

CCS Compendium: A template and tool support for Open and Transparent Science

Wouter van Atteveldt, Felicia
Loeberbach, Mickey Steijaert, Mariken
A.C.G. van der Velden, Kasper Welbers
Department of Communication Science, VU
Amsterdam

Anne C. Kroon, Joanna Strycharz,
Damian Trilling
Amsterdam School of Communication Research

The growing movement towards transparent and open communication science is hindered by practical difficulties in sharing research data and analysis scripts. Moreover, many data sets used in communication research cannot be openly shared due to privacy and copyright restrictions. To remedy this, we introduce the CCS Compendium, consisting of a standardized folder structure, workflow automation support, and a web-based manual and open science checklist. This stimulates researchers to adopt reproducible and transparent analysis practices, and it allows third parties to reproduce research with a single command that installs and runs the required software. Finally, it mitigates the issues preventing data sharing by facilitating the sharing of encrypted or desensitized intermediate data. By making open science as fun and easy as possible, we contribute to the adoption of open science practices in the field of communication and aim to be part of the broader credibility revolution in the quantitative social sciences.

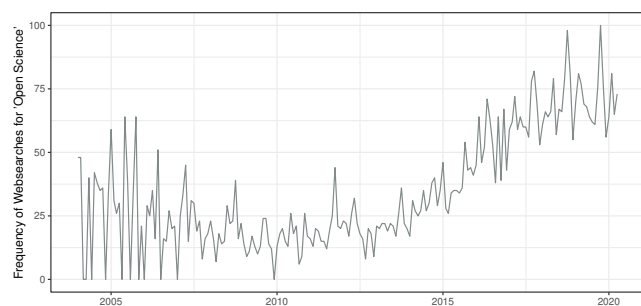
Keywords: open science, sensitive data, reproducibility, workflow

In the past decade, the trustworthiness of social scientific research has been under fire in both scientific and public debate, resulting from the inability to reproduce certain key findings. While the psychological discipline in particular has been subject to scrutiny (Anderson et al., 2015), the reproducibility crisis is more widespread: “Data on how much of the scientific literature is reproducible are rare and generally bleak” (Baker, 2016, p. 452). In the wake of these alarming observations, we are reminded of the importance of the “show me” norm of science (Merton, 1973). If we are not forthcoming in making our data and research process transparent and reproducible, we not only risk undermining trust in the conclusions of specific studies, but possibly entire disciplines (Aczel et al., 2020).

Communication science is no exception in this regard (Dienlin et al., 2020).

To improve the standards for transparency and reproducibility, scholars have rallied under the banner of open science (Miguel et al., 2014; Nosek et al., 2015; Spies, 2013). In its essence, open science is about “making the content and process of producing evidence and claims transparent and accessible to others” (Munafò et al., 2017, p.5). The call for open science has not only been widespread, but became louder, and better received by the scientific community over time. Examples are services such as pre-registration websites as Open Science Framework, AsPredicted, or Center for Open Science, academic journals’ recognition hereof by their implementation of ‘Open Science

Figure 1. Google Trends on ‘Open Science’



Last visited <https://trends.google.com> on March 2, 2021.

Badges’, open science centers at university level, as well as an upward trend in web searches of open science, as Figure 1 demonstrates.

There is also no shortage of ideas and guidelines for how to promote transparency and reproducibility (Klein et al., 2018; LeBel, McCarthy, Earp, Elson, & Vanpaemel, 2018; Nosek et al., 2015; Vazire, 2018). Some of these guidelines, such as the FAIR principles (Wilkinson et al., 2016) have also gathered such wide support that they acquire some institutional leverage, for instance in grant applications. In the field of communication, scholars have voiced calls to embrace these open science practices, and have identified particular threats and opportunities to our discipline (Dienlin et al., 2020; McEwan, Carpenter, & Westerman, 2018; van Atteveldt, Strycharz, Trilling, & Welbers, 2019).

Still, despite the support for the open science movement, the adoption of open science practices is in its infancy in communication science. One of the biggest obstacles is that following open science guidelines can be quite an investment. It requires technical skills and sometimes even legal knowledge, and the additional steps go at the cost of valuable research time. As long as journals and grant committees do not require it, and practices such as data and tool citation are not commonplace, there are simply little incentives for making this invest-

ment (van Atteveldt et al., 2019). Especially for young scholars and less affluent research institutions this means that open science can be a difficult ideal to live up to (see Bowman & Keene, 2018). Another obstacle is that properly following the guidelines of open science is often not the most enjoyable of tasks, and it would be a mistake to dismiss this as inconsequential to the scientific process. Rouder (2016) offers an honest and probably widely shared sentiment that one of his reasons for not making some data open is simply that “it was a pain” (p. 1063). Even if you firmly support the open science movement, risking your career and sacrificing your work enjoyment is a hard bargain.

