

# Resource papers as registered reports: a proposal

Emiel van Miltenburg, Tilburg University, The Netherlands

c.w.j.vanmiltenburg@tilburguniversity.edu

---

**Letter abstract** This is a proposal for publishing resource papers as registered reports in the Northern European Journal of Language Technology. The idea is that authors write a data collection plan with a full data statement, to the extent that it can be written before data collection starts. Once the proposal is approved, publication of the final resource paper is guaranteed, as long as the data collection plan is followed (modulo reasonable changes due to unforeseen circumstances). This proposal changes the reviewing process from an antagonistic to a collaborative enterprise, and hopefully encourages NLP resources to develop and publish more high-quality datasets. The key advantage of this proposal is that it helps to promote *responsible resource development* (through constructive peer review) and to avoid *research waste*.

---

## 1 Introduction

A common sentiment in NLP is that the creation of corpora and benchmarks is under-appreciated (Rogers, 2020; Sambasivan et al., 2021), even though resources are one of the driving factors of progress in our field. Moreover, the measurement of progress critically depends on having solid benchmarks. If authors are weary of producing new resources, we all suffer the consequences. How can we avoid this?

### 1.1 Barriers to resource production

Generally speaking, there seem to be two barriers to resource production: funding and appreciation. Building resources requires time and money, and researchers may only be willing to invest time in a project if it could lead to a publication in a respectable venue.

To make resource-building an attractive proposition, we somehow need to convince potential resource authors that their time will be well-spent. One way to do this is to provide a guarantee that their paper will be published. Of course, we would need to have some form of quality control, to make sure that the final resource will be useful to our community. Luckily, such a process already exists in the form of registered reports.

### 1.2 Registered reports

Registered reports are papers that are reviewed in two phases (Chambers, 2019; Henderson and Chambers, 2022). First, authors submit a research proposal, with a clear motivation and outline of the methodology. (Sim-

ilar to a preregistration, see van Miltenburg et al. 2021.) This proposal is reviewed until authors and reviewers agree on the research plan. This agreement means that the paper is accepted in principle. Once the approval is in, authors carry out their study and report their results as specified in the proposal. Then they submit their final paper for the second review phase. In this phase, reviewers check whether the authors followed their proposed methodology. Any changes should be indicated by the authors, with a clear motivation for why those changes were made. Reviewers may not criticize the methodology anymore, but can only comment on the quality of the reporting. Once this is approved, the paper is published.

### 1.3 Earlier discussion in NLP

Van Miltenburg et al. (2021) proposed preregistration and registered reports as potentially helpful innovations in NLP. They suggested that virtually all paper types in NLP are amenable to preregistration.<sup>1</sup> In response, Søgaard et al. (2023) argued that there are also some downsides to preregistration that may outweigh the benefits.<sup>2,3</sup> Nevertheless, they also see

---

<sup>1</sup>The argument is mostly based on what Lakens (2019) calls the *positive externalities* of preregistration. He argues that the core value of preregistration is “to allow others to transparently evaluate the capacity of a test to falsify a prediction, or the severity of a test.” This idea is often not applicable in NLP, but many benefits remain. See Sarafoglou et al. 2022 for a survey among researchers to determine the benefits of preregistration.

<sup>2</sup>E.g., preregistration may increase administrative workload, as also pointed out by Sarafoglou et al. (2022); Hostler (2023).

<sup>3</sup>A full discussion of the authors’ arguments goes beyond the scope of this letter, especially since the authors agree preregistra-

enough value in the idea of registered reports to make a counter-proposal: “limit preregistration to research for which our risk tolerance is low” (p.90).

Søgaard et al. (2023) roughly define risk as *the cost of being wrong*, which in NLP often means that we lose compute and human hours. They argue that this cost is often acceptable, especially in comparison with the human tragedy that may result from clinical trials, so we do not need to burden ourselves with the overhead that risk minimization strategies (such as registered reports) typically bring. On reflection, it does seem true that the risk in NLP is often lower than in the medical field, but the cost of being wrong can still be significant.

