Crowd Editions: Participation, Expertise and Automation in Digital Scholarly Editing

By ORCID Anna Busch (University of Potsdam), anna [dot] busch [at] uni-potsdam [dot] de; ORCID Martin Prell (Saxon Academy of Sciences and Humanities in Leipzig), martin [dot] prell [at] klassik-stiftung [dot] de; ORCID Torsten Roeder (University of Würzburg), torsten [dot] roeder [at] uni-wuerzburg [dot] de.

DOI: 10.18716/ride.a.21.0

Crowd Editions: Participation, Expertise and Automation in Digital Scholarly Editing

The idea that scholarly knowledge is generated by the many rather than the few has gained renewed momentum in the early 2000s. As an alternative to small, selective expert editorial teams, projects such as Wikipedia and Wikisource have demonstrated that large, loosely organised digital communities can catalogue, structure and make extensive bodies of knowledge publicly accessible. Such forms of collective knowledge production have since been widely discussed under the term *crowdsourcing* and have given rise to new ways of working, particularly in the fields of cultural heritage research and the digital humanities (Terras 2015; Ridge 2017). In the field of digital editions, too, projects have emerged since the 2010s that open up tasks within the editorial process to a wider public.

However, with the increasing digitisation of humanities work processes, it is not only the technical tools of editorial practice that have changed, but also the institutional and social context of scholarly practice. Keywords such as *Open Science*, *Citizen Science*, *Public Humanities* or *Participation* signal a shift towards viewing research as a social process and towards exploring new forms of collaboration between academic institutions and the public. In the field of digital editions in particular, collaborative working methods have developed in this context, discussed under terms such as *crowdsourcing*, *social editing* or *crowd editing*.

Currently, the question also arises as to what extent crowdsourcing, as a method of digital knowledge aggregation, is falling behind due to competition with generative artificial intelligence processes (Christoforou, Demartini, and Otterbacher 2025) – or what advantages it continues to offer as an effective method of data extraction. Advances in automated processes in areas such as OCR, handwriting recognition or text-based analysis have triggered a new debate on the relationship between algorithmic and human forms of knowledge production. Crowd editions thus also illustrate the general transformation of editorial practice, in which questions of expertise, participation and automation are being renegotiated.

This special issue of the review journal RIDE is dedicated to these developments under the heading of *Crowd Editions*. The aim is to review digital editorial projects that employ participatory working methods and make parts of the editorial process accessible to a wider public. The reviews collected in this volume examine such projects from various perspectives and explore the extent to which participatory methods are compatible with the scholarly requirements of digital editions.

Concepts and Models of Participatory Editing

The discussion of participatory forms of editing is closely linked to a range of terms whose usage in scholarly discourse is not always unambiguous. A distinction is often made between *social editions* and *crowd(sourced) editions*. Whilst social editions usually refer to collaborative projects in which a community makes editorial decisions and edits texts collectively (Crompton, Arbuckle, and Siemens 2013), crowdsourcing generally refers to a form of division of labour in which clearly defined tasks are delegated to a larger group of participants (Martin, Lessmann, and Voß 2008).

In practice, these models often overlap. Many projects combine different forms of participation and cannot easily be assigned to a single category. The term *citizen science* is also increasingly used in the context of humanities research, although its applicability to the humanities is not without controversy (Smolarski, Carius, and Prell 2023).

Historical Development

The involvement of a wider public in academic projects is not a fundamentally new phenomenon. In the natural sciences, forms of voluntary data collection or observation have long existed, which are now summarised under the term *citizen science*. In the digital humanities, the concept of *crowdsourcing* in particular has played a central role. Melissa Terras describes crowdsourcing in the digital humanities as a method whereby tasks traditionally carried out by a small circle of experts are outsourced to a larger group of voluntary or semi-professional contributors (Terras 2015).

The use of crowdsourcing in the digital humanities began to emerge in the early 2000s, in parallel with the development of collaborative platforms such as Wikipedia. An example is the *Distributed Proofreaders* project (Distributed Proofreaders Foundation 2026), founded in 2000 by Charles Franks, in which digitised texts are proofread by volunteers before being uploaded to *Project Gutenberg* (Project Gutenberg Literary Archive Foundation 2026), originally founded by Michael S. Hart in 1971. The work is broken down into small units – usually individual pages – and processed in parallel by many contributors. A key motivation for participants is to be an active part of the crowd and the growing collaborative effort.