The goal of the current paper is therefore to make open science “*fun & easy*” by addressing the issues that can make engaging in open science hard, time consuming or a pain. Specifically, we present the *CCS Compendium*¹, consisting of a template with tool support and easy documentation. The proposed compendium support the analysis and collaboration workflow during the research and function as an appendix to a published paper (Nüst, Boettiger, & Marwick, 2018). It provides “a standard and easily recognisable way for organising the digital materials of a project to enable others to inspect, reproduce, and extend the research” (Marwick, Boettiger, & Mullen, 2018, p.

¹For the purpose of peer review, the link to the GitHub repository and checklist website are blinded. The source code of the *ccs-compedium* tool can be inspected at <https://anonymous.4open.science/r/f5210f64-68d7-4653-85f7-c70222aa058a/>. To view the website or install the tool, download the and unpack the *ccs-compedium-zip* file at <https://www.dropbox.com/s/feta6tbq5wht6qo/ccs-compedium.zip?dl=0>. The website *checklist.html* can be directly opened in a browser (although some of the additional information links to the GitHub page and is not available anonymously). The tool can be installed by opening a terminal in the unpacked folder and calling `pip3 install -e .`, after which the tool can be used by running `compedium` as will be described below.

4), containing the study's data as well as the scripts used to process and analyze this data. This greatly increases the transparency of the choices made by the researchers, and makes it easy to reproduce the results, possibly with different settings or data.

CCS Compendium's Contribution to Open Science

There are three main ways in which the CCS Compendium makes engaging in open science easier and more productive. First, we provide a *standardized folder structure* for data, analysis scripts, and results which utilizes the open-source `doit` tool. By using this template to share their results, other users can automatically install the required software and correct package versions and automatically run the scripts to reproduce the published results.

Second, we present a new open source tool called `ccs-compendium`. This tool supports researchers to set up, configure, and document their compendium, and connect it with GitHub or another version control repository. Additionally, it allows researchers to easily automate their workflow, as well as check whether the results can be automatically reproduced. Moreover, this tool allows researchers to encrypt any data that cannot be shared such as copyrighted or privacy-sensitive data. Sharing the encrypted files makes sure that the whole process from raw data to final results is still transparent, even if the research can only be fully reproduced if the encryption key is available.

Finally, we provide an online open science *checklist* with step-by-step instructions tailored to the needs and skills of the individual researcher. We offer researchers an online checklist that inquires about their study and experience with open science practices. Depending on their experience, skills and wishes, they receive a step-by-step plan that guides them through the folder structure and optional technical solutions. For example, the researcher will be asked questions about their study design, software used for data collection and analysis and their experience with ver-

sion control software to determine whether preregistration makes sense for this study, and to provide further guidance in using sharing technologies for researchers with less experience in the necessary tools. Based on the collected information, instructions with pointers to further readings and a skeleton for the standardized folder structure will be provided, where researchers can move their data, code and materials to the designated folders.

These tools are not intended to be an additional step at the end of a research project, but are designed to be integrated within the research process. By standardizing the folder layout, encouraging version control for analysis scripts, and providing the `doit` build tool for workflow automation, we support and stimulate the adoption of best practices for open science. Even for a single scientist working on a project by themselves, the use of version control and a standardized build tools offers a layer of transparency and sustainability. Knowing that your scripts and data are stored securely, that previous versions of scripts can always be reviewed or reverted to, and that your results can indeed be automatically recreated from your data and scripts provides an ease of mind that you will also be able to find, reproduce, understand, and explain your analyses in the future.

Relation to existing work

Our proposed CCS Compendium is not the only tool to support reproducible research. For instance, Nüst et al. (2018) mention a tool called `ReproZip` (Chirigati, Rampin, Shasha, & Freire, 2016), but also point out that a compendium can be build by simply using standard packaging mechanisms in R and Python. More recently, Van Lissa et al. (2020) have developed a "Workflow for Open Reproducible Code in Science (WORCS)". There are also online services such as `ResearchBox` that allow sharing so-called "boxes", which are essentially research compendia, online.

