



# Introduction to the Special Issue on Digital Natural and Cultural Heritage: Opportunities and Challenges

**ANDREAS WEBER**, University of Twente, Department of Technology, Policy and Society (TPS),  
Drienerlolaan 5, The Netherlands

**MAARTEN HEERLIEN**, KB - National Library of the Netherlands, Department of Collections, Prins  
Willem-Alexanderhof 5, The Netherlands

**EULÀLIA GASSÓ MIRACLE**, Teyler's Museum, Collections Department, Spaarne 16, The Netherlands

**KATHERINE WOLSTENCROFT**, Leiden University, Leiden Institute of Advanced Computer Science, Niels  
Bohrweg 1, The Netherlands

---

This introduction to the special issue “Digital Natural and Cultural Heritage” has two aims: First, it briefly introduces individual contributions and explains how they relate to one another. Second, it explains our motivation for producing a special issue aimed at bridging the perceived gap between the fields of digital natural and cultural heritage. More concretely, it presents different computational approaches to tackle heterogeneity in digital natural and cultural heritage collections. Here the goal is to make information retrievable, interlinked, and interpretable for researchers and the general public. However, instead of offering final solutions, this special issue rather hopes to highlight the enormous potential that collaborations across domains have. We hope that we offer a starting point for transdisciplinary research in which researchers, curators, and infrastructure developers of both domains communicate and work much tighter together.

CCS Concepts: • **Applied computing** → **Arts and humanities**; • **Human-centered computing** → **Accessibility**; • **Information systems** → **Information retrieval**; **Digital libraries and archives**;

Additional Key Words and Phrases: Digital cultural heritage, digital natural heritage, natural history collections, artificial intelligence, semantic web

## ACM Reference format:

Andreas Weber, Maarten Heerlien, Eulàlia Gassó Miracle, and Katherine Wolstencroft. 2023. Introduction to the Special Issue on Digital Natural and Cultural Heritage: Opportunities and Challenges. *ACM J. Comput. Cult. Herit.* 16, 1, Article 1e (June 2023), 4 pages.

<https://doi.org/10.1145/3597459>

---

## 1 INTRODUCTION

Large and important parts of the world's natural and cultural heritage collections are stored in museums and archives that are difficult to access, even after digitization. The Dutch National Archives in The Hague alone stores more than 142 km of documents, many of them digitized.<sup>1</sup> Europe's natural history museums house more

---

<sup>1</sup><https://www.nationaalarchief.nl/en/our-story>, retrieved January 30, 2023.

---

Author's addresses: A. Weber, University of Twente, Department of Technology, Policy and Society (TPS), Drienerlolaan 5, 7522 NB Enschede, The Netherlands; email: a.weber@utwente.nl; M. Heerlien, KB – National Library of the Netherlands, Department of Collections, Prins Willem-Alexanderhof 5, 2595 BE Den Haag, The Netherlands; email: maarten.heerlien@kb.nl; E. Gassó Miracle, Teyler's Museum, Collections Department, Spaarne 16, 2011 CH Haarlem, The Netherlands; email: e.gasso@teylersmuseum.nl; K. Wolstencroft, Leiden University, Leiden Institute of Advanced Computer Science, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands; email: k.j.wolstencroft@liacs.leidenuniv.nl.

© 2023 Copyright held by the owner/author(s).

1556-4673/2023/06-ART1e

<https://doi.org/10.1145/3597459>

than a billion specimens (plants, animals, minerals) and metadata records including archival material documenting their provenance and collection circumstances [1, 2].<sup>2</sup> The inaccessibility of many of these natural and cultural heritage collections hampers research in a variety of scholarly fields ranging from cultural, colonial, and art history to the biodiversity sciences. Next to palaeographic skills and language proficiency, missing links among handwritten documents, illustrations, and other forms of tangible and intangible heritage form an important obstacle in the field. Moreover, the interpretation of such interwoven collections requires knowledge about their historical contexts and digital preservation strategies [3]. By building upon the results of the international conference “Collect and Connect: Archives and Collections in a Digital Age,” held in Leiden on November 23 and 24, 2020, this special issue focuses on the question of how different computational technologies can help curators, librarians, and archivists to enrich heterogeneous digital natural and cultural heritage collections with contextual information to make them retrievable, interlinked, and interpretable for researchers and the general public.