The *Transcribe Bentham* project (Bentham Project 2026), launched in 2010, represented a decisive step towards scholarly editorial work. The project invited volunteers to transcribe manuscript pages from the extensive papers of the philosopher *Jeremy Bentham* and to carry out basic TEI encoding. This marked the first time that a demanding editorial task had been systematically opened up to a wider public.

The success of this project has significantly influenced the discussion on participatory forms of editing. Tens of thousands of manuscript pages were transcribed and encoded, and the resulting data is being incorporated into the preparation of the printed complete works of Bentham. At the same time, the project has demonstrated that crowdsourcing can be not only a method for generating large volumes of data, but also a way of involving interested laypeople in scholarly processes (Busch and Roeder 2023).

In this context, Ben Brumfield coined the term *accidental editors* to describe those voluntary contributors who, through their participation in transcription projects, acquire editorial skills and thus effectively become contributors to scholarly editorial projects (Brumfield 2017). This observation suggests that the boundary between professional and voluntary editorial work in participatory projects is becoming increasingly blurred.

Crowdsourcing in the Digital Humanities and in Digital Editorial Projects

In the field of digitised cultural heritage and scholarly digital editions, such approaches have been established since the 2000s. Mia Ridge has demonstrated in several studies that crowdsourcing projects in libraries, archives and museums often serve not only to generate data, but also play an important role in scholarly communication and social participation (Ridge 2017; Ridge, Blickhan, and Ferriter 2021). Examples in this volume include the *Crowdsourcing Wien* project (reviewed by Christian Erlinger) and *What's on the Menu* (reviewed by Janosch Förster). Participatory projects can thus both generate research data and establish new forms of interaction between academic institutions and the public. This is evident not only in historical everyday documents such as postcards and menus, but particularly in material where collective, intangible knowledge is the central focus of scholarly investigation, as in *Making and Knowing* (reviewed by Sarah Lang).

In the field of scholarly editing, this approach may initially seem unusual. Editions are traditionally regarded as highly specialised scholarly products, the creation of which requires extensive specialist knowledge. Accordingly, opening up this process to a wider public was initially viewed with scepticism. In some cases, however – particularly in smaller and geographically dispersed specialist communities – the 'crowd' is also understood as a potentially globally distributed collective of experts. An example of this in this volume is the *Papyrological Editor* (reviewed by Lavinia Ferretti and Elisa Nury), whilst *#everynamecounts* (reviewed by Daniel Burckhardt) demonstrates that initial expert platforms can indeed transform into citizen science portals.

Opportunities and Challenges

Integrating a crowd into editorial workflows promises several advantages. One of the most important aspects is the ability to process large volumes of source material that small research teams could scarcely manage on their own. Particularly in projects involving large manuscript collections or vast amounts of source material, the participation of many individuals can significantly accelerate data collection. However, ethical concerns are sometimes raised when crowdsourcing is viewed merely as a means of obtaining cheap labour and the public engagement aspect is disregarded (Busch, Roeder, and Prell 2025).

Furthermore, participatory projects can play an important role in scholarly communication and social participation. The involvement of interested citizens can help to raise the profile of humanities research and establish new forms of interaction between research institutions and the public. However, this process is by no means one-sided: in addition to recognition for their contribution, citizen scientists also gain in-depth knowledge of the material, thereby creating forms of knowledge distribution that in some respects anticipate traditional forms of scientific public engagement.

At the same time, crowdsourcing projects involve considerable organisational challenges. As several studies have shown, the successful use of participatory methods requires careful project management as well as continuous support for the participating community (Ridge 2017). Crowdsourcing is therefore not a simple tool for reducing workload; on the contrary, it often requires additional resources for community management, technical infrastructure and quality assurance. Worth mentioning in this volume is the review of *What's on the Menu*, not least because the service has since been discontinued by its operators, and the review thus also serves a documentary function.

