A large-scale dataset on peer review in scientific journals

Giangiacomo Bravo[1], Flaminio Squazzoni[2], Francisco Grimaldo[3]

1. Linnaeus University, Växjö, Sweden, *giangiacomo.bravo@lnu.se*
2. University of Brescia, Brescia, Italy, *flaminio.squazzoni@unibs.it*
3. University of Valencia, Valencia, Spain, *francisco.grimaldo@uv.es*

The TD1306 COST Action *New frontiers of peer review* (hereinafter, PEERE) is a EU funded project aiming to improve the efficiency, transparency and accountability of peer review through a trans-disciplinary, cross-sectorial collaboration. The objectives of this action are: (i) to analyse peer review in different scientific areas by integrating quantitative and qualitative research and incorporating recent experimental and computational findings; (ii) to evaluate the implications of different models of peer review and to explore new incentive structures, rules and measures to improve collaboration in all stages of the peer review process; (iii) to involve science stakeholders in data-sharing and testing initiatives, (iv) to define collaboratively a joint research agenda that points to an evidence-based peer review reform.[1]

To achieve these objectives, PEERE has established collaborations with a group of major publishers (Elsevier, Springer, Wiley) to define a data sharing strategy aimed at providing data on the internal peer review processes of a large sample of journals. This will allow to analyse the dynamics of peer review across different fields and to identify important factors that could help to improve the quality, transparency and accountability of the process.

The data will be acquired according to a data sharing policy that has been developed by PEERE partners and stored in a server of the School of Engineering at the University of Valencia, under the responsibility of Prof. Francisco Grimaldo. Descriptive meta-data will be developed and updated and will be linked with the datasets. This will help the researchers involved in the project to understand the data structure and content. Data access will be granted to PEERE members following a procedure that will ensure full compliance with the PEERE data sharing policy by anyone.

The dataset currently is under construction. Once finished, it will include over fifty variables per interaction (Tab. 1), where an interaction is defined as any exchange between a journal editor, referees and authors of submissions. Taking into account that the number of interactions recorded for each each journal has a a magnitude of 10^3 and that the publishers that have signed or are signing an agreement for data sharing with PEERE host a large number of journals, we expect to be able to build a dataset having an approximate size of at least 50×10^5, and possibly 50×10^6, which represents the larger collection on data on peer review ever created. The dataset will be completed by the late 2016-early 2017 and will be available for analysis in the following months after a preliminary work of recoding and anonymisation.

---

[1] For further details on the project, please contact *info@peere.org*

1. Date Reviewer Invited	19. Reviewer Stop Date	34. Title
2. Reviewer Has Not Responded	20. Total Days with Reviewer	35. Degree
3. Agree Date	21. Reviewer Recommendation	36. Reviewer Role
4. Review in Progress	22. Review Rating	37. Person is a Board Member
5. Reviewer Recommendation In Progress	23. Reviewer Comments to Author nltk	38. Position
6. Reviewer Declined Invitation	24. Reviewer Comments to Author corenlp	39. Country
7. Reviewer Terminated	25. Reviewer Comments to Author length	40. Revision Number
8. Reviewer Uninvited	26. Reviewer Comments to Editor nltk	41. Article Title and Abstract nltk
9. Alternate Reviewer	27. Reviewer Comments to Editor corenlp	42. Article Title and Abstract corenlp
10. Alternate Reviewer Selection Date	28. Reviewer Comments to Editor length	43. Article Type
11. Reviewer Promoted	29. Days Reviewer has to Submit Review	44. Initial Date Submitted
12. Date Review Due	30. Number of days review was late	45. Final Decision Date
13. Date Last Reminder Sent	31. Days between Invite and Review Completion	46. Date of First Decision
14. Total Number of Reminders Sent	32. Days between Agreement and Review Completion	47. First Decision Term
15. Number of Invitation Reminders Sent	33. ReviewerID	48. Date Revision Submitted
16. Number of Reminders Sent Before Due Date		49. Editorial Status
17. Number of Reminders Sent After Due Date		50. Final Disposition Term
18. Review Complete		51. Date Final Disposition Set
		52. All Authors
		53. Number of Required Reviewers
		54. Number of Days Reviewer has to Respond

Table 1: Variables in the current version of the PEERE dataset. The name and characteristics of each journal can be linked to the variable above.

The dataset will have multiple uses. These include (but are not limited to):

- testing the impact of peer review models (e.g., confidential vs. open) on the quality and efficiency of peer review, including style of reviewing, referee and editorial bias, and referee motivation;
- measuring disciplinary and community, context-specific standards of peer review (see Casnici et al. 2016)
- the study of collaboration networks among editors, referees and authors;
- providing a test on different (reputational/material) incentive schemes for referees (see Bohannon 2013; Squazzoni et al. 2013; Squazzoni and Gandelli 2013) based on a large-scale comparison among journals using different systems;
- a textual analysis of referee reports as a tool to control referee bias and measure the quality of referee reports.

By the Symposium time, we hope to be able to present at least a clearly-defined picture of the dataset, possibly together with a preliminary analysis based on a subset of records derived from the journals edited by one of the major publishers involved in PEERE.

References

Bohannon, J. (2013). Who's afraid of peer review? *Science*, 342(6154):60–65.

Casnici, N., Grimaldo, F., Gilbert, N., and Squazzoni, F. (2016). Attitudes of referees in a multidisciplinary journal: An empirical analysis. *Journal of the Association for Information Science and Technology*, In press.

Squazzoni, F., Bravo, G., and Takács, K. (2013). Does incentive provision increase the quality of peer review? An experimental study. *Research Policy*, 42(1):287–294.

Squazzoni, F. and Gandelli, C. (2013). Opening the black-box of peer review: An agent-based model of scientist behaviour. *Journal of Artificial Societies and Social Simulation*, 16(2):3.