Organized in the middle of the coronavirus pandemic, the “Collect and Connect” conference attracted a substantial international audience, including researchers, curators of digital heritage collections, developers of large-scale digital infrastructures, and publishers. More than 240 registered participants from all over the world contributed to three research paper sessions, one demo lab, a round table discussion, and three keynote lectures.<sup>3</sup> Over the two conference days, a variety of heterogeneous digital natural and cultural heritage collections were discussed: Next to an examination of natural history collections in Europe and the U.S., speakers also focused on archaeological collections, collections of art, newspaper collections, an Aegean seal collection, and large serial collections of handwritten documents produced in European and colonial administrative settings [4]. Each of these articles also entailed a detailed evaluation of the computational technologies that were used to tackle heterogeneity in such collections. For this special issue, the conference organizers selected five articles that we consider representative of current discussions in the field. To fit with the special issue, all articles have been substantially revised and enriched with follow-up research.

We are, of course, not the first to focus on the heterogeneity of digital natural and cultural heritage. Since the early 2000s, this journal as well as a large number of colleagues all over the world have made major efforts to develop and evaluate technology geared toward increasing the accessibility of digital cultural and natural heritage collections. In particular, mass digitization, semantic web, and AI-driven technologies in combination with the rise of vast digital infrastructures and a strong emphasis on relying on volunteers to annotate collections have changed the landscape fundamentally. The available scholarship and projects are too vast to mention here, so we simply refer to a number of survey articles and recent research we were aware of at the moment of writing this introduction (e.g., References [5–12]). However, despite enormous efforts, we also notice an unfortunate divide between research into digital natural and cultural heritage collections. Instead of working jointly toward a transdisciplinary information infrastructure, as had been envisioned by Karl-Heinz Lampe et al. in the first issue of this journal in 2008, research in both fields has taken a siloed approach leading to a fragmented landscape different infrastructures [13]. Only recently, infrastructure developers and researchers in both fields have started to envision how a Common European Data Space, integrating natural and cultural collections and data according to the FAIR principles, could serve both research communities [14, 15]. We sincerely hope that work in this direction will continue leading to digital infrastructures that allow for the contextualization and enrichment of heterogeneous natural *and* cultural heritage across domains and disciplines. This special issue cannot offer final solutions to the above-mentioned issues. However, by clustering recent scholarship in the field, we aim to showcase and evaluate best practices of how researchers have used different computational technologies to tackle heterogeneity and the lack of links between collections across domains.

<sup>2</sup><https://www.dissco.eu/>, retrieved January 30, 2023.

<sup>3</sup>The conference’s full program can be accessed here: <https://sites.google.com/naturalis.nl/makingsenseproject/conference/program>, accessed January 30, 2023.

Collected, ordered, and inventoried in different formats and modalities over a long time span, many digital natural and cultural heritage collections require curators, researchers, and infrastructure developers to employ tailor-made and labor-intensive assemblages of different computational technologies. Koolen et al. (first article of this special issue) and Ameryan and Schomaker (second article of this special issue) illustrate how AI-driven handwriting recognition tools, named entity recognition, and humans in the loop can be efficiently combined to increase the accessibility of heterogeneous archival collections. Both articles capitalize on heterogeneity in different ways: While the first article focuses on extracting and structuring information in a large serial archive in which handwritten and published text is intertwined, the second article forms part of a wider attempt to rapidly index heterogeneous archives in the natural heritage domain by making optimal use of human labellers. Both articles tellingly show that the heterogeneity of many digital natural and cultural heritage collections is an ideal playground for researchers to develop, test, and scale up new assemblages of computational technologies.

