods, protocols, and findings, but in general it does not allow a complete assessment of the research flow. Some solutions have reproducibility of science as their main objective, rather than peer review, hence they focus on the executable representation of digital objects that encode final successful experiments. Research is still peer-reviewed out of its original context (the digital laboratory) and the concepts of machine-assisted peer-review and ongoing peer-review are not being considered. It would be desirable to have tools for machine-assisted peer-review, built on the very same digital laboratory assets that were used to generate the research products. Although humans would still play a central role in the peer-review process with regards to the evaluation of novelty and impact of a research flow and its final products, such tools would support reviewers facing challenges going beyond their capabilities, like checking the quality of each record in a database, or the conformance to structural and semantic requirements [18]. The ultimate goal should be that of an ongoing peer review of the research flow. In contrast with traditional peer review models, which assess scientific results only once the research activity has been successfully completed, ongoing peer review could also be applied as a sort of monitoring and interim evaluation process. The sharing of intermediate research flow experiments and steps would open up the possibility of publishing negative results. This practice could have a twofold positive effect: on the one hand, the researcher might receive comments and advice from colleagues, on the other hand, she would help the community by suggesting to avoid the same “mistakes” [19].

3 A Framework in Support of Peer-Review of the Research Flow

3.1 Overview

The implementation of a fully-fledged methodology for the peer review of the research flow has requirements (tools and practices) that differ from those identified in Open Science for reproducibility. Reproducibility of science and its underlying principles are indeed crucial to support transparent peer review, but existing practices are not enough to fully address research flow peer review. In order to support this kind of peer review, reviewers should evaluate science by means of a user-friendly environment that transparently relies on the underlying digital laboratory assets, hides their ICT complexity, and gives those guarantees of repeatability and reproducibility recognized by the community.

Depending on the tools and technology available in a digital laboratory, scientists may generate products whose goal is not just sharing “findings” but also sharing “methodologies”. Methodology products are digital objects encoding experiments or the research flow itself. As such, they are generated to model the actions performed by the scientists and enable their machine-assisted repetition. The availability of research products at various stages of the research flow (see Fig. 2) makes it possible to introduce peer review stages while the research activity is ongoing. Specifically, depending on the kind of products made available, different degrees of peer review may be reached, to support manual but also machine-supported reproducibility and consequently enforce more transparent and objective research flow peer review practices:

- Manual reproducibility: the digital laboratory generates, or supports researchers at generating:
 - Literature, defined as narrative descriptions of research activities (e.g. scientific articles, books, documentation);
 - Datasets, defined as “digital objects used as evidence of phenomena for the purpose of research or scholarship” [20];
 - Computational products (e.g. software, tools), intended as digital objects encoding business logic/algorithms to perform computational actions over data; Reviewers are provided with the products generated by a research flow, whose steps are reported in an article together with references to the digital laboratory. Reproducibility and research flow assessment strongly depends on humans, both in the way the research flow is described and in the ability of the reviewers, and in general of other researchers, to repeat the same actions.
- Machine reproducibility of experiments: the digital laboratory generates literature, datasets and computational products together with:
 - Experiments, intended as executable digital objects encoding a sequence of actions (e.g. a methodology) that make use of digital laboratory assets to deliver research products. Reviewers are provided with an experiment, inclusive of products and digital assets. Reproducibility can be objectively supported by a machine and finally evaluated, but the assessment of methodology as a whole still depends on humans.

- Machine reproducibility of research flows: the digital laboratory generates literature, datasets, computational products, experiments together with
 - Research flows, intended as digital objects encoding a flow, inclusive of experiments, intermediate, and final products, and their relationships; the research flow may be encoded as a sharable and possibly reproducible digital product. Reviewers are provided with technology to reproduce experiments and research flows. In this scenario, human judgment is supported by machines, which can provide a higher degree of transparency.

