

IJDC | Conference Pre-print

Do Open Data Badges Influence Author Behaviour? A Case Study at Springer Nature

Rebecca Pearce
Springer Nature

Rebecca Grant
Springer Nature

Abstract

Digital badges have previously been shown to incentivise journal authors to share their data openly. In this paper we introduce an Open data badging project at the Springer Nature journal *BMC Microbiology*. The development of the Open data badge is described, as well as the challenges of developing standard badging criteria and ensuring authors' awareness of the badges. Next steps for the badging project are outlined, which are based on the experiences of the team assessing the badges, the number of badges awarded at the journal to date, and the results of an author survey.

Submitted 16 December 2019 ~ *Accepted* 19 February 2020

Correspondence should be addressed to Rebecca Grant, Springer Nature, London N1 9FN. Email: Rebecca.grant@springernature.com

This paper was presented at International Digital Curation Conference IDCC20, Dublin, 17-19 February 2020

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

In this paper we introduce the Open data badge pilot project at Springer Nature, and describe the implementation of digital badges which aim to recognise and reward authors for sharing their data in publicly-available repositories. We discuss the practicalities of implementing a workflow for awarding badges, as well as the challenges that arise when assessing non-standardised narrative data availability statements for compliance with the badging checks. Finally, we present initial findings on how the Open data badges have impacted author and reader behaviour, and the next steps for the project.

The use of digital badging to encourage and reward authors for practicing open science is not new: the Center for Open Science (CoS) currently supports Open Science Badges for 53 publishers.¹ Initial analysis of the effectiveness of digital badging in motivating data sharing behaviour found that data sharing in the journal *Psychological Science* increased from 3% to 23% (Kidwell et al., 2016). Further research found that digital badges were the only incentive that dramatically increased data-sharing practices among medical and health researchers (Rowhani-Farid, Allen and Barnett, 2017). It is for that reason Springer Nature sought to implement a digital badging program for their journals, building upon the foundation set by the CoS.

In this pilot a set of badging criteria was created in collaboration with CoS, which could be applied consistently across a range of journals and disciplines. The Open data badge is currently being piloted with authors at the journal *BMC Microbiology*, and requires that data availability statements for all published papers are evaluated by a member of the Research Data team (part of Springer Nature's professional staff). If the badging criteria are met, the paper (rather than the author) then receives a badge which is displayed on the landing page for that paper.

Implementing Open Data Badges and Assessing Data Availability

The badging criteria were initially developed with reference to the CoS Open Science Badges. The Open Data badge created by CoS require that authors provide two disclosures: that a DOI or other persistent identifier is provided to access the data in a public, open access repository; and that they provide sufficient information for an independent researcher to reproduce the reported results.² The Springer Nature badges require that a persistent identifier is provided and that the data are deposited in an appropriate repository, but acknowledge that the repository may allow controlled access, and that open licences may not be available. A criterion around the reproducibility of the reported results was not included in the checks, as it was felt that this could not be accurately assessed in a rigorous and consistent manner at scale.

In September 2018, the pilot badging project was launched at the journal *BMC Microbiology*. This journal was chosen because it is a relatively broad-scope journal with an in-house Editorial team, which would make coordination of the pilot more

¹ COS: Open Science Badges: <https://cos.io/our-services/open-science-badges/>

² Badges to Acknowledge Open Practices Wiki: Awarding Badges: <https://osf.io/tvyxz/wiki/2.%20Awarding%20Badges/>

manageable. Additionally, the journal's data policy requires that a statement describing how data can be accessed is included in every manuscript. The number of manuscripts accepted at the journal (around ten per week) was also sufficient to allow us to gather data on the efficacy of the pilot, but low enough to ensure that the team had sufficient time to assess the papers, which was done manually during the pilot.

The aims of the initial pilot were to assess whether Open data badges motivated authors to share their data more frequently; and whether reader engagement increased when published papers had an Open data badge. The pilot also intended to establish what resources are required to undertake badging checks and to what extent this would be scalable across additional journals. The pilot was initially planned to run for 3-6 months, but based on the results of the author survey described below, the pilot period was extended in order to further refine the workflows and to increase the visibility of the badges. During the pilot period, authors were informed of the Open data badges via the journal's website and submission guidelines, and in the letters they received from the journal's editorial office.