Grainger et al. (2020) coin the term *research waste*, and highlight different ways in which we may produce such waste. If you take the wrong approach, you lose researcher and GPU time, and waste the efforts of the volunteers, crowd workers, or consultants involved in your research. Registered reports can be used to prevent this situation. At the same time, they also enable us to carry out ethics review where it is most relevant: in the preparation stage. This immediately solves the problem of after-the-fact ethics reviewing, where we may spot issues, but authors may no longer be able to resolve them.<sup>4</sup>

**Contribution.** This letter proposes registered reports to support the creation of resources (through peer feedback early on in the process) and to avoid research waste. The proposal is painted in broad brush strokes to emphasise the big picture. If this proposal is successful, we can work out the finer details.<sup>5</sup>

## 2 Process outline

The general writing process for registered reports has been described elsewhere (e.g., Henderson and Chambers 2022; Kiyonaga and Scimeca 2019). What would the process look like for resource papers? Here is a brief sketch of what this process could look like if NEJLT would accept registered reports.

### 2.1 Review phase 1

The first review phase is all about your plans. This means that authors will have to write about:

1. **The purpose of the resource.** Why do you want to collect the data? What secondary purposes could the resource also be used for? These

tion/registered reports can be beneficial for our field—we should just work out the proper conditions and guidelines.

<sup>4</sup>Lakens (2023) makes a similar argument, but his solution is to make institutional review boards also review research methodology, as part of their ethics approval procedure.

<sup>5</sup>The Center for Open Science provides a useful set of resources to get started with registered reports: <https://www.cos.io/initiatives/registered-reports>

questions serve as a guide to inform your answers to the other questions. After having listed the different use cases that you (don't) want to support, you can carry out a requirements analysis to see what is needed (split up into essential or nice-to-have) to actually carry out the relevant task. For benchmark datasets it is important to have a clear definition of the skills that you want to assess or the dependent variables that you aim to operationalise. (See Schlangen 2021; Shimorina and Belz 2022 for inspiration.)

2. **The composition of the resource.** What properties should your resource have, and how do you plan to ensure that the resource will indeed have those properties? Additionally: at what level of granularity should you collect different kinds of information?<sup>6</sup> It is a good idea to prepare a draft data statement (Bender and Friedman, 2018) for your resource.<sup>7</sup>
3. **The development process.** How will you go about developing the resource? How will you ensure that the requirements are met? (Also taking practical and technological limitations into account.) If your project requires a large amount of computing power, what strategies are you using to minimise your carbon emissions? (See Lucioni et al. (2020) for recommendations.)
4. **Ethical considerations.** How are the rights and well-being of participants/crowd-workers, data subjects and other direct/indirect stakeholders taken into account, both during and after the development of the resource? Jamieson et al. (2023) provide questions and considerations to make the resource development process more reflexive. As Henderson and Chambers (2022) note, it is important to consider when to submit a proposal to your local institutional review board (IRB) for ethics approval. For most NLP studies it seems reasonable to first apply to your local IRB before submitting the proposal to a journal. This would strengthen your proposal, and any important changes that are requested during the review process could be approved via an amendment to the original IRB application.
5. **Data stewardship.** How will the data be stored, and what measures will be put in place to maintain the resource and take care of any issues that arise from the publication or use of the resource? For discussion, see for example: Peng et al. 2021; Jernite et al. 2022. As with ethics review, it is rea-

<sup>6</sup>Here one might also consider *k-anonymity* for participants/crowd-workers/data subjects (Sweeney, 2002), i.e. ensuring that each property or combination of properties is shared by at least *k* individuals.

<sup>7</sup>A step-by-step guide for writing data statements is available at this URL: <https://techpolicylab.uw.edu/data-statements/>

sonable to contact your local data steward about the measures you should take to responsibly collect and share data. (In some cases, you may be required to carry out a [Data Protection Impact Assessment](#).) At some universities, the IRB process already incorporates a form on data management to protect any data subjects.