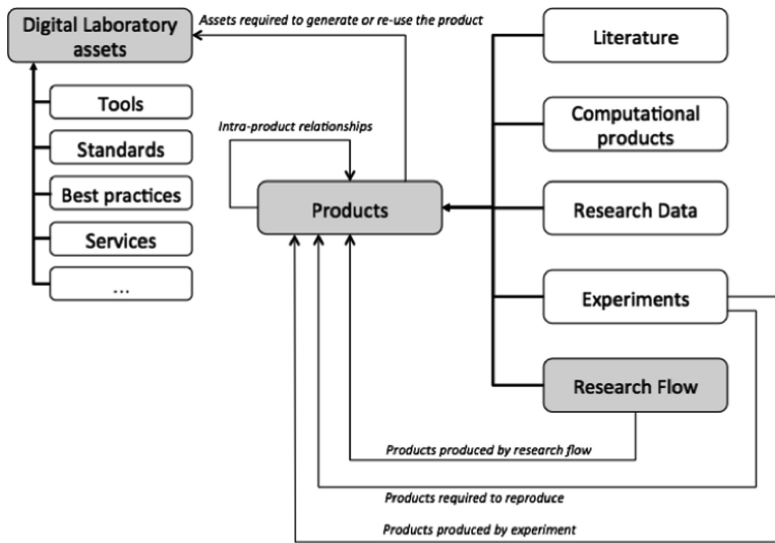


Fig. 2. Entities of the research flow

We propose a framework in support of the peer-review of the research flow. The framework is built around the notion of *research flow templates*. These are representations of the scientific processes in terms of patterns (sequences and cycles) of experiments and relative steps to be peer reviewed. Note that such templates should include only the experiments and steps that are relevant for peer-reviewing the research flow, including their inputs and outputs. In other words, research flow templates are not intended to describe the detailed experiments and steps of a research activity (as would be the case for reproducibility), but are intended to model the subset of actions that are relevant to assess the quality of the research flow.

3.2 Concepts of the Framework

In order for a research community to provide specifications for the peer-reviewable parts of its research flow and be able to build adequate tools in support of reviewers, a simple formal framework capable of describing the structure of the templates for that

community should be available. Each template should reflect one particular way of performing science, capturing the steps which should be subject to peer review (a community may have more than one template). At the same time, templates help researchers to comply with certain rules and expectations when producing science. Templates express common behaviour, determine good practices, facilitate reproducibility and transparent evaluation of science. To make an analogy, the structure of a fully-fledged template should reflect the structure of a recipe for cooking. It should specify a list of all the types of products needed from the digital laboratory at each step of the research flow (the ingredients) and should provide a detailed description, possibly machine actionable, of all the steps to be executed (the mixing and the cooking) in order to obtain the research results (the cake).

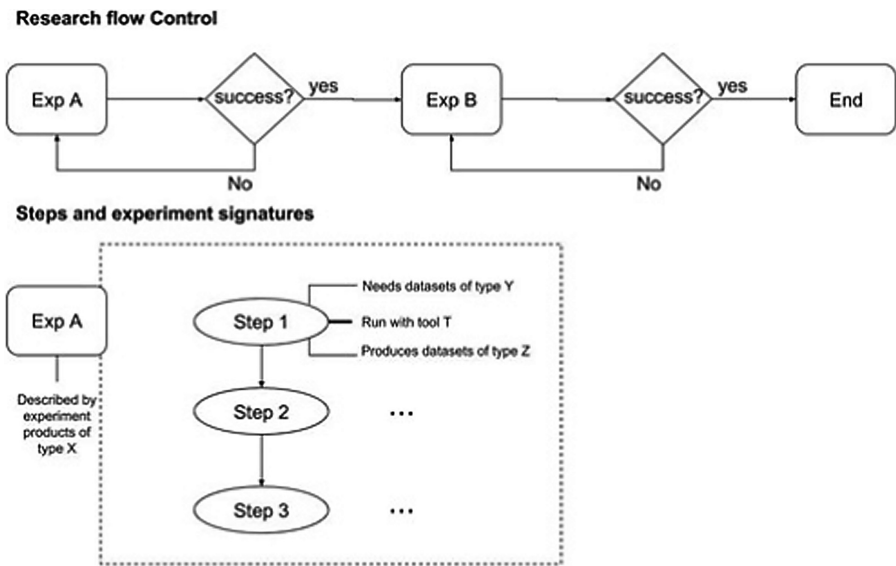


Fig. 3. Research flow templates concepts

A research flow review framework should encompass the following concepts (see Fig. 3):

