

1. Introduction

1.1 Motivation

Research compendia are an increasingly used form of publication and scholarly communication. They comprise not only the research paper’s text and figures, but also all data and software used to conduct the computational workflow and create all outputs. They provide a lot of added value by revealing more of the research process to readers, but, if not done well, they can increase the difficulty of understanding the research. To help readers better understand how to read a research compendium, we extend Keshav’s three-pass method targeted at improving skills for reading a research paper (???) with additional steps relevant to a research compendium’s content.

Unlike the first version of the original (???), we cannot draw from a long history of experience, because until recently research compendia have been relatively rare. Our intention here is to provide guidance at the outset of an emerging practice to both readers and authors of research compendia to help them understand each others’ perspectives and needs and improve their communication. Authors can use this guide to improve their research compendium’s structure and content by better anticipating their readers’ needs. They should not be held back by unwarranted concerns, like providing support (???). Readers can avoid the trap of falling too deep into technological challenges by an iterative approach to reading and using that gives attention to the scientific issues. Ultimately research compendia can enhance and deepen the reading experience, if done right. Keshav’s following introduction applies directly to research compendia:

Researchers must read papers for several reasons: to review them for a conference or a class, to keep current in their field, or for a literature survey of a new field. A typical researcher will likely spend hundreds of hours every year reading papers.

Learning to efficiently read a paper is a critical but rarely taught skill. Beginning graduate students, therefore, must learn on their own using trial and error. Students waste much effort in the process and are frequently driven to frustration.

For many years I have used a simple ‘three-pass’ approach to prevent me from drowning in the details of a paper before getting a bird’s-eye-view. It allows me to estimate

the amount of time required to review a set of papers. Moreover, I can adjust the depth of paper evaluation depending on my needs and how much time I have. This paper describes the approach and its use in doing a literature survey. (???)

The additions made in this work to accommodate for the content in a research compendium are quite extensive. This stems from the complexity that an interactive compendium has compared to a classic static “paper”, because a research compendium goes well beyond the “mere advertising of the scholarship” (???). We see the breadth of additions as a sign of potential, namely for unprecedented transparency, openness, and collaboration.

1.2 Structure

In the remainder of this paper, the excellent original work is taken over completely. It is set in *italic font* based on the most recent online version: (???). The term “paper” was not replaced with “research compendium” for better readability.

First we briefly introduce research compendia and existing conventions. We further list relevant resources for authors related to research compendia. Then, matching the original paper’s section numbering, Section 2 extends the “Three-pass Approach” to include research compendium features in the reading process. Section 3 extends “Doing a Literature Survey” with aspects relevant reviewing many research compendia.

1.3 Research compendia

The term *research compendium* was coined by (???) who “introduce[d] the concept of a compendium as both a container for the different elements that make up the document and its computations (i.e. text, code, data, . . .), and as a means for distributing, managing and updating the collection.” According to (???) it provides “a standard and easily recognisable way for organising the digital materials of a research project to enable other researchers to inspect, reproduce, and extend the research”. This standard may differ between scientific domains, yet the intentions and benefits are the same. Research compendia are practised Open Science culture and as such improve transparency (???), “make more published research true” (???), and enable enhanced review and publication workflows (???). They answer readers’ needs to understand complex analyses through inspection and manipulation (???) and enable other researchers to reproduce and extend the research (???). Research compendia improve citations since code and data are

As this article is focused on providing hands-on guidance on using, and to some extent also creating, research compendia, we refer the reader to the references for more specific details. For the remainder of this work, we assume a minimal view of a research compendium suitable for *readers* who examine a research compendium directly. A research compendium has three integral parts: text, code, and data. Text can be instructions, software documentation, or a full manuscript with figures. Code can be scripts, software packages, specifications of dependencies and computational environments, or even virtual machines. Data can be just about anything, but probably comprises plain text or binary files that are used as input to the workflow, and produced as output from executing the workflow.

The following conventions are specifically for research compendia:

- pendia, more specifically R and Python packages (???) describe the tool **ReproZip** (<https://reprozip.org>) to support capture and reproduction of a research compendium

The key idea is that you should read the paper in up to three passes, instead of starting at the beginning and plowing your way to the end. Each pass accomplishes specific goals and builds upon the previous pass: The first pass gives you a general idea about the paper. The second pass lets you grasp the paper's content, but not its details. The third pass helps you understand the paper in depth. (???)

The first pass is a quick scan to get a bird's-eye view of the paper. You can also decide whether you need to do any more passes. This pass should take about five to ten minutes and consists of the following steps: (???)

