

Sharing data, caring for collections. Open data on collection agents affiliated with the *Museum für Naturkunde Berlin*

Sabine von Mering[‡], Erik Stolze[‡], Katja Kaiser[‡], Mareike Petersen[‡]

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

Corresponding author: Sabine von Mering (sabine.vonmering@mfn.berlin)

Academic editor: Editorial Secretary

Abstract

Linked open data on collection agents contribute to increased discoverability, accessibility and transparency of natural history collections. Despite major efforts to digitise and open up museum and university object collections, related information is often stored in internal resources. This paper describes a project conducted at the Museum für Naturkunde Berlin (MfN) contributing to its collection disclosure and development initiatives. Information on historical collectors and other collection agents was transferred from the internal MfN collector wiki to Wikidata. For a total of 600 collection agents, existing Wikidata items were enriched or new items created.

Special emphasis was put on linking these people to the Museum, to document their affiliation with the MfN, its collection and its archive. Within the project, an open participatory approach was taken. Several Wikidata edit-a-thons were organised to test this collaborative and innovative format for possible future application by the Museum. By opening up institutional silos and openly sharing data on agents connected to museum holdings, these data become more widely accessible and reusable, for example, as a resource for transdisciplinary provenance research.

Keywords

citizen/community science, collectors, collection history, collection management, colonial history, edit-a-thons, identifier, institutional history, Linked Open Data (LOD), natural history collections, natural science collections, people data, provenance research, Wikidata.

Introduction and background

Through extensive digitisation initiatives, natural history museums, herbaria and other natural science collections around the world are improving access to their collections, thus mobilising huge amounts of data. This has led to new challenges for research data management, particularly to enable interoperability between the different data sources and to guarantee uniqueness of relevant entities such as collection objects, taxa, localities and people.

Research on natural history collections – especially from a biological perspective – is typically object-centred, for example, with a clear focus on (type) specimens of certain taxonomic groups. However, people play a central role in collections as they are the ones gathering the objects in the first place and are also a main starting point for provenance research. In fact, collection agents are involved in all steps of the research process (Fig. 1) – from the planning to the actual gathering event ('collect' and 'create'), further to processing, analysing ('research') and preserving specimens ('curate') and other objects, to finally sharing the results ('collaborate' and 'publish') and allowing the reuse of data. As stable and relatively well-known entities in the biodiversity knowledge graph, people are essential for making links between entities, such as specimens and other collection objects, taxon names, publications, other people, collections or institutions (e.g. Page 2016, Groom et al. 2020, Page 2022, Groom et al. 2022, Page 2023) as well as research expeditions (von Mering et al. 2023). In addition to collectors, collection agents comprise suppliers of specimens, such as dealers of naturalia, owners of menageries and zoological gardens as well as taxidermists, preparators, photographers and scientific illustrators. It is important to note that the contribution of local collectors, preparators, informants and guides often remains hidden, due to a lack of information in the historical records.

Information on historical (and recent) collectors and other collection agents increases the discoverability and accessibility of natural history specimens and other objects, as well as adding transparency on the collections and their provenance. People can provide a different way to access information about biodiversity and related historical information. Well-curated and trustworthy biographical information about collectors and itineraries of their expeditions and collecting trips also help to improve metadata from incomplete label information. Increased transparency and accessibility of collections data facilitates scientific research from a wide range of fields including provenance research and critical reflection on collections from colonial contexts. Furthermore, such data also provide opportunities for application by artists, creative industry, education and co-creation with stakeholders from all parts of society.

The Museum für Naturkunde Berlin (MfN) houses one of the largest natural history collections in Europe comprising approximately 30 million zoological, palaeontological, geological and mineralogical objects. As a result of scientific processing, description and intensive research work over several centuries, the objects at the Museum have acquired

significant historical, scientific and cultural value. This entails a huge amount of responsibility – for the safeguarding of the collection and for making it accessible to the scientific community and the wider public. Within the framework of the so-called *Zukunftssplan* ([Future Plan](#)), the MfN collection will be digitised and opened up to make it accessible for all, i.e. to create an open global knowledge infrastructure (Hoffmann et al. 2022a, Hoffmann et al. 2022b, Lentge-Maaß and Berger 2022).

The [archive of the MfN](#) holds its Historical Image and Document Collections, i.e. various kinds of documents such as correspondence, personnel files of MfN employees, expedition reports, field books and diaries, complemented by an extensive collection of historical photographs and portraits (e.g. Reimers 2022). As part of the central research infrastructure, the archive provides (in many cases, the only) evidence for the provenance and scientific examination of natural history objects. It is now increasingly becoming accessible and usable for the general public.