In their tutorial, Rinaldo<sup>†</sup> et al. take up that cue and explain how the rise of computational technologies has shaped knowledge extraction in the digital natural heritage domain in the U.S. since the early 2000s. The tutorial shows that archives in the natural heritage domain are often much less structured than in the cultural heritage domain. Next to fragmentation across institutions, the exploitation and establishment of links among libraries, archives, and specimen collections is a labor-intensive endeavour. By examining different crowdsourcing solutions as they have been used in the natural heritage domain, the tutorial highlights the role volunteers played in adding contextual knowledge, annotations, and links to other resources in the field. Instead of taking cutting-edge computational methods as a point of departure, the tutorial shows the value of modelling computational solutions according to collection-related human expertise and exploration strategies.

The special issue is concluded by articles by Chandrasekar et al. (fourth article of this special issue) and Viola (fifth article of this special issue). Both articles highlight the importance of visual aspects of natural and cultural heritage collections and their computational processing and enrichment. While Chandrasekar and colleagues focus on the (semi-)automated localization and interpretation of plants and other objects on digitized herbarium sheets, Viola problematizes the role of users in the creation of visualizations based on digital cultural heritage data. Instead of considering users as consumers of visualizations, she proposes a post-authentic framework in which visualizations are considered as part of a much longer, cyclical, and open-ended process of curation in which various stakeholders are involved. Although both articles take a different approach to “the visual,” they highlight the importance of humans in enriching natural and cultural heritage and related visual representations. Without careful alignment of humans and machines, automatically created interpretations of visual cultural and natural heritage will remain problematic and deficient. However, both articles also show how fast the fields of computer vision and visualization studies, as part of the broader field of digital humanities, have advanced over the last years.

As a way of concluding this editorial, we thank several individuals and institutions. First, we thank the Dutch Research Council (NWO) for financing the conference “Collect and Connect.” The conference marked the end of a 4-year research program titled “Making Sense of Illustrated Handwritten Archives” (2016–2020, grant number 652.001.001) financed by Brill publishers and the Creative Industries program of NWO. Second, we thank the Naturalis Biodiversity Center in Leiden, who provided the conference’s organizers with a base for two intensive conference days. Third, we thank Dr. Martha Fleming for organizing a fantastic round-table discussion that brought key issues of the conference together. Fourth, we thank Dr. Lise Stork, professor Fons Verbeek, and professor Jaap van den Herik for helping us to realize the “Collect and Connect” conference during which the seeds for this special issue were planted. Finally, we thank professor Franco Niccolucci and Dr. Karina Rodriguez-Echavarría, who offered us the opportunity to publish the conference results as a *JOCCH* special issue. Many thanks for your valuable support and advice along the way.