- Research flow template: the model of research flow to be followed by scientists in terms of experiments, including cycles, conditions, etc. to be peer reviewed;
- Signatures of steps and experiment, intended as:
 - The types of input and output products (e.g. datasets, computational products, documentation, scientific articles);
 - Asset of the digital laboratory (which is not necessary a research product) required for the execution of the step or of the experiment (e.g. tool, service, application);
 - For experiments: the format in which the experiment will be digitally represented and shared with peers.

Referring again to the example in Fig. 1 (geothermal science research), we show in Fig. 4 a more detailed view of the elements of a possible template. The template would specify the input and output of the first step (obtaining data from a Geo database); it would then specify the input (obtained data), the output (the generated model) and the tool (the 3D GeoModeller) for the second step, which could be executed several times, each time generating a “research object” containing all these data; finally, it would specify the input (the final generated model) and the output (the paper to be published) of the last step, which obtains the research results.

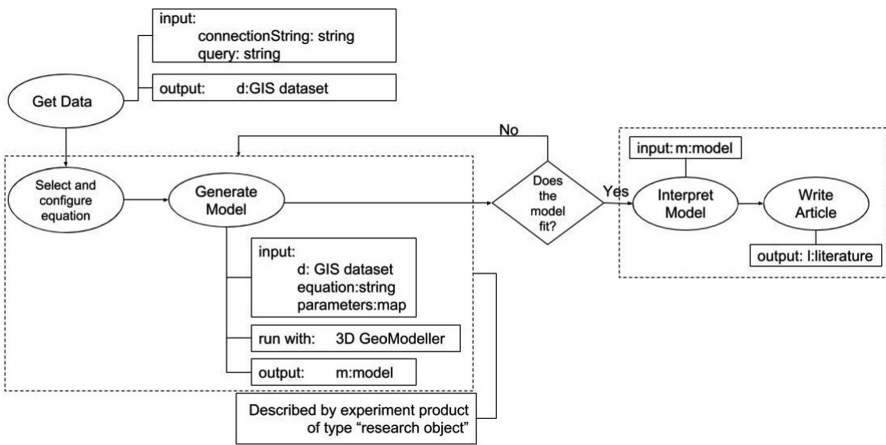


Fig. 4. A possible template for geothermal research

3.3 Building Peer-Review Tools on Top of the Framework

The framework and its templates may become the scaffolding on top of which developers can build tools to support an ongoing peer review of research flows by “real-time hooking” to the underlying digital laboratory, where scientists are carrying out their research activities. Such tools would abstract over the complexity of the research activity and offer user-friendly dashboards to examine the adopted scientific process, explore the ongoing research flow, and evaluate its intermediate experiments and relative products. In a less advanced implementation, such tools might provide scientific process and research flow to reviewers once the research activity has been terminated, inclusive of all intermediate experiments, steps and research products.

For example, consider the scientific process in Fig. 5 [21], which models one experiment repeatedly executed until the research activity is successful. At every cycle, the researcher designs (1) and collects (3) input data, instruments the digital laboratory with processing algorithms (2) and performs some computations (4) to produce output data. Finally, it publishes (5) all such products. In this model, we might assume that the only point for review is the one of “publication” (5), where input data and all the digital laboratory assets are made available. The corresponding research flow review template

would model the same cycle and be made of one experiment including a single step of peer review, the one of publication mentioned above. Tools for an ongoing peer review would allow reviewers to select a given execution of the experiment in time, explore and assess input and output data, and re-execute the given step, given the related products. Of course, such tools should be equipped with functionalities to provide feedback and evaluation.



