

Tenzing and the importance of tool development for research efficiency

Marton Kovacs^{a,b}, Alex Holcombe^{c,*}, Frederik Aust^d and Balazs Aczel^a

^a*Doctoral School of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary*

^b*Institute of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary*

^c*School of Psychology, University of Sydney, Sydney, NSW, Australia*

^d*University of Amsterdam, Amsterdam, Netherlands*

Abstract. The way science is done is changing. While some tools are facilitating this change, others lag behind. The resulting mismatch between tools and researchers' workflows can be inefficient and delay the progress of research. As an example, information about the people associated with a published journal article was traditionally handled manually and unsystematically. However, as large-scale collaboration, sometimes referred to as "team science," is now common, a more structured and easy-to-automate approach to managing meta-data is required. In this paper we describe how the latest version of *tenzing* (A.O. Holcombe et al., Documenting contributions to scholarly articles using CRediT and *tenzing*, *PLOS One* **15**(12) (2020)) helps researchers collect and structure contributor information efficiently and without frustration. Using *tenzing* as an example, we discuss the importance of efficient tools in reforming science and our experience with tool development as researchers.

Keywords: Research collaboration, *tenzing*, authorship

1. The changing research process in team science

Methodological innovations and technological advancements have changed how scientific research is conducted. Today, for many projects individual researchers do not have sufficient knowledge and resources to perform all parts of the work. Consequently, these scientific projects are conducted by large teams of researchers [2]. For example, some multi-lab replications in the social sciences have dozens or even hundreds of participating researchers. When this many contributors are involved, seemingly trivial tasks, such as listing all contributors and their affiliations on the title page of a manuscript, turns into a time-consuming and error-prone endeavor [2]. The increasing numbers of project contributors have highlighted the importance of indicating who did what in a project. This, in turn, has further increased the documentation that must be done [3] (Holcombe, 2019). To reflect the shift toward giving equal acknowledgment to all who contributed to a paper, not just its writers, we will use the term 'contributors' rather than 'authors.'

* Corresponding author: Alex Holcombe. E-mail: alex.holcombe@sydney.edu.au.

2. Why manuscript submission is so bothersome

Two main challenges arise when managing contributor information in large-scale projects. First, each contributor's metadata - names, affiliations, and funding information - must be collected and properly managed. Traditionally, the corresponding author is solely responsible for this task. Many journals now require reporting of who contributed what to a research project, and even when journals do not require it, some research teams prefer to do it. One increasingly-popular way to document this information in a standardized and machine-readable format is by using the CRediT taxonomy [4]. The taxonomy outlines fourteen different types of contributions (see <http://credit.niso.org/>). As there are no widely-used tools, guidelines, or best practices on how to manage contributors' metadata, researchers typically use inefficient *ad hoc* methods to collect this information, such as emailing each collaborator and then assembling the collected information directly in the manuscript. Such *ad hoc* approaches are tenable when the number of contributors is small, but they are error-prone and inefficient in larger projects. Furthermore, when corresponding authors do not collect contributors' metadata until just before or at the time of submission to a journal, *ad hoc* approaches can delay the submission of the manuscript.

Second, contributor information is typically collated and formatted manually in manuscript sections such as the title page, the funding information section, and the contributor information section. Creating these outputs can be frustrating as different journals require different formats. For example, some journals require full contributor names while others prefer the use of initials. To the best of our knowledge, until *tenzing* there was no tool to facilitate or automate these tasks. Another source of frustration is that researchers often have to re-enter the same metadata cell by cell in the journal's submission portal.

3. Tenzing and other tools can help

To facilitate managing contributor's metadata and automate their reporting, we developed *tenzing* [1], an online app (<https://martonbalazskovacs.shinyapps.io/tenzing/>) and R package (<https://github.com/marton-balazs-kovacs/tenzing>) for documenting contributorship. Since the release of the first version, we have received feedback and feature suggestions from the research community, and professional user interface advice from ChaiOne (ChaiOne.com). Based on these inputs, we have updated the design and code of the app and released *tenzing* v0.2.0. (available at the previous link)¹.

To collect the contributors' metadata, research teams are provided with a standardized Google spreadsheet that they can edit simultaneously (see Fig. 1). In the spreadsheet, each row contains the information for a different researcher. The columns are for their various metadata, presently: name, email address, affiliations, funding acknowledgment, and ORCID ID.