## REFERENCES

- [1] P. Brandon, J. Hedrick, Mason Heberling, Emily K. Meineke, Kathryn G. Turner, Christopher J. Grassa, Daniel S. Park, Jonathan Kennedy, Julia A. Clark, Joseph A. Cook, David C. Blackburn, Scott V. Edwards, and Charles C. Davis. 2020. Digitization and the future of natural history collections. *Bioscience* 70, 3 (2020), 243–251.
- [2] Andreas Weber. 2021. Natural history collections and empire. In *Routledge Handbook of Science and Empire*, Andrew Goss (Ed.). Routledge, London, 80–86. <https://doi.org/10.4324/9780429273360-8>
- [3] Gerben Zaagsma. 2022. Digital history and the politics of digitization. In *Digital Scholarship in the Humanities* (2022). <https://doi.org/10.1093/llc/fqac050>
- [4] Andreas Weber, Maarten Heerlien, Eulalia Gassó Miracle, and Katherine Wolstencroft. 2020. *Proceedings of the International Conference Collect and Connect: Archives and Collections in a Digital Age*. 23–24.
- [5] Melissa Terras. 2022. Inviting AI into the archives: The reception of handwritten recognition technology into historical manuscript tradition. In *Archives, Access and Artificial Intelligence*, Lise Jaillant (Ed.). Bielefeld University Press, Bielefeld. <https://doi.org/10.1515/9783839455845-008>
- [6] Giovanni Colavizza, Tobias Blanke, Charles Jeurgens, and Julia Noordegraaf. 2022. Archives and AI. *J. Comput. Cult. Herit.* 15 (1), 4 (2022), 1–15. <https://doi.org/10.1145/3479010>
- [7] Elisabetta Cilli, Francesco Fontani, Marta Maria Ciucani, Marcella Pizzuto, Pierangelo Di Benedetto, Sara De Fanti, Thomas Mignani, Carla Bini, Rocco Lacovera, Susi Pelotti, Filippo Spadola, Donato Luiselli and Sabrina Lo Brutto. 2023. Museomics provides insights into conservation and education: The instance of an african lion specimen from the museum of zoology “pietro doederlein”. *Diversity* 15, 1 (2023), 87. <https://doi.org/10.3390/d15010087>
- [8] Lise Stork, Andreas Weber, Jaap van den Herik, Fons Verbeek, Aske Plaat, and Katherine Wolstencroft. 2021. Large-scale zero-shot learning in the wild: Classifying zoological illustrations. *Ecol. Inf.* 62 (2021), 101222. <https://doi.org/10.1016/j.ecoinf.2021.101222>
- [9] Andreas Weber, Mahya Ameryan, Katherine Wolstencroft, Lise Stork, Maarten Heerlien, and Lambert Schomaker. 2018. Towards a digital infrastructure for illustrated handwritten archives. In *Lecture Notes in Computer Science (LNISA)*, vol. 10605 (2018), 155–166. [10.1007/978-3-319-75826-8\\_13](https://doi.org/10.1007/978-3-319-75826-8_13)
- [10] Lambert Schomaker. 2020. Lifelong learning for text retrieval and recognition in historical handwritten document analysis. In *Recognition and Retrieval—State of the Art and Future Trends*, World Scientific, Singapore, 221–248. [https://doi.org/10.1142/9789811203244\\_0012](https://doi.org/10.1142/9789811203244_0012)
- [11] Marco Fiorucci, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Travigila, Alessio Del Bue, and Stuart James. 2020. Machine learning for cultural heritage: A survey. *Pattern Recogn. Lett.* 133 (2020), 102–108. <https://doi.org/10.1016/j.patrec.2020.02.017>
- [12] Alex Hardisty, Elizabeth R. Ellwood, Gil Nelson, Breda Zimkus, Jutta Buschbom, Wouter Addink, Richard K. Rabeler, John Bates, Andrew Bentley, José A. B. Fortes, Sara Hansen, James A. Macklin, Austin R. Mast, Joseph T. Miller, Anna K. Monfils, Deborah L. Paul, Elycia Wallis, and Michael Webster. 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the internet. *BioScience* 72, 10 (2022), 978–987. <https://doi.org/10.1093/biosci/biac060>
- [13] Karl-Heinz Lampe, Klaus Riede, and Martin Doerr. 2008. Research between natural and cultural history information: Benefits and IT-requirements for transdisciplinarity. *J. Comput. Cult. Herit.* 1, 1 (2008), 1–22. <https://doi.org/10.1145/1367080.1367084>
- [14] Lise Stork. 2021. *Knowledge Extraction from Archives of Natural History Collections*. Ph.D. Dissertation, Leiden University.
- [15] Sharif Islam, Andreas Weber, and Erzsébet Tóth-Czifra. 2022. From green deal to cultural heritage: FAIR digital objects and european common data spaces. *Res. Ideas Outcomes* 8 (2022), e93815. <https://doi.org/10.3897/rio.8.e93815>