This is more or less equivalent to writing an introduction, theoretical framework, methodology, and ethical considerations section.

### When to submit a proposal?

What is the right time to submit a research proposal? This is an open question, as we know that the annotation process is often cyclical, with multiple rounds of revision before an appropriate model and a set of guidelines has been developed (see [Pustejovsky et al. 2017](#), for example). However, most project parameters are likely to be known after a small-scale pilot study. (By keeping the pilot small, we are still minimising research waste.) Even if the exact model and annotation scheme are not fully fixed yet, the methodology and feasibility of the study are clear. At that point, research proposals may be submitted for review.<sup>8</sup>

### Reviewing

Reviewing the proposal is similar to how it is currently done at NEJLT: you submit the paper to the journal, and an editor assigns reviewers to your proposal. The reviews themselves should be constructive, focusing mainly on the methodological and ethical issues:

1. Does this resource address a current need in NLP research?
2. Is the proposed dataset representative of the intended genre or domain?
3. Is the methodology appropriate, valid, and described in sufficient detail?
4. Will the data be responsibly collected and maintained?

What sets registered reports apart from regular submissions is that reviewers can actively contribute to the methodology; they can propose changes to improve the quality of the dataset to be more considerate of any stakeholders, or to make it more broadly us-

<sup>8</sup>A related and common question is: what happens if authors want to change the design of their study, after their research proposal has been accepted? The answer depends on the nature of the changes. Small modifications should be noted and motivated in the final report. Larger modifications may need to be reviewed, or at least flagged to the editor. The Center for Open Science notes in their [Frequently Asked Questions](#) that it is also possible to carry out *sequential registrations* for studies where the design and hypotheses for each subsequent study in a paper is based on previous results.

able. Authors can then refine their proposal before the manuscript is provisionally accepted.

### Should we publish Stage 1 protocols?

An open question here is whether the proposal should be published at this stage, or only when the final report has been accepted for publication. Publication policies differ between different journals: [The Royal Society \(ND\)](#) does *not* publish Stage 1 registered reports before the final manuscript is approved. [Nature Scientific Reports \(ND\)](#) does not publish Stage 1 registered reports either, but *does* require authors to preregister their study in a recognised repository. The preregistration can either be made public, or put under embargo until Stage 2. This matches the recommendations from [Chambers et al. \(2023\)](#), who note that “the journal can also perform the Stage 1 registration process on behalf of authors.” Finally, the publisher [Wiley \(2018\)](#) recommends that journals publish registered reports after passing Stage 1 peer review, but also allows its journals to instead require authors to preregister their study design (similar to Nature Scientific Reports).

[Wiley \(2018\)](#) notes that publishing registered protocols has the advantage of providing transparency and accountability both for journals (showing what reports are in principle accepted, publicly committing to the publication of the final result) and authors (showing *what* they are working on, *when* they developed the ideas for their final publication, and publicly committing to finish the resource). Of course researchers could also feel uncomfortable sharing their research-in-progress for all sorts of reasons, so it may be good to at least offer them the option to put their research proposal under embargo.

## 2.2 Review phase 2

With an in-principle acceptance in hand, authors should aim to carefully follow their original proposal. Deviations from the original plans are possible, but these should be clearly indicated in the report and well-motivated by the authors. Once the dataset has been collected and a full report has been written, the paper can undergo the final review.

### Reviewing

Having already approved the methodology, reviewers now comment on the execution of the project:

1. Has the resource been compiled according to plan, with all deviations clearly marked?
2. Does the report contain all relevant details about the creation and composition of the resource?
3. Is the presentation clear and accessible?
4. Is the resource accessible and easy-to-use?

We should expect resources to be publicly available, unless there are strong arguments in favor of limited accessibility (for example: copyright issues, or privacy of the data subjects).