Assessing Papers for Compliance

The Research Data team at Springer Nature are involved in research data related projects including the roll out of standard data policies, the data curation service Research Data Support, and the publication of the journal *Scientific Data*. Working with the Open data badges team, the Research Data team developed and tested a set of criteria which could be used to assess whether a published paper should be awarded a badge. The checks were intended to be applied only to the data availability statement published as part of the paper, and any other descriptions of the data within the body of the paper were not checked.

The criteria were as follows:

- A data availability statement must be included with the paper, stating how the data can be accessed.
- The dataset must be deposited in a publicly accessible repository (or an appropriate disciplinary repository, where community mandates apply).
- A DOI, accession code or other persistent identifier must be supplied for the dataset.
- The dataset provided is checked and confirmed as relevant to the related paper.

The checks were made on every paper published at *BMC Microbiology* during the pilot period, and every criterion needed to be fulfilled in order for a badge to be awarded. Papers were assessed by the Research Data team in weekly batches. Each check had to be passed before the next check was conducted, for example if a paper did not include a persistent identifier for the dataset then the relevance of the dataset was not checked and the paper was not awarded a badge. If a paper was approved for a badge, the badge was issued via the badgr.io platform, and an Open data icon was added to the landing page for the paper on the journal's website (Figure 1). The badgr.io platform allows users to create digital badges and to issue them based on a set of standard criteria. The Open data badge record is linked to the DOI of the published paper, and narrative evidence of data sharing is added by including the text of the paper's data availability statement to the badgr.io record. Issuing badges is a manual

process undertaken by the Research Data team, and during the pilot period authors did not receive any notification that their paper had received a badge.

Research article | [Open Access](#) |  [Open data](#) [↗] | Published: 03 October 2018

Air and waterborne microbiome of a pharmaceutical plant provide insights on spatiotemporal variations and community resilience after disturbance

[Filippo Pacchioni](#), [Alfonso Esposito](#), [Elisabetta Giacobazzi](#), [Clotilde Bettua](#), [Paolo Struffi](#) & [Olivier Jousson](#) [✉]

Figure 1. A highlighted example of the Open data badge icon on a paper at BMC Microbiology.

The badging pilot was announced in July of 2018 and manuscripts were assessed from September 2018, to ensure that authors would be aware of the badge criteria before the assessments began. From September 10, 2018 to May 31, 2019, 210 papers have been assessed for badges.

Assessing the Impacts of the Badges

In February 2019, the Open data badges team conducted an online survey of authors who had published in *BMC Microbiology* and whose papers had been evaluated for data badges, to determine what impact the pilot had on data sharing practice. Authors were asked about their awareness of the pilot and their opinions on data sharing generally. In addition, data on the number of badges awarded during the pilot period were compiled as well as comparative data on data sharing in papers published before the pilot began. We also assessed whether there was an increase in the deposition of non-mandated data types to repositories during the pilot period (e.g. data for which there is no community expectation to share in suitable repositories, as there would be for genomics data for example).

Lastly, we examined usage statistics for papers with and without badges to determine if there was a different level of engagement with papers awarded badges (based on paper accesses and downloads) and engagement with the data. To estimate engagement, click-through rates on the digital badges were collected (these lead to the badge's landing page on badgr.io which provides information about the deposited data). Taken together, these data were intended to inform potential adjustments to the pilot project to allow us to better evaluate the impact of the badging on author behaviour.

Impact of the Open Data Badges to Date

From September 10, 2018 to May 31, 2019, 210 papers have been assessed and 65 badges awarded, meaning that 31% of papers were eligible for a badge. In comparison, 20% of the papers published in the three month period before the pilot would have been eligible for a badge, indicating that the badging pilot has impacted positively on

data sharing. The majority of badges awarded were for papers where there are existing community expectations that data should be deposited in a publicly available repository however (e.g. genetic data). We have awarded two badges to papers where data sharing was not mandated by a research community, funder, or institution, in comparison to the pre-pilot assessment where no authors had shared their data without being mandated to do so.