While these existing tools offer impressive features, they seem to be targeted at audiences that only partly overlap with the audience of this jour-

nal – that is social scientists, and communication science scholars in particular. For instance, ReproZip (Chirigati et al., 2016) assumes quite some familiarity with several tools and techniques used in computer programming. WORCS (Van Lissa et al., 2020), in contrast, is “targeted towards scholars working in R on a private computer” (p. 3) and “offers an RStudio project template that is, effectively, a ‘one-click solution’” (p. 3).

As these examples illustrate, there is a tension between more technical solutions and “one-click” solutions that promise to take care of all required steps automatically. Low level approaches excel in offering a lot of flexibility, but require a certain level of technological skills. One-click solutions, nonetheless, lower the barrier for researchers to use them, but also restrict them in their choices and result in “black-box” solutions which workings are not always fully understandable to the researcher. Communication Science is a very heterogeneous field, which means that our compendium should not be primarily tailored towards one computer language or statistics package, and also not (implicitly) assume one specific research method. At the same time, while computational skills are on the rise within the community, our compendium should be easy to use without requiring extensive technical skills.

In particular, we argue that there are three ways in which CCS Compendium offers a unique contribution to open and reproducible communication science. First, it directly addresses a core problem in data sharing in communication science: We often cannot share our raw data because it is privacy-sensitive, proprietary or copyrighted. Our CCS Compendium deals with this problem in two ways: (a) by providing an easy way to (de)encrypt sensitive files; and (b) by including an *intermediate data* folder which can contain aggregated or desensitized versions of the sensitive files. This way, when users have access to the decryption key (or the sensitive files themselves), they can reproduce the results from raw data. Even without this access to the encrypted data, the research can still be re-

produced, starting from the intermediate data files. The intermediate data folder thereby also gives a solution for files which are accessible online but which cannot be redistributed, such as twitter message details.

Second, the CCS Compendium is platform and tool independent. Although the main tools (`doit` and `ccs-compendium`) are Python-based, the compendium works with any analysis environment that can be automatically executed, such as R, Stata or SPSS. Hence, we offer something that goes beyond the existing solutions by explicitly catering to the diversity in methodological approaches as well as preferences for computer programs in our field.

Finally, the CCS Compendium does not try to hide the underlying collaboration and open science tools such as version control (e.g. Github), package management (e.g. Python and R virtual environments), and process management (e.g. `doit`). Although CCS Compendium sets up the initial configuration for these tools, the user still directly uses those underlying tools to achieve their respective goals. Many users already use these tools, and hiding or mirroring their functionality would just cause confusion and frustration. Moreover, we believe learning about these tools is essential for developing and disseminating the culture of open and reproducible communication science and helping researchers work and collaborate more efficiently and sustainably (Klein et al., 2018).

Relation to other Open Science Practices

Pre-Registration. While our online research compendium aims to open up the choices a researcher makes in the process from raw data to final analyses, this would not prevent the process of ‘cooking’: “One of its numerous processes is to make multitudes of observations, and out of these to select those only which agree or very nearly agree. If a hundred observations are made, the cook must be very unlucky if he can not pick out fifteen or twenty that will do up for serving” (Babbage, 1830). Selectively reporting results that

‘work’ – in contrast to those that ‘do not work’ – inflates the false positive rate (or Type 1 error) of published results. Giving a full-fledged plea on why to use pre-registration is beyond the scope of our paper. Furthermore, we do not want to advocate that exploratory research, typically less suited for pre-registration, should not be conducted nor reproducible. Yet, our compendium is an ideal means to demonstrate that a pre-registered plan has indeed been followed. Our interactive compendium website therefore explicitly asks whether the researcher has considered pre-registration, and provides a “show-me tutorial” that describe the purpose and advantages, but also the cost of, pre-registration. In addition, it links to a consensus-based checklist for writing a pre-registered report (Aczel et al., 2020). To integrate pre-registered reports with our online compendium, a researcher can either use a hyperlink to the pre-registration platform of the researcher’s choice and/or store the pre-registered report in the online research compendium.

Open Access Publishing. There exists a broad international consensus these days that advocates for free and open online access to academic information, both for publications and data. Our online research compendium facilitates this movement in multiple ways. First, for everyone, regardless of their levels of coding skills, our generalized folder structure includes a `results/report` folder where the researcher can store the latest version of the research paper that (s)he is legally allowed to upload. The compendium checklist also reminds the researcher to upload the legally allowed latest version. Second, for interested users, the compendium offers workflow support for writing the research paper using systems that enable to embed analysis code in the paper itself, such as (R)Markdown. In this way, there is a full integration of the theory generating hypothesis with the (empirical) data testing them or exploring potential alternatives. For those users that aim to learn how to work with (R)Markdown, we have added a “show-me tutorial” to our website.