- At the end of the first pass, you should be able to answer the seven Cs:

- 2

4. *Contributions: What are the paper's main contributions?*
5. *Clarity: Is the paper well written?*
6. *Construction: What are the building blocks of the analysis workflow and how accessible are they (data set(s), programming language(s), tools, algorithms, scripts)? Under what licenses are code and data published?*
7. *Complexity: What is the scale of the analysis (e.g. HPC, required OS/cores/memory, typical execution time, data size) and the software (number of dependencies and is installation possible with dependency management tools)?*

Using this information, you may choose not to read further (and not print it out, thus saving trees). This could be because the paper doesn't interest you, or you don't know enough about the area to understand the paper, or that the authors make invalid assumptions. (???)

You may also choose not to pursue the parts of the research compendium further, i.e. not running the workflow or looking at data or code, thus saving resources. Reasons to not read further that relate specifically to code and data may be that you don't have the expertise or access to resources to re-use the data and code.

The first pass is adequate for papers that aren't in your research area, but may someday prove relevant. (???)

This first pass suits research compendia comprising potentially re-usable components, like workflows or algorithms using data sets or generic software that are directly transferable to your field of research. After the first pass, you should be able to judge if the software is useful, if it works.

Incidentally, when you write a paper, you can expect most reviewers (and readers) to make only one pass over it. Take care to choose coherent section and sub-section titles and to write concise and comprehensive abstracts. If a reviewer cannot understand the gist after one pass, the paper will likely be rejected; if a reader cannot understand the highlights of the paper after five minutes, the paper will likely never be read. For these reasons, a 'graphical abstract' that summarizes a paper with a single well-chosen figure is an excellent idea and can be increasingly found in scientific journals. (???)

When you write a paper, take care to add instruc-

tions on how a reader can reproduce your work and provide all required parts, i.e. publish a research compendium. The instructions should start with a "blank" system and be specific, i.e. ready for copy & paste, including expected or experienced execution times and resources. Such instructions give readers a good idea about what is needed to recreate your environment and execute the analysis. If your work requires specialised or bespoke hardware (HPC, specific GPUs), consider creating an exemplary, reduced analysis that runs in regular environments.

Also ensure your code and data are properly deposited, citable and licensed. If you don't do this, these core parts of your work will likely never be properly evaluated or re-used. See the section "Research Compendia", above, for recommendations and further reading on how to make your reviewers' and readers' lives easier.

2.2 The second pass

In the second pass, read the paper with greater care, but ignore details such as proofs. It helps to jot down the key points, or to make comments in the margins, as you read. Dominik Grusemann from Uni Augsburg suggests that you "note down terms you didn't understand, or questions you may want to ask the author." If you are acting as a paper referee, these comments will help you when you are writing your review, and to back up your review during the program committee meeting. (???)

1. *Look carefully at the figures, diagrams and other illustrations in the paper. Pay special attention to graphs. Are the axes properly labelled? Are results shown with error bars, so that conclusions are statistically significant? Common mistakes like these will separate rushed, shoddy work from the truly excellent. (???)*
2. *Remember to mark relevant unread references for further reading (this is a good way to learn more about the background of the paper). (???)*
3. *Skim over data and source code files without opening them. Are they reasonably named (???)? Do they follow a well-defined structure (e.g. a Python package or a research compendium convention)? Is there a README file and/or structured documentation for functionalities?*
4. *Visit the online source code repository, if available. Is it established and well maintained, or orphaned? Is there only one author or are there*

contributors? How responsive are they to issues? Does the repository have signs of public recognition (i.e. GitHub “stars” and “forks”)? Are there regular releases, using semantic versioning?

5. Follow the instructions to install the required software and execute the research compendium’s workflow with the provided parameters and input or sample data. Note down errors or warnings but do not try to fix any but trivial or known problems (e.g. fixing a path or installing an undocumented dependency).
6. Compare the outputs with the expected ones reported in the paper. Also check for differences in output figures: Do labels, legends etc. match those in the paper?

Points 3 and 4 above hint at how to estimate the quality of a software, but we recommend to be realistic as to what to expect and be careful not to judge too fast. The software project you evaluate might be done by a single researcher who is not a professional programmer, working under a lot of pressure to write code for a single use case. In these situations one might find low levels of code documentation, but further documentation might be quickly provided by the authors once you as an external user show interest. Also, no recent changes or releases at a source code repository can also mean the software is stable and simply works with no problems!

The second pass should take up to an hour for an experienced reader. (???)

This does not include the computation time of workflows in a research compendium. Use this time to complete first passes for one or several other compendia. If the software used is familiar, you may attempt to reduce the computation time by sub-setting data or simplifying the workflow. As an author, consider adding a reduced example to your research compendium for easier access by readers.

After this pass, you should be able to grasp the content of the paper. (???)