The spreadsheet also contains fourteen columns corresponding to the fourteen contribution types of the CRediT taxonomy [4]. Researchers can indicate their (intended²) contributions according to the specific roles by clicking a checkbox. The definitions of each CRediT role, which are displayed when hovering over the label, help users report their contributions in a standardized manner.

¹Development is continuing, and we encourage others to provide feedback or contributions on the GitHub page of the project (<https://github.com/marton-balazs-kovacs/tenzing/issues>) or by contacting us by email (marton.balazs.kovacs@gmail.com).

²We encourage users to use the contributors table as a planning tool early in research projects and iteratively update the record as the project progresses (Holcombe et al. 2020).

Order in publication	Firstname	Middle name	Surname	Conceptualization	Data curation	Formal analysis	Funding acquisition
1	Marton		Kovacs	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Alex		Holcombe	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Frederik		Aust	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Balazs		Aczel	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 1. A snippet of the standardized spreadsheet that serves as the only input for *tenzing*. The spreadsheet is available at: <http://bit.ly/tenzingTemplate>.

The screenshot shows the *tenzing* app interface with three main steps:

- 1. Create your infosheet**: A button labeled "Duplicate and edit the [infosheet template](#)".
- 2. Load your infosheet**: A section with two tabs: "Local file" and "URL". Under the "URL" tab, there is a text input field containing the URL "https://docs.google.com/spreadsheets/d/1o6jIX1OeQpFgDac0jdiqHj5-Z1JzFn3". Below the input field are two buttons: "Upload from url" with an upload icon and "Review infosheet" with an eye icon.
- 3. Download the output**: A list of five buttons: "Show author contributions text", "Show author list with affiliations", "Show XML file (for publisher use)", "Show *papaja* YAML", and "Show funding information".

Fig. 2. A screenshot of the *tenzing* app. In the second step, users can choose between uploading a local file or pasting the URL of the filled-out spreadsheet.

When it is time to create a manuscript file, an author can load the filled-out contributors table into *tenzing* by pasting a link to the Google sheet into the *tenzing* app or, if they have downloaded it as a file,

Marton Kovacs^{1,2}, Alex Holcombe³, Frederik Aust⁴, Balazs Aczel¹

¹Institute of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary, ²Doctoral School of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary, ³School of Psychology, University of Sydney, Sydney, NSW 2006, Australia, ⁴University of Amsterdam



Fig. 3. Screenshot of a title page created by *tenzing*. The app automatically detects if there are multiple first authors and creates a note by listing the names and highlighting the corresponding author.

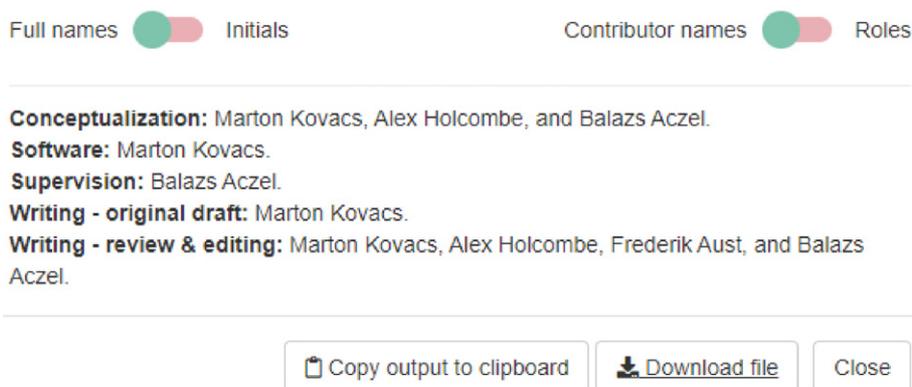


Fig. 4. A screenshot of the contributorship output according to the CRediT taxonomy. Users can choose to list the CRediT roles after each corresponding name as well.

by uploading that file (e.g., in XLSX-format; see Fig. 2). *Tenzing* then validates the contributors table to ensure that the metadata are properly structured and no values are missing or duplicated.

Based on the contributors table, *tenzing* can create five different outputs with just one click for each. Three of these outputs are human-readable and designed for inclusion in manuscripts. They are the title page, the list of contributorship information, and one or more sentences that report funding information. The other two outputs are structured, machine-readable outputs, one in YAML [5] format and one in JATS-XML [6] format. Each output can be downloaded or copied to the clipboard by clicking a button.