CCS Compendium Design and Workflow

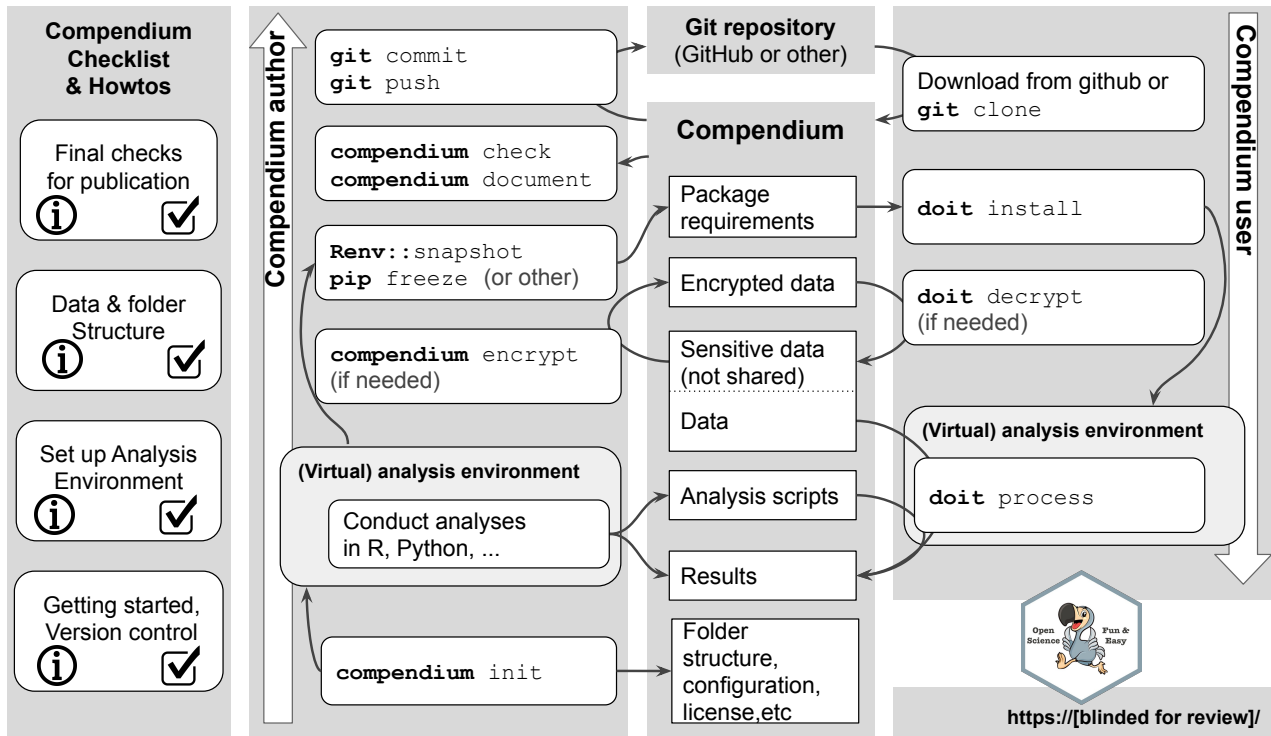
A guiding principle for designing the compendium was to limit additional work for the user as much as possible. In particular, integrating the compendium into existing research workflows should be as easy as possible. Figure 2 summarizes its design and demonstrates how users and authors of the compendium can integrate this to their existing research workflows. In the following sections, we will describe the CCS Compendium in more detail from two perspectives: (a) the perspective of a user who wants to reproduce existing research that uses the compendium (right side of the figure); and (b) the perspective of an author who wants to create a compendium for their research.

Compendium users: just do it

In principle, having access to (1) the data set and (2) the analysis script is enough to reproduce any result reported in a given study. Yet, often between raw data and final analyses lie pre-processing or cleaning steps, and not all conducted analyses may make it to the final paper. As a replicator, you therefore not only require the (original) data set, as well as the scripts to create the variables and conduct the analyses. The replicator is also in need of instructions on how the author(s) exactly walked the path from raw data to final analyses. This is where the compendium comes into play. In contrast to written or verbal instructions, a compendium contains this information in a machine-readable form.