### Should reviewing be anonymous?

An open question here is whether reviewing in the second phase should proceed anonymously, or whether it is also OK for author names to be revealed at this time. This would certainly make it easier to assess the final resource (which may be hard to anonymise), but may unduly influence the reviewing process.

## 2.3 Publication

Once the paper is ready for prime-time, it can be published as usual. An open question is whether the reviews should be published as well and, if so, whether the reviews should be kept anonymous or not. For transparency reasons, it would be really insightful to publish all the correspondence between authors and reviewers along side the final report. This way, we would get to see the original intentions of the authors, and how the approach was transformed during the review process. Reviewer names could be published on an opt-in basis, so that they might claim credit for the provided service. (This avoids the issue of reviewers holding back their criticism for fear of retribution if their name is published alongside their review, see e.g. [Ali and Watson 2016](#) for discussion.)

## 3 Eligibility

What kinds of resources should be eligible for publications through registered reports? So far this proposal has not set any strict requirements to determine what makes a resource worth publishing in a journal like NE-JLT. To some extent, we can be pragmatic about this issue: authors tend to prefer conferences for smaller contributions, and journals for larger contributions. The administrative hassle for smaller projects may just not be worth the effort of writing a registered report (in [Søgaard et al.](#)'s terms: there is less 'risk' involved), so authors of small studies are not very likely to submit a research proposal.<sup>9</sup> What matters is that authors have clearly thought through their proposal, and are not just letting reviewers do their work. In the latter case, desk rejection seems appropriate. If we do need more guidelines, we can always fall back on the existing ones, that easily carry over to (and indeed overlap with some reviewing questions in) this proposal.<sup>10</sup>

<sup>9</sup>But if authors think that their work should be published as a registered report, there is little harm in letting them carry out a small but high-quality study.

<sup>10</sup><https://www.nejlt.org/review/>

## 4 Conclusion

This letter proposed to offer potential resource authors the opportunity to publish their resources as registered reports, as an addition to the existing paper types. (The resource category would not need be removed.)

The proposal outlined here is more modest than the one put forth by [Van Miltenburg et al. \(2021\)](#), who suggest that *all* types of NLP papers (except position papers) could in theory be published as registered reports. This modesty is not for a lack of ambition; instead, this proposal is offered as a first step, to see if registered reports could actually work for NLP research. And what better way to start, than to support the creation of fundamental resources?

## References

- Ali, Parveen Azam and Roger Watson. 2016. Peer review and the publication process. *Nursing Open*, 3(4):193–202.
- Bender, Emily M. and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.
- Chambers, Chris. 2019. What's next for registered reports? *Nature*, 573:187–189.
- Chambers, Chris, George Christopher Banks, Dorothy Bishop, Sara Bowman, Kate Button, Molly Crockett, Zoltan Dienes, Timothy M. Errington, Agneta Fischer, Alex O. Holcombe, Kai Jonas, Edward Miguel, Marcus Munafò, Brian A. Nosek, Brendan J. Nyhan, David Rand, Daniel J. Simons, Carien van Reekum, Andrew Sallans, Steven Rogelberg, , and David Thomas Mellor. 2023. Registered reports implementation checklist. Available through: <https://osf.io/2m4ct> Originally published on February 4, 2014. The current version was modified on May 8, 2023. Retrieved 13 July 2023.
- Grainger, Matthew J., Friederike C. Bolam, Gavin B. Stewart, and Erlend B. Nilsen. 2020. Evidence synthesis for tackling research waste. *Nature Ecology & Evolution*, 4(4):495–497.
- Henderson, Emma L. and Christopher D. Chambers. 2022. Ten simple rules for writing a registered report. *PLOS Computational Biology*, 18(10):1–9.
- Hostler, Thomas J. 2023. The Invisible Workload of Open Research. *Journal of Trial & Error*. <https://journal.trialanderror.org/pub/the-invisible-workload>.