Fig. 5. Research lifecycle adapted from the research process model by Kraker & Lindstaedt

Sharing a framework of this kind allows the realization of research publishing tools and review tools that allow scientists to produce products as expected by the community and allow other scientists to access such products, for reuse, reproducibility, and review. As mentioned above, to be effective and used in practice, such tools should be:

- Integrated with digital laboratory assets used to perform science: scientists should focus on developing their science rather than publishing it; the process of creating research products and methodology products should be delegated as much as possible to machines, together with tracking the history of the actual research flow; digital laboratory assets require research publishing tools (e.g. wrappers, mediators) capable of flanking the experiment functionality they support with functionality for packaging and publishing the relative products, so that review tools can benefit from those;
- Easy to use: user-friendly enough for scientists to access machine-assisted review tools without development skills; reviewers should be able to view the actual research flow, to view its current stage of development, and to apply machine-assisted validation from end-user interfaces;
- Trustworthy: easy to use is a property that should come with guarantees of fairness, typically endorsed by the community adopting research publishing and review tools.

Implementing this vision raises serious challenges, among which a major one is the realization and maintenance of tools for publishing and review, whose cost do not easily find a donor in communities that are typically formed by scientists rather than institutions.

In addition, endorsement of communities and cultural convergence are required towards scholarly communication practices that enable to share, cite, evaluate and assign scientific reward (i.e. author credit) for all the types of research products. In particular, the research communities should understand that, together with the advantages of peer reviewing the research flow and all the research products, there would be an increased burden for the researchers and the reviewers, mitigated as much as possible by the facilities provided by the infrastructure.

4 Conclusion and Future Work

The Open Science paradigm calls for the availability, findability and accessibility of all products generated by a research activity. That practice is a prerequisite for reaching two of the main goals of the Open Science movement: reproducibility and transparent assessment of research activities. In this paper we have described the current practices for the peer review of research flows, which range from traditional peer-review via scientific literature to peer review by reproducibility of digital experiments. We have argued that current practices have reproducibility of science as their main objective and they do not fully address transparent assessment and its features like publishing negative results, supporting peer-review while the research activities are ongoing and enabling machine-assisted peer review.

Foundations of a framework for the peer-review of research flows have been presented. The goal of the framework is to be the bridge between the place where the research is conducted (i.e., the digital laboratory) and the place where the research is published (or in general, made available and accessible). The framework aims at providing the scaffolding on top of which reviewers can evaluate science by means of a user-friendly environment that transparently relies on the underlying digital laboratory assets, hides their ICT complexity, and gives guarantees of repeatability and reproducibility recognized by the community. One of the building blocks of the framework is the notion of research flow template, through which a community can model the research flow to be followed by scientists in terms of experiments, including cycles, conditions, etc. to be peer reviewed. The framework allows communities to define one or more research flow templates, each capturing the steps which should be subject to peer review for a specific type of research activity. Templates are not only useful to peers willing to evaluate a research activity, but also enforce researchers at complying with certain expectations of their community, like best practices and common behaviour.

The framework is theoretically applicable to any field of research adopting digital objects and/or producing digital research outputs. Detailed analysis on the applicability of the framework is ongoing. Specifically, the fields of geothermal energy science and archeology have been considered as representatives of non-fully digital disciplines, which may pose challenges from the modelling point of view, as not all the research assets and products may be available in a digital laboratory.

Acknowledgement. This work is partially funded by the EC project OpenUP (H2020-GARRI-2015-1, Grant Agreement: 710722). The content of this work reflects the views of the author(s). The European Commission is not responsible for any use that may be made of the information it contains.