One of the guiding design principles for our compendium was that it should have as little as possible dependencies: it runs on Windows, MacOS, and Linux; and it is unlikely to depend on some specific OS version now or in the future. The compendium is based on a system called `doit`. `Doit` is written in Python, which nowadays is pre-installed on almost all computers and can be installed easily and free of charge if it is not. Installing the `doit` package takes just a one-line terminal command: `pip3 install doit` (Win-

Figure 2. CCS Compendium Design and Workflow



dows) or `sudo pip3 install doit` (Linux and MacOS).

How does this work practically? Suppose that to reproduce a specific paper, you need to (a) install several R packages; (b) decrypt encrypted data; (c) re-shape the data; and (d) conduct the final analysis. The compendium contains a configuration file that describes the various tasks representing these steps.

The first step the user would take is to download the compendium, probably from GitHub. Second, they navigate to the downloaded folder, open a terminal, and enter `doit`. This will run all necessary tasks will be run in the correct order. In this example, it will first run `doit install`, which will install the needed R packages in the correct version (a). Then it will run `doit decrypt` (for b), and run `doit process` which will run all data cleaning and processing steps (c,d) in the order determined by the compendium author.

As a user, you don't have to know which packages to load, which additional resources to down-

load from the Internet, and how to run the processing scripts. `doit` also makes sure that everything is just installed *for this specific compendium only* – it uses a virtual environment to make sure it does not affect the rest of your computer. Of course, because all these steps are described in the configuration file, you can inspect and run every step (either with a specific `doit` command or manually), ensuring complete transparency.

Our compendium website includes “show-me tutorials” for both cloning a Github repository as well as a set of how-to documents for (first-time) users of the online research compendium. We also provide an example project to illustrate all steps.

Compendium authors: `ccs-compendium`

The purpose of the open source `ccs-compendium` tool (GitHub link blinded for peer review) is to make the creation, documenting, testing, and publishing of a compendium as easy as possible. In addition to the tool, we offer an

interactive online guide and checklist (*link blinded for peer review*) that guides the author through the process. Again, the only requirement is that a single Python package needs to be installed, in this case via `pip3 install [package]` (Windows) or `sudo pip3 install [package]` (Linux and MacOS).

As Figure 2 illustrates, a compendium author starts by running `compendium init` in their command line. This will initialize the folder structure, add configuration files, a license file, etc. This happens at the very beginning of a project. Then, the researcher conducts the research itself. This can be done using any set of tools the researcher wants to use. No adaptations to one’s workflow are necessary: it is only expected to save the files in the folder structure that has been created.

After finishing the analysis, the researcher can encrypt data that cannot be publicly shared using `compendium encrypt`. Then, the researcher can create a snapshot of the packages and their versions used in the analysis using the normal command for the tool (e.g. `pip freeze` for Python, `Renv::snapshot()` for R). `compendium check` can be used to verify the contents of the compendium, and will also prompt the researcher to run the snapshot command if the file is not present or outdated. The `compendium document` command will generate documentation files and a process graph based on the headers as described above. Finally, the author can publish the compendium using `git commit` and `git push`. Thus, the whole workflow from creating the compendium structure to publishing it on GitHub consists of 5 simple steps, as summarize in Figure 2 above.

To make the threshold as low as possible, our accompanying website guides the researcher through all steps and offers additional help where needed. It also makes it easy to quickly check whether all requirements for a successful replication of the compendium are met. To make it “fun & easy” to reproduce others’ work, the site also provides a *dodo reproduction badge* (shown in the corner of

Figure 2). Introducing open science badges have shown to increase the reporting of open data, open materials in a complete and comprehensive way, as well as pre-registration plans in psychological sciences (Grahe, 2014; Kidwell et al., 2016).

Table 1 gives an overview CCS Compendium standardized folder structure created by the `compendium init` command. Generally speaking, `data` stores data files, `source` stores analysis scripts and other source code, and `results` stores the final results (papers, figures, cleaned data sets, etc.). The sections below will elaborate on the `raw-private`, `intermediate`, and `tmp` folders

Private data. When data used by the researcher has no identifiable or identified information and is not protected by Intellectual Property laws, the data can be publicly shared online in the online compendium. Examples hereof are speeches given by politicians in parliament or in other public domains or fully anonymized survey data collected by the researcher herself. The initial version of such data should be stored in the folder `data/raw`.