You should have re-executed the provided workflow or understand why you could not. You should be able to complete the second pass even if you are unfamiliar with the actual language the software is written in or if you are not a developer yourself. However we do recommend not to dive too deep, i.e. not going beyond the provided instructions for the research compendium’s workflow. At this stage, it is the author’s responsibility to guide you through their work.

Still, you may also face unsolvable problems, like access to specific infrastructure. But if you encounter

issues or have questions, you should communicate these to the author, for example in the software’s public code repository, if available. It is important to do this respectfully, and give the authors a chance to fix bugs or respond to issues (???). Also let the authors know if your reproduction was successful, especially if you used a different operating system or software version than reported.

At this point you should be able to judge whether the software works and if it is sustainable. Based on this evaluation you can decide to re-use parts of the analysis, i.e. software, data, or method, for your own work.

You should be able to summarize the main thrust of the paper, with supporting evidence, to someone else. This level of detail is appropriate for a paper in which you are interested, but does not lie in your research speciality. Sometimes you won’t understand a paper even at the end of the second pass. This may be because the subject matter is new to you, with unfamiliar terminology and acronyms. Or the authors may use a proof or experimental technique that you don’t understand, so that the bulk of the paper is incomprehensible. The paper may be poorly written with unsubstantiated assertions and numerous forward references. (???)

The research compendium may have incomplete documentation, rely on unavailable software (e.g. proprietary) or data (e.g. sensitive), or require infrastructure not available to you (e.g. high-performance computing, HPC). It may use a programming language or programming paradigms unfamiliar to you.

Or it could just be that it’s late at night and you’re tired. You can now choose to: (a) set the paper aside, hoping you don’t need to understand the material to be successful in your career, (b) return to the paper later, perhaps after reading background material or (c) persevere and go on to the third pass. (???)

2.3 The third pass

To fully understand a paper, particularly if you are a reviewer, requires a third pass. The key to the third pass is to attempt to virtually re-implement the paper: that is, making the same assumptions as the authors, re-create the work. By comparing this re-creation with the actual paper, you can eas-

ily identify not only a paper's innovations, but also its hidden failings and assumptions. This pass requires great attention to detail. (???)

If a best practice or established convention for structuring data and code was followed, familiarise yourself with it now.

You should identify and challenge every assumption in every statement. Moreover, you should think about how you yourself would present a particular idea. This comparison of the actual with the virtual lends a sharp insight into the proof and presentation techniques in the paper and you can very likely add this to your repertoire of tools. (???)

Take a close look at data, metadata, source code including the embedded code comments, and further documentation. You now leave the realm of the mere software user to the developer's perspective. This can be a time consuming very close study of the materials. If data is not publicly available, e.g. because it contains information about human subjects, decide if you have a reasonable request to contact the original authors and ask for data access. Work through the examples and analysis scripts included in the research compendium. Play close attention not only to code, but also to code comments as they should include helpful information. A good entry point for your code read may be a "main" script (if provided by the author), makefile, or literate programming document (e.g. an R Markdown file or Jupyter Notebook). If neither of these are available, then start with the code creating the figures for the article (e.g. look for "plot" statements in the code) and trace your way back through the code until you reach a statement where the input data is read. Your impression of the code can help to inform your impression of the article's quality.

If you did not succeed before but the work is relevant for you, spend more time on getting the analysis to run on your computer. Do not hesitate to contact the authors of the paper or authors of the software for help, but follow common error reporting guidelines (e.g. (???) or (???)). For authors it is a great experience to be contacted by an interested and respectful reader!

With regard to the analysis, you may re-implement core parts or the full workflow with a different software. For example, using a tool you know but which was not used in the research compendium. Does your

code lead to the same results, or does it give different ones? Can the differences be explained or are they not significant? Note that such a replication is of very high value for science and you should share your findings with the research compendium's authors and also with the scientific community. Depending on the efforts you put in, write a blog post or even publish a replication research compendium for one or more evaluated research compendia.

If a full replication is not feasible, explore the assumptions you challenge with data and code. Play around with input parameters to get a feel for the changing results. Create exploratory plots for the data as if you would want to analyse it from scratch, without the knowledge of the existing workflow. With your understanding of the code you can extend the method to a new problem or apply it to a different dataset. This deep evaluation of code and data increases your understanding of the authors' reasoning and decisions, and may lead to new questions.

To make sure you can trace your own hands-on changes with the original code and configuration. We recommend initiating a local git repository when starting this pass. You can create branches for specific explorations and easily reset to the original functional state.

During this pass, you should also jot down ideas for future work. This pass can take many hours for beginners and more than an hour or two even for an experienced reader. At the end of this pass, you should be able to reconstruct the entire structure of the paper from memory, as well as be able to identify its strong and weak points. In particular, you should be able to pinpoint implicit assumptions, missing citations to relevant work, and potential issues with experimental or analytical techniques. (???)