Issues of academic quality also play a central role. Contributions from a large and heterogeneous group of participants must be reviewed and integrated consistently. In many projects, multi-stage control mechanisms are used to this end, such as editorial checks, peer validation within the crowd, or automated verification procedures. However, it is also the case that a crowd community evolves over time into an *expert crowd*, which in turn can contribute to increasing self-regulation and quality assurance within the community, provided that adequate opportunities for exchange exist.

The discussion on crowdsourcing is also influenced today by advances in the field of artificial intelligence. Automatic handwriting recognition, OCR processes and large language models open up new possibilities for processing historical texts. In the Digital Humanities, such processes now take on tasks that were previously considered typical applications of crowdsourcing, such as the transcription and annotation of historical documents. At the same time, the question arises as to whether human crowd work, or parts thereof, could be replaced by automated processes in the long term.

Within the Digital Humanities, however, this development is increasingly understood not as a simple substitution of human labour by machines, but as a transformation of the work processes themselves. Although automated methods enable the faster processing of large volumes of text, they remain dependent on training data, which often originates from previously manually generated or curated datasets. Crowdsourcing projects have contributed significantly to the creation of such *ground truth* corpora, which serve as training and evaluation data for handwriting recognition, OCR or other forms of automated text analysis. The crowd thus functions not only as a workforce in the editorial process, but also as a producer of key data infrastructures upon which many AI applications in the digital humanities are built.

Current research, therefore, suggests that the relationship between human and machine work is more complex than a simple competitive perspective would suggest. Whilst automated methods now achieve remarkable results in many areas, well-trained citizen scientists can still achieve high levels of accuracy in certain tasks (Brumfield and Evans 2025). This applies in particular to material that exhibits a high degree of diversity (e.g. typescripts and manuscripts) and consists of small or abstract documents (e.g. index cards). Automated methods continue to reach their limits, particularly with highly variable layouts, damaged documents or idiosyncratic handwriting (Werner and Cugliana 2026), whereas human editors can contribute contextual knowledge and interpretative skills.

Furthermore, crowdsourcing and AI methods differ not only in terms of their technical capabilities but also in terms of their scientific functions. In the current DH debate, it is repeatedly pointed out that algorithmic methods often operate as ‘black boxes’, whose decision-making processes are only partially comprehensible to users. Participatory projects, on the other hand, can make editorial decisions visible and enable scholarly work processes to be designed transparently. In this sense, crowd editions contribute not only to data generation but also to the epistemic traceability of editorial work.

At the same time, crowdsourcing projects often fulfil additional roles within the research landscape. They serve, for instance, as instruments of scholarly communication, education or social participation, and make it possible to make historical source collections accessible to a wider public. This aspect can only be replaced to a limited extent by automated processes.

This suggests that future working models may rely more heavily on *hybrid approaches*, combining automated processes with human collaboration. In such models, AI systems can, for example, handle initial transcription or recognition steps, whilst the crowd is deployed for corrections, contextualisation or more complex interpretative tasks. The role of the crowd thus potentially shifts from primary data collection towards tasks of validation, curation and interpretation of algorithmically generated results.

Conclusion and Outlook

The reviews collected in this volume examine crowd-sourced editions from various perspectives. Some projects place particular emphasis on the aspect of *citizen science* and use participatory methods as a tool for educational work and social engagement. Other projects focus more strongly on the *indexing of previously neglected source materials*, such as postcards, menus or other forms of everyday documents. Still others can be understood as *platform-based research infrastructures* that primarily serve specialised academic communities.

The contributions in this volume demonstrate that crowd editions now constitute a diverse and dynamic field within digital scholarly editing. At the same time, it becomes clear that many projects have so far been inadequately documented or scientifically evaluated. Reviews can play an important role here by making projects visible and comparable, situating them within broader research contexts, and ensuring their long-term documentation.

In the long term, the question arises as to what role participatory methods will play in future editorial practice. Crowd editions are not only a tool for managing large volumes of data, but also a testing ground for new forms of collaboration between academia and the public. In a research landscape increasingly characterised by openness, interdisciplinarity and digital infrastructure, they could contribute to a greater understanding of editorial work as a collective knowledge process.

References

Bentham Project. 2026. *Transcribe Bentham. A Participatory Initiative*. https://web.archive.org/web/20260508072837/https://transcribe-bentham.ucl.ac.uk/td/Transcribe_Bentham.