Respondents to our author survey largely supported data sharing practices (17 of 18 stated it was very or moderately important to them). Moreover, 10 of 18 stated that the Open data badges would make them more likely to submit their research to the journal. However despite authors' positive attitude towards Open data badges generally, those surveyed were largely unaware of the pilot or their paper's badge status. Where authors were aware of the pilot, most had learned about it on the journal's website rather than via the manuscript submission system or author letters.

To assess engagement, 31 papers with badges were compared to a randomly selected group of 31 papers without badges. We found that overall, papers with badges had higher engagement with a lower bounce rate (60% bounce rate for papers with a badge, and 70% for papers without) and a higher download rate (53 clicks to download papers with badges versus 12 clicks). Papers with badges also had a higher access rate (695 accesses on average) than those without badges (402 accesses on average). Of the 31 papers with badges, users clicked through to view the badge's landing page from 19.

Discussion

Author Visibility of the Pilot Project

Despite the efforts of the Editorial teams to ensure the authors were aware of the pilot, it became clear that the placement of messaging and the lack of post-badge follow up resulted in low awareness of the pilot among authors, thereby making it difficult for us to firmly determine the effect of the badging on author behaviour. In fact, only one of the 18 survey respondents stated they had seen messaging about the pilot in the letters they received from the journal. Additionally, eight stated they were unaware of the pilot, and six stated they had seen it mentioned on the website or submission guidelines (one respondent did not answer this question). Lastly, only one of the eight of the badge recipients surveyed indicated that they were aware that their paper had received a badge.

Another unanticipated challenge was the increased volume of genetics-related submissions to the journal during the pilot period. As there are community mandates and well established repositories for this type of data, this could have largely been responsible for the increase in the volume of shared data in the published papers, independent of the badging pilot. It is worth noting that respondents to our author survey did not necessarily adhere to existing mandates however, and many appeared to be uncertain as to what the mandates entailed. In spite of this uncertainty, half of surveyed authors (9 of 18) believed that data sharing is important to progress research in their field.

Without clear acknowledgement from the authors that they were aware of the badging criteria and that this influenced their decision to deposit their data in a repository, it is not possible to determine the exact cause for the increased number of

papers with openly shared data. As part of the next steps for the project we intended to implement the Open data badges at a journal in a discipline which does not have such well-established community data sharing mandates in order to assess their impact further.

Testing and Refining the Badging Criteria

The checks for badging were defined at the beginning of the project, and tested on a set of papers which had been published before the pilot began. In spite of the testing, the unstructured nature of data availability statements meant that challenges arose when applying the checks to a larger volume of papers over an extended period. The checks were initially intended to apply consistently to each paper, but over the duration of the pilot it became clear that flexibility was required. For example, where only a part of the overall dataset had been shared, we decided that a badge should still be awarded. This applied to many of the papers we assessed, and the stricter assessment we had planned (that all data must be shared) would have meant a much lower rate of badging.

In some cases private links to data were included in the data availability statements of published papers, which can occur when an author shares a private link to an unpublished dataset for the purposes of peer review, but does not update it with a DOI or accession code during a later revision. Some repositories provide private links that expire over time, meaning that the data become inaccessible. Where these links were identified during the badging checks, the editorial team was notified, but this issue, and the expertise necessary to identify it, must be taken into account when scaling the badging checks to staff outside the Research Data team.

Additionally some authors reported the availability of their data in other sections of the paper, for example the materials and methods section. In these cases we decided not to award a badge, as reviewing each paper in full would be time-consuming and would also impact on scaling the pilot in future. Focusing compliance assessment on data availability statements could also enable more automated approaches to compliance checking in future. For example, simple natural language processing (NLP) could be applied to the papers and data availability statements to assist in identifying those which should be awarded a badge.

Conclusion

This pilot has brought to light several challenges that will need to be addressed in providing Open data badges, or any other digital badging initiative we launch.