You should be able to come up with useful extensions of the used software stack and be able to judge the transferability and reusability of the analysis' building blocks. You should most certainly have improved your programming skills by reading and evaluating other people's code or even trying to extend or improve it.

3. Doing a literature survey

Paper reading skills are put to the test in doing a literature survey. This will require you to read tens of papers, perhaps in an

unfamiliar field. What papers should you read? Here is how you can use the three-pass approach to help. First, use an academic search engine such as Google Scholar or CiteSeer and some well-chosen keywords to find three to five recent highly-cited papers in the area. (???)

No search capability comparable to scientific articles exists for research compendia, though you can of course use generic and academic search engines. More and more journals encourage reproducible research and software and data publication, so that extending your search regular search with keywords such as “reproduction”, “reproducible”, “open data/software/code” may improve your results.

In addition, you can search online platforms where research compendia have been published and tagged as a research compendium (**research-compendium**):

- GitHub label: <https://github.com/topics/research-compendium>
- Zenodo community: <https://zenodo.org/communities/research-compendium>
- OSF tag: <https://osf.io/search/?q=tags:research-compendium>

There is no journal specifically for research compendia yet, but the following ones feature reproducibility, computational studies, or openness in a prominent way and can be a starting point for finding research compendia, if they fit your topic:

- *ReScience*: <https://rescience.github.io/>
- *Information Systems* has a reproducibility editor and special track for invited reproducibility papers: <https://www.journals.elsevier.com/information-systems/>

A lateral approach takes advantage of the parts of a research compendium. If you work with a specific software (tool, extension package, library) or data, find out the recommended way to cite it (and follow it yourself). Most scientific software provides this information in their FAQ or might have a built-in function to generate a citation. Scientific data is often accompanied by a “data paper” or published in repositories with citeable identifiers. Then search for recent publications which cite the referenced software or data.

Do one pass on each paper to get a sense of the work, then read their related work sections. You will find a thumbnail summary of the recent work, and perhaps, if you are lucky, a pointer to a recent survey paper. If you can find such a survey, you are done.

Read the survey, congratulating yourself on your good luck. Otherwise, in the second step, find shared citations and repeated author names in the bibliography. These are the key papers and researchers in that area.

You can also find shared software or data and use them as a seed for a next iteration.

Download the key papers and set them aside. Then go to the websites of the key researchers and see where they’ve published recently. That will help you identify the top conferences in that field because the best researchers usually publish in the top conferences.

Also check where they publish their code and data. It will give you an idea where this community interacts online and can even lead you to research compendia under development.

The third step is to go to the website for these top conferences and look through their recent proceedings. A quick scan will usually identify recent high-quality related work. These papers, along with the ones you set aside earlier, constitute the first version of your survey. Make two passes through these papers. If they all cite a key paper that you did not find earlier, obtain and read it, iterating as necessary. (???)

If a majority cites or uses a key software, technology, or dataset, then evaluate it and include it in the next iteration.

4. Related work

If you are reading a paper to do a review, you should also read Timothy Roscoe’s paper on “Writing reviews for systems conferences” (???). If you’re planning to write a technical paper, you should refer both to Henning Schulzrinne’s comprehensive web site (???) and George Whitesides’s excellent overview of the process (???). Finally, Simon Peyton Jones has a website that covers the entire spectrum of research skills (???). Iain H. McLean of Psychology, Inc. has put together a downloadable ‘review matrix’ that simplifies paper reviewing using the three-pass approach for papers in experimental psychology (???), which can probably be used, with minor modifications, for papers in other areas. (???)

We are working on an extended version of this matrix to provide space for notes about software, data, results of the reproduction, and application of the methods. See the corresponding repository issue for details and provide your feedback: <https://github.com/nuest/how-to-read-a-research-compedium/issues/2>

If you are reviewing a research compedium, a more detailed checklist is given in the “rOpenSci Analysis Best Practice Guidelines” (???), which are partially even automated for R-based research compendia (???), and the Journal of Open Research Software’s guidelines for reviewing research software (???).

5. Acknowledgements

The first version of this document was drafted by my students: Hossein Falaki, Earl Oliver, and Sumair Ur Rahman. My thanks to them. I also benefited from Christophe Diot’s perceptive comments and Nicole Keshav’s eagle-eyed copy-editing. I would like to make this a living document, updating it as I receive comments. Please take a moment to email me any comments or suggestions for improvement. Thanks to encouraging feedback from many correspondents over the years. (???)

In the spirit of the original paper, we would like to make this a living document and invite readers to provide comments or suggestions for improvement via email, as part of this preprint, or on the GitHub repository: <https://github.com/nuest/how-to-read-a-research-compedium>. The repository also includes open questions and is where the paper’s authors openly discuss.

References