- Jamieson, Michelle K., Gisela H. Govaart, and Madeleine Pownall. 2023. Reflexivity in quantitative research: A rationale and beginner's guide. *Social and Personality Psychology Compass*, 17(4):e12735.
- Jernite, Yacine, Huu Nguyen, Stella Biderman, Anna Rogers, Maraim Masoud, Valentin Danchev, Samson Tan, Alexandra Sasha Luccioni, Nishant Subramani, Isaac Johnson, Gerard Dupont, Jesse Dodge, Kyle Lo, Zeerak Talat, Dragomir Radev, Aaron Gokaslan, So-maieh Nikpoor, Peter Henderson, Rishi Bommasani, and Margaret Mitchell. 2022. Data governance in the age of large-scale data-driven language technology. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, page 2206–2222, New York, NY, USA. Association for Computing Machinery.
- Kiyonaga, Anastasia and Jason M. Scimeca. 2019. Practical considerations for navigating registered reports. *Trends in Neurosciences*, 42(9):568–572.
- Lakens, Daniël. 2019. The value of preregistration for psychological science: A conceptual analysis. *Japanese Psychological Review*, 62(3):221–230.
- Lakens, Daniël. 2023. Is my study useless? Why researchers need methodological review boards. *Nature*, 613(7942):9–9.
- Luccioni, Alexandra, Alexandre Lacoste, and Victor Schmidt. 2020. Estimating carbon emissions of artificial intelligence [opinion]. *IEEE Technology and Society Magazine*, 39(2):48–51.
- van Miltenburg, Emiel, Chris van der Lee, and Emiel Kraemer. 2021. Preregistering NLP research. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 613–623, Online. Association for Computational Linguistics.
- Nature Scientific Reports. ND. Registered reports. Retrieved from <https://www.nature.com/srep/journal-policies/registered-reports> on 13 July 2023.
- Peng, Kenneth, Arunesh Mathur, and Arvind Narayanan. 2021. Mitigating dataset harms requires stewardship: Lessons from 1000 papers. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1. Curran.
- Pustejovsky, James, Harry Bunt, and Annie Zaenen. 2017. Designing annotation schemes: From theory to model. *Handbook of Linguistic Annotation*, pages 21–72.
- Rogers, Anna. 2020. Peer review in nlp: resource papers. Published on the Hacking Semantics blog on 16 April 2020, retrieved 13 July 2023 from: <https://hackingsemantics.xyz/2020/reviewing-data/>.
- Sambasivan, Nithya, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. “everyone wants to do the model work, not the data work”: Data cascades in high-stakes ai. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.
- Sarafoglou, Alexandra, Marton Kovacs, Bence Bakos, Eric-Jan Wagenmakers, and Balazs Aczel. 2022. A survey on how preregistration affects the research workflow: Better science but more work. *Royal Society Open Science*, 9(7):211997.
- Schlangen, David. 2021. Targeting the benchmark: On methodology in current natural language processing research. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 670–674, Online. Association for Computational Linguistics.
- Shimorina, Anastasia and Anya Belz. 2022. The human evaluation datasheet: A template for recording details of human evaluation experiments in NLP. In *Proceedings of the 2nd Workshop on Human Evaluation of NLP Systems (HumEval)*, pages 54–75, Dublin, Ireland. Association for Computational Linguistics.
- Søgaard, Anders, Daniel Herscovich, and Miryam de Lhoneux. 2023. A two-sided discussion of preregistration of NLP research. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 83–93, Dubrovnik, Croatia. Association for Computational Linguistics.
- Sweeney, Latanya. 2002. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570.
- The Royal Society. ND. Registered reports. Retrieved from <https://royalsocietypublishing.org/rsos/registered-reports> on 13 July 2023.
- Wiley. 2018. Registered reports policy on publishing at stage 1 and stage 2. Version 1.0 dated 10 January 2018, retrieved from <https://authorservices.wiley.com/author-resources/Journal-Authors/submission-peer-review/>

[registered-reports-policy.html](#) on 13 July  
2023.