The title page output provides the names and affiliations of all contributors in the annotated list format expected by many journals (see Fig. 3).

The contributorship output lists each CRediT role together with the corresponding researchers' full name or initials (see Fig. 4). In many journals, this text is expected to be part of a section with a name such as "Author Note" or "Authors' Contributions."

The funding output lists the full names or initials of contributors who received funding for the project and the name of the funder (see Fig. 5). This output is suitable for the funding information article section used by many journals.

The YAML output (see Fig. 6) is for researchers who write their manuscript in R Markdown and use the *papaja* package [7] to create PDFs and Microsoft Word files in APA format.

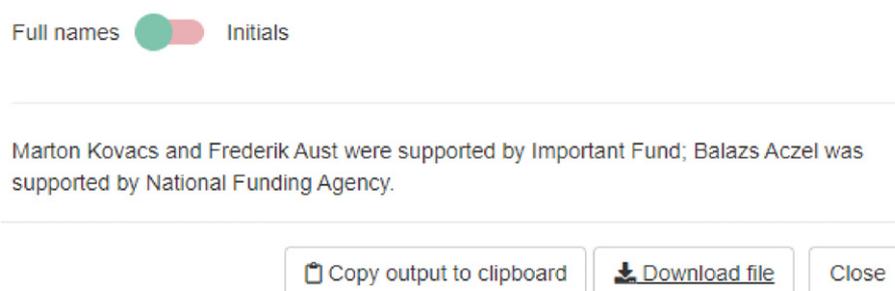


Fig. 5. A screenshot of the funding acknowledgment output. The app automatically lists the names of multiple contributors if they were supported by the same fund.

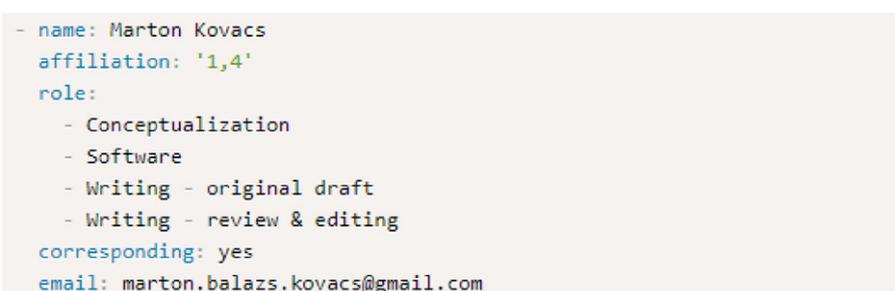


Fig. 6. A snippet of the YAML output which can be used in manuscripts written with the R package *papaja*.

The Journal Article Tag Suite (JATS) is an XML format used to describe scientific literature published online. [Find out more about JATS XML](#)



Fig. 7. A snippet of the JATS-XML output created with *tenzing*.

The JATS-XML output is the format that publishers use to make CRediT information available alongside articles in a machine-readable fashion (see Fig. 7). Publishers generate it from the information that contributors enter into journal management and submission systems, which, as mentioned above, can be very laborious. Unfortunately, no publisher currently allows contributors to skip this step by accepting a pre-generated JATS-XML file such as that provided by *tenzing*. We consulted with several journals, publishers, and preprint servers who indicated that they support the concept, but at the time of writing, none of them has added this feature to their submission system.

4. Reforming science with efficient tools

Whereas *tenzing* focuses specifically on documenting contributorship, it also showcases a general need for new tools that address the inefficiencies that emerge from changing research pipelines. By efficiency, we do not necessarily mean an increase in the number of scientific outputs, but rather allowing researchers to spend more of their time on tasks that improve the quality of their research. Efficient research practices may thereby reduce errors, as well as researcher frustration [8].

High-risk organizations and for-profit companies allocate substantial resources to automatization and to the development of standard guidelines to avoid mistakes, increase efficiency, and thereby increase their productivity [9,10]. Academic researchers in many sciences, in contrast, may not normally think of their work in terms of efficiency. This is unfortunate because efficiency may be an ethical requirement in light of the fact that researchers' time is often funded by taxpayers [11].