- Authors must be made aware of the badging at submission and there must be follow up communication about the badge criteria to ensure they have the opportunity to comply and to receive a badge.
- Messaging around the badging should provide authors with additional resources they can access for further guidance with respect to appropriate repositories for their data and best practices.
- Authors should be notified of their paper's badge status.
- Visibility of papers with badges should ideally be improved within the online platforms.

- If possible, data availability statements should be reviewed before a manuscript is accepted so it can be queried if necessary, and authors can be given the opportunity to refine or amend it.

Our next steps in the pilot will focus on ensuring author awareness of the badge criteria, which we will address through improvements to the visibility of the Open data badge information in the journal's manuscript submission system. Additionally we intend to create a workflow for notifying authors when their paper receives a badge. Although badgr.io does provide this functionality, it is limited to notifying a single author and therefore an alternative method is required. Addressing the challenges relating to the visibility of the badge on published papers, and checking data availability statements before papers are accepted, will rely on the success of the current pilot. Changes to manuscript submission workflows and journal landing pages require substantial effort, with the manuscript workflow in particular having an impact on editorial processes and how the pilot can be scaled. When improvements to visibility and author notifications have been made we intend to add a second journal to the pilot, in a subject area where no community mandates exist regarding data sharing. We will then reassess the effectiveness of the badging by measuring the number of badges issued, and surveying authors to assess their awareness of the badges.

The pilot project has added further context to existing research on the impact of badges on author behaviour. We identified an 11% increase in data sharing during the pilot period, although the challenges described above have impacted the conclusions we can draw from this. The pilot also indicated an increase in engagement with published papers with badges, including a 53% difference in how frequently the papers were accessed on average, in comparison to papers without badges. Based on the results to date, and positive author feedback on badges when surveyed, the Open badges team intends to continue to explore badging as an incentive for authors. If the addition of a second journal to the pilot shows similarly positive results, Open data badges may be made available for any journal across Springer Nature that wishes to offer them. In addition, the team will consider the development of additional digital badges which reward and recognise Open Science practice.

Data Availability

The dataset tracking the Open badge checks and awarding of badges is available in the figshare repository at [doi:10.6084/m9.figshare.8268161](https://doi.org/10.6084/m9.figshare.8268161).

Acknowledgements

Thank you to Iain Hrynaszkiewicz for comments on the draft of this paper. The pilot project would not have been possible without the support and enthusiasm of the *BMC Microbiology* Editors: Cecilia DeVoto and Akila Sridhar. The Open data badge working group has worked tirelessly to drive this initiative forward -this includes Sowmya Swaminathan, Iain Hrynaszkiewicz, Sadie Forrester, Amye Kenall, and Thomas Appleyard. Our OSCAR web team (Paul Rissen, Moshen Husen) have been critical in implementing and supporting the badging. Our colleague Dave Mellor at the Center for Open Science provided valuable insights into their badging program and helpfully

shared their expertise with us in support of our own badging efforts. In addition, the support team at badgr.io (Nate Otto and Kim Hammond) were incredibly helpful when we were setting up the badges and troubleshooting any technical glitches along the way. The Market Intelligence team led by Dan Penny and Gregory Goodey coordinated the author survey. We thank our colleagues in the Journals Editorial Office (Ruth Baker, Anna Shattles) for helping implement the changes to our submission systems that were necessary for the running of this pilot. Lastly, we thank our production colleagues who helped us track published manuscripts to be assessed by Research Data Support team (Rene Aanstoot and Irina Masalagiu).

References

- Grant, R. & Pearce, R. (2019). *Dataset tracking Open data badges awarded at the journal BMC Microbiology, between October 2018-May 2019* [Data set]. figshare. doi:10.6084/m9.figshare.8268161
- Kidwell, M.C, Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S. & Nosek, B.A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*. doi:10.1371/journal.pbio.1002456
- Rowhani-Farid, A., Allen, M., Barnett, A.G. (2017). What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review* 2:4. doi:10.1186/s41073-017-0028-9