

Modelling Research Expeditions in Wikidata: Best Practice for Standardisation and Contextualisation

Sabine von Mering[‡], Paul Jean-Charles Braun[§], Robert W. N. Cubey^l, Quentin Groom[¶], Elspeth M Haston^l, Annika Hendriksen[#], Rukaya Johaadien[▫], Siobhan Leachman[«], Luke Marsden[»],[^], Heimo Rainer[˘], Joaquim Santos^l, Dag Endresen[▫][?]

[‡] Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

[§] Musée national d'histoire naturelle Luxembourg, Luxembourg, Luxembourg

^l Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

[¶] Meise Botanic Garden, Meise, Belgium

[#] Naturalis Biodiversity Center, Leiden, Netherlands

[▫] Natural History Museum, University of Oslo, Oslo, Norway

[«] Independent researcher, Wellington, New Zealand

[»] University Centre in Svalbard, Longyearbyen, Norway

[^] Norwegian Meteorological Institute, Oslo, Norway

[˘] Natural History Museum Vienna, Vienna, Austria

^l Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal

[?] GBIF Norway, Oslo, Norway

Corresponding author: Sabine von Mering (s.vonmering@gmail.com)

Abstract

Expeditions and other collecting events are a major source of objects in natural history museums (e.g., Mesibov 2021). Historically, these trips were often transdisciplinary: biological and Earth science specimens were collected at the same time as ethnological or anthropological objects. As a result, specimens and other material gathered during the same expedition, as well as the related data and metadata, are often distributed across multiple institutions. Many expeditions were driven by colonial agendas, aiming to discover new resources to exploit, and their findings were seldom shared with the source countries and local people. Understanding these expeditions illuminates the colonial origins of museum collections, and contributes to recognizing and addressing their impacts (e.g., Das and Lowe 2018, Ashby and Machin 2021).

Research expeditions continue to contribute to natural history collections. There is a need to link historical or contemporary research expeditions to other entities, requiring the unambiguous labelling (and persistent identifiers) of such events. Stable identifiers for expeditions plus the sharing of metadata and descriptions in a wide range of languages will facilitate access to scattered information about the event, the institutions housing specimens and objects, the participants and the locations visited, and assist with the linking of distributed material and related research data. However, structured data for scientific expeditions are currently lacking. While identifier systems have been created for many entities over the last few decades, there is no dedicated identifier for research

expeditions and similar events. Several studies have shown the importance of people identifiers for linking collection data (e.g., Groom et al. 2022), and we argue the same is true for expeditions.

[Wikidata](#) is a multilingual community-curated knowledge base containing data structured in a human- and machine-readable format. It allows easy creation, updating and enriching of items on expeditions, and provides stable identifiers for them that can be used in collection management systems. Expeditions can be linked to participants and other agents, regions, localities, objects, archival material, maps, publications, field notebooks, documentary footage and art works resulting from the expeditions, thus making historical information more easily accessible and assisting with the acknowledgment of any imperial or colonial impact that may have resulted from the expedition. Expeditions in Wikidata can be hierarchical, e.g., linking a series of related events or under an umbrella project together providing a machine-readable way to harvest all project data. Wikidata also can provide links between present day countries and historical names for locations (e.g., former colonial names). Expeditions published as Linked Open Data make datasets more [FAIR](#) (Findable, Accessible, Interoperable, Reusable), and are also useful in data transcription and validation processes. Visualisation of itinerary data and travel routes also facilitate data quality checks.

An informal working group of people interested in the topic was formed to discuss standards and share best practices and recommendations regarding terminology, data modelling and contextualisation. Building upon previous work (e.g., Bauer et al. 2022, von Mering et al. 2022, Leachman 2023), we aim to work towards the enrichment, linking and standardisation of data about research expeditions. If the Wikidata identifiers of these expeditions and participants are added to the records of the corresponding entities in the collection management system, institutions can link from their own collection metadata to the relations made in Wikidata, including to collections in other institutions. The participants of the expedition can be further linked to specimens gathered during the expedition with the use of tools, such as [Bionomia](#), which can facilitate data round-tripping between these collections and specimen records, the Global Biodiversity Information Facility ([GBIF](#)) and Wikidata (Shorthouse 2020). Other initiatives such as the Distributed System of Scientific Collections ([DiSSCo](#)) are also interested in incorporating these identifiers as links and annotations.

Keywords

collecting events, collection management, Linked Open Data (LOD), natural history/science collections, natural history expeditions, provenance research, recommended properties

Presenting author

Dag Endresen

Presented at

TDWG 2023

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Ashby J, Machin R (2021) Legacies of colonial violence in natural history collections. *Journal of Natural Science Collections* 8: 44-54. URL: <http://www.natsca.org/article/2631>
- Bauer J, Burkhalter R, Karim T, Krimmel E, Landis M, Leachman S, Little H, Lorente M, Mills S, Neu-Yagle N, Norton B, Paul D, Shorthouse D, Utrup J, Van Veldhuizen J, Walker L (2022) Guidelines for Using Wikidata to Mobilize Information about People in Collections: A Paleontology Perspective. Zenodo. <https://doi.org/10.5281/zenodo.6977243>
- Das S, Lowe M (2018) Nature Read in Black and White: decolonial approaches to interpreting natural history collections. *Journal of Natural Science Collections* 6: 4-8208. URL: <http://www.natsca.org/article/2509>
- Groom Q, Bräuchler C, Cubey R, Dillen M, Huybrechts P, Kearney N, Klazenga N, Leachman S, Paul DL, Rogers H, Santos J, Shorthouse D, Vaughan A, von Mering S, Haston E (2022) The disambiguation of people names in biological collections. *Biodiversity Data Journal* 10: e86089. <https://doi.org/10.3897/bdj.10.e86089>
- Leachman S (2023) Schemas for Wikidata items for scientific collectors and research expeditions (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.8088216>
- Mesibov R (2021) An Australian collector's authority file, 1973–2020. *Biodiversity Data Journal* 9: e70463. <https://doi.org/10.3897/bdj.9.e70463>
- Shorthouse D (2020) Slingshot With Four Giants on a Quest to Credit Natural Historians for our Museums and Collections. *Biodiversity Information Science and Standards* 4: e59167. <https://doi.org/10.3897/biss.4.59167>
- von Mering S, Kaiser K, Petersen M (2022) Transforming Closed Silos into Shared Resources: Opening up data on historical collection agents affiliated with the Museum für Naturkunde Berlin. *Biodiversity Information Science and Standards* 6: e93787. <https://doi.org/10.3897/biss.6.93787>