However, the type of data typically studied by communication scholars often does not meet these criteria. In fact, media content is often copyrighted, data collected about individuals through experiments, survey research, but also through social media as well as transcripts of qualitative interviews can be privacy-sensitive, and data acquired from third parties often cannot be shared or published under terms of use (Van Atteveldt, Althaus, & Wessler, 2020). Thus, sharing such data often exposes researchers to the dilemma between privacy and openness. To facilitate sharing personal or copyrighted data, our compendium includes encryption functionalities that enable researchers to share their data in an encrypted form with the key only available to the researcher in question. To enable (first-time) users to encrypt their data, the accompanying website includes a “show-me tutorial” that explains the working of the encryption functionality. The encrypted data will subsequently be stored in the folder

Folder	Published	Contains
data	yes	All data files
raw	yes	Raw (unprocessed) data
raw-private	no	Sensitive or proprietary data that cannot be shared
raw-encrypted	yes	Encrypted files containing private data
intermediate	yes	Processed or anonymized data that can be shared
tmp	no	Processed data that cannot (or need not) be shared
source	yes	All analysis scripts (source code)
processing	yes	Processing scripts that produce (intermediate) data
analysis	yes	Analysis scripts that produce results
lib	yes	Utility files that do not produce data or results
results	yes	Final analysis results

Table 1

CCS Compendium Folder Structure

data/raw-private-encrypted and can only be decrypted with the key set by the owner of the compendium.

Intermediate data. The data/intermediate folder is shared as part of the repository. Crucially, this allows for splitting the processing of private or sensitive data into two parts. The first script takes the minimal steps to anonymize or aggregate the data such that it is no longer sensitive, yielding a desensitized data file in data/intermediate that is as close as possible to the original private data. The next script(s) then work on this intermediate data to further clean, process, or analyse the data. Since the intermediate file can be freely shared, this allows other researchers without access to the private data to reproduce all steps except for the first anonymization step, while still keeping the provenance and transparency of the anonymization process intact. Thus, when the research obtains access to the private data or decryption key, they can fully replicate the results from the raw data.

Temporary data. Finally, there is a folder data/tmp for processed or created data that is not shared as part of the compendium. This folder can simply be used to temporarily store the output of resource-intensive processing scripts. However, another use of this folder is to serve as a location for data that can be downloaded from pub-

lic sources (such as the Twitter API or an online data set) but that may not be shared publicly by the researcher. By including a processing script without dependencies that downloads this data to the data/tmp folder, other researchers can easily download the data for themselves (using the `doit` command as normal) and reproduce the analyses.

Workflow automation. If the analysis involves scripts that can be run automatically (such as an R or Python script, stata do-file, or SPSS production job), the researcher can also use the automated workflow provided by `doit`. For this, they need to add a small header to each script to indicate the data files it needs and creates and what command to use to run the script. For example, this is the header for the `example.py` script packaged with the tool:

```
#!/usr/bin/env python3
#DEPENDS: data/raw-private/secret.txt
#CREATES: data/intermediate/upper.txt
#TITLE: Example script that ...
```

This header tells the computer that it is a python script, that it needs the `secret.txt` private data file and produces the `upper.txt` intermediate data file. Moreover, it provides a title and description to be used in documentation. For his header,

the `doit` tool can automatically determine which scripts need to be run in which order. Moreover, when the researcher calls `doit`, it will only run scripts that are edited or whose input data has changed. This makes it very convenient to use the tool also during the analysis phase, and makes it much easier to collaborate on a project as your co-author can just run `doit` and it will automatically (re)run any cleaning scripts or processing scripts needed.

Literate programming: Integrating compendium and research article. The typical workflow in communication science is to produce the various figures, tables, and statistical reports in an external tool, and copy the results into a Word (or LaTeX) manuscript or report. This means that it is not always clear which script produced which results, and moreover it creates additional work if a figure needs to be changed for e.g. a revision. Finally, it can create inconsistencies if e.g. a table is edited without also changing references to these tables or statistical tests in the main text.

A solution for these issues is to integrate the paper in the analysis workflow. A popular way to do this is by using RMarkdown, a format included in RStudio that integrates code blocks with normal text, and supports formatting, tables, figures, and even references in EndNote or BibTeX format. With RMarkdown, you include the commands to generate the various figures directly in the document. This is then *knit* into an output document in PDF or Word which can be published or submitted.

Note that even though RMarkdown is part of RStudio, it can also be used to include e.g. Python or Stata code. Alternative solutions also exist, for example SWeave and PWeave which integrates R or Python into latex. Another possibility is to create `.tex` files containing figures or other outcomes automatically from the analysis scripts and use `input` to include these in the LaTeX manuscript.

The CCS Compendium supports all these workflows by treating e.g. an RMarkdown file similarly to the other analysis scripts. So, the researcher only needs to include the header as shown above,

and `doit` can then automatically create the output file, and it will also run any preprocessing scripts needed and will re-create the output file if any of the inputs have changed. Integrating the output document into the workflow like this ensures both maximal transparency and productivity.