References

1. European Commission: Validation of the results of the public consultation on Science 2.0: Science in Transition [report]. Brussels: European Commission, Directorate-General for Research and Innovation (2015). http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf
2. European Commission's Directorate-General for Research & Innovation (RTD): Open Innovation, Open Science and Open to the World (2016). <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>
3. FOSTER: Open Science Definition. <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>
4. Bechhofer, S., et al.: Why linked data is not enough for scientists. *Future Gener. Comput. Syst.* **29**(2), 99–611 (2013). <https://doi.org/10.1016/j.future.2011.08.004>. ISSN 0167-739X
5. Smagorinsky, P.: The method section as conceptual epicenter in constructing social science research reports. *Writ. Commun.* **25**, 389–411 (2008). <http://journals.sagepub.com/doi/pdf/10.1177/0741088308317815>
6. Teytelman, L.: We've been itching to share this! Integration of GigaScience and protocols.io is an example of how science publishing should work. *Protocols.io news* (2016). <https://www.protocols.io/groups/protocolsio/news/weve-been-itching-to-share-this-integration-of-gigascience>
7. Cotos, E., Huffman, S., Link, S.: A move/step model for methods sections: demonstrating rigour and credibility. *Engl. Specif. Purp.* **46**, 90–106 (2017). <https://doi.org/10.1016/j.esp.2017.01.001>. ISSN 0889-4906
8. Center for Open Science: Registered Reports: peer review before results are known to align scientific values and practices. <https://cos.io/rr/>
9. FORCE11: Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing Version B1.0 (2014). <https://www.force11.org/fairprinciples>
10. Wilkinson, M.D., et al.: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
11. Candela, L., Castelli, D., Manghi, P., Tani, A.: Data journals: a survey. *J. Assoc. Inf. Sci. Technol.* **66**, 1747–1762 (2015). <https://doi.org/10.1002/asi.23358>
12. Assante, M., Candela, L., Castelli, D., Tani, A.: Are scientific data repositories coping with research data publishing? *Data Sci. J.* **15**, 6 (2016). <https://doi.org/10.5334/dsj-2016-006>
13. Mayernik, M.S., Callaghan, S., Leigh, R., Tedds, J., Worley, S.: Peer review of datasets: when, why, and how. *Bull. Amer. Meteor. Soc.* **96**, 191–201 (2015). <https://doi.org/10.1175/BAMS-D-13-00083.1>
14. Carpenter, T.A.: What Constitutes Peer Review of Data: a survey of published peer review guidelines (2017). arXiv preprint [arXiv:1704.02236](https://arxiv.org/abs/1704.02236). <https://arxiv.org/pdf/1704.02236.pdf>
15. Protocols.io team: How to make your protocol more reproducible, discoverable, and user-friendly (2017). <http://dx.doi.org/10.17504/protocols.io.g7vbn6>
16. Tang, A.: ArrayExpress at EMBL-EBI - quality first! Repositve blog (2017). <https://blog.repositve.io/arrayexpress-at-embli-ebi-quality-first/>

17. De Roure, D., Goble, C., Stevens, R.: The design and realisation of the myExperiment Virtual Research Environment for social sharing of workflows. *Future Gener. Comput. Syst.* **25**(5), 561–567 (2009). <https://doi.org/10.1016/j.future.2008.06.010>
18. Shanahan, D.: A peerless review? Automating methodological and statistical review (2016). <https://blogs.biomedcentral.com/bmcblog/2016/05/23/peerless-review-automating-methodological-statistical-review/>
19. Di Leo, A., Risi, E., Biganzoli, L.: No pain, no gain... What we can learn from a trial reporting negative results. *Ann. Oncol.* **28**(4), 678–680 (2017). <https://doi.org/10.1093/annonc/mdx065>
20. Borgman, C.L.: *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, Cambridge (2015)
21. Kraker, P., Bachleitner, R., et al.: Deliverable D4.1 – Practices evaluation and mapping: methods, tools and user needs (2017). http://openup-h2020.eu/wp-content/uploads/2017/01/OpenUP_D4.1_Practices-evaluation-and-mapping.-Methods-tools-and-user-needs.pdf