Brumfield, Ben. 2017. “Accidental Editors.” In *Advances in Digital Scholarly Editing: Papers Presented at the DiXiT Conferences in The Hague, Cologne, and Antwerp*, edited by Peter Boot, Anna Cappellotto and Wout Dillen, 69–83. Leiden: Sidestone Press. https://web.archive.org/web/20260410231134/https://www.sidestone.com/books/advances-in-digital-scholarly-editing.

Brumfield, Ben, and Connor Evans. 2025. What’s the Character Error Rate of a Volunteer? Analyzing accuracy in cultural heritage crowdsourcing projects. Digital Humanities Conference 2025, Lisbon, 18 July 2025. Program Agenda, DH2025 LOC. https://doi.org/10.5281/zenodo.16084200.

Busch, Anna, and Torsten Roeder. 2023. „Crowdsourcing in digitalen Editionen – ein Themenband der Rezensionszeitschrift RIDE.“ In *Partizipative Transkriptionsprojekte in Museen, Archiven und Bibliotheken*, edited by Diana Stört and Anita Hermannstädter, 47–49. Berlin: Museum für Naturkunde Berlin (MfN) – Leibniz Institute for Evolution and Biodiversity Science. https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-926062.

Busch, Anna, Torsten Roeder, and Martin Prell. 2025. „‘Die Hölle, das sind die anderen.‘ Crowdsourcing in digitalen Editionen.“ DHd 2025: Under Construction (DHd2025), Bielefeld, Germany. Zenodo, 2025. https://doi.org/10.5281/zenodo.14944585.

Christoforou, Evgenia, Gianluca Demartini, and Jahna Otterbacher. 2025. “Crowdsourcing or AI-Sourcing? Considering the impact of generative artificial intelligence on data annotation tasks.” *Communications of the ACM*, 4 March 2025. https://cacm.acm.org/opinion/crowdsourcing-or-ai-sourcing/.

Crompton, Constance, Alyssa Arbuckle, and Raymond Siemens. 2013. “Understanding the Social Edition Through Iterative Implementation: The Case of the Devonshire MS (BL Add MS 17492).” *Scholarly and Research Communication*. 4 (3). https://doi.org/10.22230/src.2013v4n3a118.

Distributed Proofreaders Foundation. 2026. *Distributed Proofreaders. Preserving History One Page at a Time*. https://web.archive.org/web/20260507101008/https://www.pgdp.net/c/.

Martin, Nicole, Stefan Lessmann, and Stefan Voß. 2008. „Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte.“ Multikonferenz Wirtschaftsinformatik, MKWI 2008, Munich, 26 February 2008 – 28 February 2008, Proceedings.

Project Gutenberg Literary Archive Foundation. 2026. *Project Gutenberg*. https://web.archive.org/web/20260502174313/https://gutenberg.org/.

Ridge, Mia, ed. 2017. Crowdsourcing Our Cultural Heritage. First issued in paperback. Digital Research in the Arts and Humanities. London New York: Routledge, Taylor & Francis Group.

Ridge, Mia, Samantha Blickhan, and Meghan Ferriter, eds. 2021. *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage – community review version*. PubPub. https://doi.org/10.21428/a5d7554f.1b80974b.

Smolarski, René, Hendrikje Carius, and Martin Prell, eds. 2023. “Citizen Science in den Geschichtswissenschaften: methodische Perspektive oder perspektivlose Methode?” *DH & CS - Schriften des Netzwerks für digitale Geisteswissenschaften und Citizen Science*, Vol. 3. Göttingen: V&R unipress.

Terras, Melissa. 2015. “Crowdsourcing in the Digital Humanities.” In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens and John Unsworth, 420–38. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118680605.ch29.

Werner, Nicolas, and Elisa Cugliana. 2026. “Transcribing the Untranscribable: Automating Recognition of Text and Image in Multimodal Medieval Manuscripts from Law to Divination.” *DHd 2026: Not Just Text, Not Just Data (DHd2026)*, Vienna, Austria. Zenodo 2026. https://doi.org/10.5281/zenodo.18696365.

Licence: http://creativecommons.org/licenses/by/4.0/

Published: 2026-05 · build 833c03b (2026-11-10T21:33:03+01:00)