As a final note, even though it is possible to create an APA compliant MS Word file using either RMarkdown or LaTeX, it would be an added step towards transparent and open science if journals would accept submissions directly as Markdown or LaTeX PDF files (the latter already being the norm in the natural and computational sciences).

Conclusion and Discussion

With this paper, our compendium template at *link blinded for peer review*, and the accompanying website *link blinded for peer review*, we aimed at providing easy to use tools to enhance transparency and reproducibility of communication science research. In particular, we hope to have lowered the barrier for those who want to make their research transparent and reproducible, but did not know where to start or lacked technical skills.

Our main argument is that ensuring transparency and reproducibility should not be considered an extra task at the end of a research project, but be a guiding principle from the start. Extending an argument by van Atteveldt et al. (2019), who suggest using a version control service like Github from the start of a project, we suggest to organize not only the code, but also data and output in a structure that ensures transparency and reproducibility. Moreover, the workflow automation stimulates and rewards researchers to see the compendium as an integral part of their analysis workflow, rather than as an afterthought. Notably, we do not advocate a rogue “just put everything online” mentality, but explicitly provide guidance for handling data that – legally and/or ethically – cannot be publicly shared with everyone.

When following our proposed approach, we believe that researchers can easily comply with the increasing demand – by funders, by their institu-

tions, by their peers – for transparency and reproducibility.

Limitations

We do not want to claim that our proposed compendium is the only or even the best way to make open and transparent science fun and easy. While we believe that our standardized structure can be beneficial for very different types of studies, there are use cases where its applicability may be less obvious. One example for this may be qualitative studies and data. While there is some work that discusses how to implement Open Science practices in qualitative research such as interviews (Haven & Van Grootel, 2019; Steinhardt, 2020), it seems fair to say that qualitative studies are rarely fully replicable—nor do they aspire to be—as researchers’ subjectivity is considered an inseparable element of data analysis. Researchers’ interpretation of the situations, persons, or dynamics under study might be impossible to recreate, and computer tools for qualitative analysis often do not lend themselves to automatic reproduction. Regardless, we believe that sharing codebooks, instrumentation, field notes, and possibly encrypted transcripts can help others understand the analysis choices and claims made in the final research paper. This will help understand how scholars reached their conclusions, ultimately increasing the transparency of the research process. However, one would need to critically assess to what extent the different elements of the compendium fit the community’s needs.

A second limitation is that – despite our efforts to make using our compendium as easy as possible – some additional skills that have not been part of the typical skillset of communication scientists in the past may be necessary. To make use of all our suggestions, the user will need to perform actions like running some scripts or editing some configuration files. All of this is not difficult and does not require specific training, but may feel unusual for users who mostly work with graphical interfaces.

Third limitations are the legal and ethical risks

involved in data sharing that are not fully tackled by the compendium. While the encryption function for data that due to privacy considerations or copyright cannot be publicly shared solves some issues related to making data public, it does not offer a solution to handling the "data minimization" principle in open science (as introduced in the General Data Protection Regulation). More specifically, this principle states that only the strictly necessary data shall be stored (which can be achieved in the compendium by storing only data strictly necessary for reproducibility), but also that data should be stored not longer than necessary for the original aim with which the data was collected. This second requirement is at odds with the aim of the research compendium.

Future work and perspectives

We hope that, as more and more studies use our compendium, it will become even clearer which needs it fulfills well, and where it can still be improved. As the compendium we present is fully open, we hope that future work by both interested colleagues and ourselves will lead to gradual improvements to ensure a best-possible fit with the discipline’s need. For example, we can see value in supporting e.g. integration with Docker containers or cloud services such as Google Colab or Code Ocean that allow code to be run without any local installation.

In the longer run, we also see possibilities for integrating research compendiums into the submission (and publication) process with journals. Especially in combination with formats like Markdown or LaTeX, one could envision that in the manuscript of the future, tables, figures, etc. are not inserted by hand, but dynamically created from a research compendium. This could also automatically support a semantic annotation of research results in such a way that it becomes possible to search for, for example, all experimental studies with voter turnout as a dependent variable, and automatically compare all experimental setups and outcomes. Enabling this is more of an organi-

zational than a technical matter. After all, some journals already require replication packages (Key, 2016), and it is mainly a matter of journal policy and workflows to decide on how they should look like.