The uptake of new tools that address inefficiency can be slow, because of the burden of learning how to use them. It is therefore important that new tools both facilitate the completion of routine tasks and are easy to master. Tools that facilitate routine tasks free up time and can facilitate the adoption of new best practices. In the case of *tenzing*, researchers may initially be attracted to the app for its ability to generate a title page and funding information, but continue on to use its functionality to document each author's specific contributions using the CRediT taxonomy.

5. Tools developed by researchers for researchers

Tool development by researchers faces a number of obstacles. While researchers are often the ones who recognize the need for new or better tools, they often lack the skills, resources, and infrastructure to develop them. For-profit entities, on the other hand, have the necessary resources, but are less interested in keeping their services free-to-use for the research community [12,13] (Taylor, 2016; Posada & Chen, 2018). For researchers, acquiring the required skills, such as programming languages or user interface design, can be time-consuming and the contribution of this investment to career advancement is uncertain. Currently, tool development is not adequately rewarded by academia. Developers do not receive the same credit as researchers who conduct experiments or develop research methodologies, even though a good tool can have a substantial impact, by speeding knowledge accumulation thanks to the research efficiency gains. One reason developers often do not receive credit is that journal articles are the currency of academia, but few journals publish tool-focused articles. Instead, tools are often popularized via other means of communication. Furthermore, if a tool is introduced through the traditional route of a journal publication, these publications often are not cited by researchers who use the tools in their research [14,15].

Another challenge for developers is that tools, unlike research projects, can require continued attention long after the publication of the associated journal article - in particular, when the tool is widely-adopted. The work associated with a scientific study is completed upon publication in a journal. For tools, in contrast, publication in a journal may increase adoption and increase the demand for upgrades and continued maintenance to ensure continued stability and usability. Such maintenance work occupies developers without providing an opportunity to publish new journal articles. Thus, tool development is a long-term commitment that is rarely honored in the academic reward system and therefore can put developers at a career disadvantage - with the incurred disadvantage paradoxically more pronounced for widely-adopted tools than for less-adopted ones.

In the case of web applications, web hosting is another obstacle facing tool developers. For-profit companies continuously invest money to support their interactive web content. Researchers, on the other

hand, often struggle to secure comparably-sustained funding. As a consequence researchers often rely on free services, which limit the speed or the number of concurrent users (e.g., the free Shiny server³ hosted by RStudio that serves *tenzing*).

Finally, the ability of new tools to increase research efficiency is often restricted by the lack of interfaces with other applications that researchers use. For example, *tenzing* should ideally interface directly with publishers' submission management portals. While some of these portals have an application programming interface (API) that allows other entities to programmatically communicate with it, we do not know of any that support adding contributor metadata, not to mention CRediT information. Most submission management portals are owned by for-profit entities that may not be motivated to facilitate interfacing with free-to-use external tools [7,8].

6. Conclusions

In the present paper, we showcased how the *tenzing* app helps researchers to collect and structure contributor information. We argued that tools such as *tenzing* can play an important role in improving the efficiency and indirectly the reliability of scientific work. Presently, scientific tool development faces several challenges. It is either done by the researchers who are rarely trained for this task and struggle to guarantee the long-term maintenance of their tools, or it is done by professional corporations who are less interested in keeping their services free for the research community and interoperable. We hope that the issues this paper highlights would generate further discussions among scientific stakeholders and usher in a new age of research efficiency.

Acknowledgements

We thank Julien Colomb (<https://orcid.org/0000-0002-3127-5520>) and the ChaiOne team for their contribution to the new version of *tenzing*. We also thank Barnabas Szaszi (<https://orcid.org/0000-0001-7078-2712>) for his suggestions regarding the manuscript. Finally, we would like to thank all those who provided feedback on the first version of *tenzing*.

Authors' contribution

Conceptualization: Marton Kovacs, Alex Holcombe, and Balazs Aczel.

Software: Marton Kovacs.

Supervision: Balazs Aczel.

Writing - original draft: Marton Kovacs.

Writing - review & editing: Marton Kovacs, Alex Holcombe, Frederik Aust, and Balazs Aczel.

About the Authors

Alex Holcombe is a professor of psychology at the University of Sydney. In his laboratory, he investigates visual perception and cognition. Outside the lab, for over fifteen years Alex has worked on open access,

³See: <https://www.rstudio.com/products/shiny/shiny-server>, accessed July 13, 2021.

research transparency, and improving reproducibility in academic roles with organizations such as PLOS and the Association for Psychological Science. E-mail: alex.holcombe@sydney.edu.au.