Last but not least, we hope that our approach can be a useful tool in teaching. When educating the next generation of communication scientists, it may serve as a practical guideline that helps students to implement advice like “keep a copy of the original data” and “make all steps transparent and reproducible” without needing to reinvent the wheel. By learning to implement the compendium from the beginning of the research process, following principles of open and reproducible science can become a natural part of research of the future.

References

- Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., ... others (2020). A consensus-based transparency checklist. *Nature human behaviour*, 4(1), 4–6. <https://doi.org/10.1038/s41562-019-0772-6>
- Anderson, J. E., Aarts, A. A., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., ... others (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Babbage, C. (1830). *Reflections on the decline of science in england: And on some of its causes, by charles babbage (1830). to which is added on the alleged decline of science in england, by a foreigner (gerard moll) with a foreword by michael faraday (1831)*. (Vol. 1). B. Fellowes.
- Baker, M. (2016). Reproducibility crisis? *Nature*, 533(26), 353–66. <https://doi.org/10.1038/533452a>
- Bowman, N. D., & Keene, J. R. (2018). *A layered framework for considering open science practices*. Taylor & Francis. <https://doi.org/10.1080/08824096.2018.1513273>
- Chirigati, F., Rampin, R., Shasha, D., & Freire, J. (2016). Reprozip: Computational reproducibility with ease. In *Proceedings of the 2016 international conference on management of data* (pp. 2085–2088). ACM. <https://doi.org/10.1145/2882903.2899401>
- Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., ... others (2020). An agenda for open science in communication. *Journal of Communication*. <https://doi.org/10.1093/joc/jqz052>
- Grahe, J. E. (2014). *Announcing open science badges and reaching for the sky*. Taylor & Francis. <https://doi.org/10.1080/00224545.2014.853582>
- Haven, T. L., & Van Grootel, L. (2019). Preregistering qualitative research. *Accountability in Research*, 26(3), 229–244. Retrieved from <https://doi.org/10.1080/08989621.2019.1580147> <https://doi.org/10.1080/08989621.2019.1580147>
- Key, E. M. (2016). How are we doing? Data access and replication in political science. *PS - Political Science and Politics*, 49(2), 268–272. <https://doi.org/10.1017/S1049096516000184>
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., ... others (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS biology*, 14(5), e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Hofelich Mohr, A., ... Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, 4(1), 1–15. <https://doi.org/10.1525/collabra.158>
- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018).

Manuscript under review, please do not distribute

- A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1(3), 389–402. <https://doi.org/10.1177/2515245918787489>
- Marwick, B., Boettiger, C., & Mullen, L. (2018). Packaging data analytical work reproducibly using r (and friends). *The American Statistician*, 72(1), 80–88. <https://doi.org/10.1080/00031305.2017.1375986>
- McEwan, B., Carpenter, C. J., & Westerman, D. (2018). On replication in communication science. *Communication Studies*, 69(3), 235–241. <https://doi.org/10.1080/10510974.2018.1464938>
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... others (2014). Promoting transparency in social science research. *Science*, 343(6166), 30–31. <https://doi.org/10.1126/science.1245317>
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, 1(1), 1–9. <https://doi.org/10.1038/s41562-016-0021>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... others (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Nüst, D., Boettiger, C., & Marwick, B. (2018). How to read a research compendium. *arXiv preprint arXiv:1806.09525*.
- Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavior research methods*, 48(3), 1062–1069. <https://doi.org/10.3758/s13428-015-0630-z>
- Spies, J. R. (2013). *The open science framework: improving science by making it open and accessible*. University of Virginia.
- Steinhardt, I. (2020). Learning Open Science by doing Open Science. A reflection of a qualitative research project-based seminar. *Education for Information*, 36(3), 263–279. <https://doi.org/10.3233/efi-190308>
- Van Lissa, C. J., Brandmaier, A. M., Brinkman, L., Peikert, A., Struiksmā, M. E., & Vreede, B. (2020). *WORCS: A Workflow for Open Reproducible Code in Science*. <https://doi.org/10.17605/OSF.IO/ZCVBS>
- Van Atteveldt, W., Althaus, S., & Wessler, H. (2020). The trouble with sharing your privates: Pursuing ethical open science and collaborative research across national jurisdictions using sensitive data. *Political Communication*, 1–7.
- van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Toward open computational communication science: A practical road map for reusable data and code. *International Journal of Communication (19328036)*, 13. Retrieved from <https://ijoc.org/index.php/ijoc/article/view/10631>
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, 13(4), 411–417. <https://doi.org/10.1177/1745691617751884>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... others (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3. <https://doi.org/10.1038/sdata.2016.18>