Marton Kovacs is a Ph.D. student at ELTE University in Budapest, Hungary, where he works in the Decision Lab. His Ph.D. focuses on the measurement and improvement of research efficiency. He participates in the development of tools for helping researchers, such as *tenzing* or the SampleSizePlanner. He also takes part in several metascience studies focusing on the credibility and transparency of psychological research findings.

Frederik Aust is a post-doctoral fellow in the Psychological Methods department at the University of Amsterdam. He completed his Ph.D. at the University of Cologne, Germany, in 2020. In his research, Frederik develops and tests mathematical models of human learning and memory, studies Bayesian statistics, and develops tools to improve the computational reproducibility of psychological science.

Balazs Aczel is an associate professor at ELTE University in Budapest, Hungary, where he serves as the vice-dean for research. Dr. Aczel received his Ph.D. at the University of Cambridge, UK, in 2010. Since then, he has led the Decision Lab at ELTE. His research activity covers decision biases, the psychology of stupidity, and the mechanisms of cognitive control. In meta-research, his team develops user-friendly tools and guidelines in several areas, including reporting contributorship and conducting multi-analyst projects.

References

- [1] A.O. Holcombe, M. Kovacs, F. Aust and B. Aczel, Documenting contributions to scholarly articles using CRediT and *tenzing*, *PLOS One* **15**(12) (2020), <https://pubmed.ncbi.nlm.nih.gov/33383578>, accessed July 13, 2021.
- [2] D.H. Sonnenwald, Scientific collaboration, *Annual Review of Information Science and Technology* **41**(1) (2007), 643–681.
- [3] A.O. Holcombe, Contributorship, not authorship: Use CRediT to indicate who did what, *Publications* **7**(3) (2019), 48.
- [4] L. Allen, J. Scott, A. Brand, M. Hlava and M. Altman, Publishing: Credit where credit is due, *Nature News* **508**(7496) (2014), 312.
- [5] YAML, *Wikipedia*, <https://en.wikipedia.org/wiki/YAML>, accessed July 13, 2021.
- [6] D. Lapeyre, Introduction to JATS (Journal Article Tags Suite, XML.com, October 12, 2018, <https://www.xml.com/articles/2018/10/12/introduction-jats/>, accessed July 13, 2021.
- [7] F. Aust and M. Barth, papaja: Prepare APA journal articles with R Markdown [Computer software]. R package version 0.1.0.9987, July 7, 2020, retrieved from <https://github.com/crsh/papaja>, accessed July 17, 2021.
- [8] B. Aczel, M. Kovacs and R. Hoekstra, The role of human fallibility in psychological research: A survey of mistakes in data management, November 5, 2020, PsyArXiv Preprints, <https://doi.org/10.31234/osf.io/xcy kz>, accessed July 13, 2021.
- [9] J.N. Rouder, J.M. Haaf and H.K. Snyder, Minimizing mistakes in psychological science, *Advances in Methods and Practices in Psychological Science* **2**(1) (2019), 3–11.
- [10] K.E. Weick, K.M. Sutcliffe and D. Obstfeld, Organizing for high reliability: Processes of collective mindfulness, *Crisis Management* **3**(1) (2008), 81–123.
- [11] E. Giesen, Ethical and efficient research management: A new challenge for an old problem, *International Journal of Metrology and Quality Engineering* **6**(4) (2015), 406.
- [12] M. Taylor, Elsevier's increasing control over scholarly infrastructure, and how funders should fix this, Saupod Vertebra Picture of the Week, May 22, 2016, <https://svpow.com/2016/05/22/elseviers-increasing-control-over-scholarly-infrastructure-and-how-funders-should-fix-this/>, accessed July 13, 2021.
- [13] A. Posada and G. Chen, Inequality in knowledge production: The integration of academic infrastructure by big publishers, *ELPUB* (2018), <https://hal.archives-ouvertes.fr/hal-01816707v1>, accessed July 13, 2021.
- [14] Editorial: Giving software its due, *Nature Methods*, published February 27, 2019, citation is: *Nat Methods* **16**, 207 (2019). <https://doi.org/10.1038/s41592-019-0350-x>, accessed July 13, 2021.
- [15] A.M. Smith, D.S. Katz and K.E. Niemeyer, Software citation principles, *PeerJ Computer Science* (2016), <https://peerj.com/articles/cs-86>, accessed July 13, 2021.