The MfN holds considerable collections from the former German colonial regions. Following the Federal Council decision (*Bundesratsbeschluss*) from 1889 (Centralblatt 1890), the MfN was – together with the Ethnological and Botanical Museum Berlin – the primary recipient of objects from expeditions funded by the German Reich, as well as materials collected by colonial officials. As a result, the collections of the Museum grew to an unprecedented extent. The MfN and, especially, the centre for the *Humanities of Nature*, have conducted research focusing on the political and, especially, the colonial history of the institution and its holdings (e.g. Heumann et al. 2018, Kaiser et al. 2023). Currently, different [projects](#) examine colonial provenances and try to establish a more systematic approach to deal with natural history objects from colonial contexts that pay respect to the multilayered meanings of these holdings. The project described in this publication is part of these efforts and faces up to the broad academic, political and public debate about the re-appraisal of the colonial past and the need for more transparency regarding the colonial holdings of museums (DMB 2021). Since the 1970s, actors in the countries of origin, as well as civil society initiatives and representatives of the diaspora, have been demanding transparency with regard to collections acquired during the colonial era (Savoy 2022). On a governmental level, in 2019, the cultural sector in Germany agreed on “Framework Principles for Dealing with Collections from Colonial Contexts” which state that more transparency and documentation are needed regarding objects from colonial contexts. These framework principles further demand museums to increase provenance research on collections from colonial contexts and cooperations with countries of origin (Framework Principles 2019). The principles are complemented by the federal “3-road strategy”, meaning Access – Transparency – Cooperation as three parallel roads to be taken and aiming at make information on collections from colonial contexts held in Germany accessible in a central data repository. Therefore, access to digital information on collections from the colonial context has become a national priority in Germany. Even though natural history museums are not the central focus of the current political and public debates on the colonial heritage of museums, they cannot be separated from the interdisciplinary and cumulative practices of colonial collecting. Their collections and archives provide essential information on how natural history collections

expanded massively during the period of colonial expansion. Recently, the role of the natural sciences and their institutions for colonial rule has been scrutinised in various studies (Alberti 2018, Ashby and Machin 2022, Cisneros et al. 2022, Gladstone and Pearl 2022).

In the past, many collectors were active in several fields and acquired diverse material, for example, zoological and botanical as well as palaeontological or mineralogical objects, but often also ethnological or anthropological artefacts. This transdisciplinary collecting practice, in addition to the tradition of exchanging duplicates between institutions (e.g. Kaiser 2022), resulted in a large or at least partial overlap of people affiliated with different (international) collections. While an increasing number of institutions are providing access to internal information, such as accession books and lists of collectors (e.g. Döring and Dressler 2018, SMB 2023), a concerted and collaborative effort could speed up the disambiguation of collection agents. Larger institutions with archives could act as role models and support other institutions by opening up information about agents from their collection management systems and connected information from archival systems. Progress made during collection disclosure depends not only on the size of the institution and its collection, but also on standardisation and automation of related digitisation processes and workflows to openly share information. Nevertheless, collections of all sizes and fields may hold crucial “puzzle pieces” to disambiguate specific collectors, for example, through annotated duplicate specimens, institutional archival records or information from connected local or regional natural history societies in which these collectors were members. Therefore, sharing information and making it openly available for everyone could turn out to be beneficial for many institutions. Opening up available data will help to improve data quality in collection databases, support the disambiguation of collection agents (Groom et al. 2020, Groom et al. 2022, von Mering et al. 2022, Meeus et al. 2023) and allow cross-disciplinary reuse of the data. Once extensive collection data and linked data about archival material become openly available, the whole process of collection disclosure and data integration is accelerated. Required is a central, openly accessible platform that enables collaborative editing and sharing of interoperable data.

[Wikidata](#) has developed into a widely used system that could serve this purpose. It is a free and open knowledge base that can be edited by anyone in the world. Wikidata stores structured data that can be read and edited by humans and machines alike (e.g. Vrandečić and Krötzsch 2014, Waagmeester et al. 2020, Shafee et al. 2023). Since Wikidata serves as a hub for other identifiers, many Wikidata items for people provide dozens of identifiers from digital libraries, specialised databases etc. As a collaborative and multilingual platform, the already large and active user community of Wikidata is growing steadily. One prominent example of cultural heritage or GLAM institutions (GLAM stands for galleries, libraries, archives and museums) using Wikidata include the collaboration between the Swedish Nationalmuseum and the National Historical Museums with Wikimedia Sverige (Fagervig 2023).

At the MfN, a number of past projects used internal wikis to collaboratively collect information and to make it available, as well as to test new methods in knowledge

transfer. Within the project "WIKI-Ansatz und kollaboratives Arbeiten im Forschungsmuseum" from 2013 until 2016 (Patzschke et al. 2016), project staff and other MfN colleagues interested in the topic collected, added and edited information on collectors and other collection agents into a wiki, the so-called "MfN [Sammler-Wiki](#)" (hence MfN collector wiki, Fig. 2). This wiki was developed to serve as a support tool for collection staff, facilitating investigation on collectors and other collection agents and other tasks linked to collection management. The information is only available internally, it is only in German, not fully structured and lacks updates since 2017. To improve the accessibility of these data, a new project was conceived and approved for internal funding to address these issues and to enhance global access to this relevant and valuable information.

The main objectives of the project were:

1. to transfer the information from the MfN collector wiki to Wikidata, thus making the data openly available and more accessible,
2. to enrich the dataset from the wiki, to analyse the included collection agents and to identify gaps and bias and
3. to test a new interactive and collaborative format (so-called edit-a-thons) in the Museum with professionals and interested volunteers or citizen scientists.

Following the motto of the original wiki project, the project aspires to "bring together, complement and share the knowledge of many" on collectors and other collection agents (see Fig. 2 with German text "Das Wissen der Vielen zusammenführen, ergänzen und teilen").

The process of disambiguating collection agents requires expertise, time and care (e.g. Groom et al. 2022). Open and participatory approaches can help to involve people with knowledge in different fields. Stable identifiers (persistent identifiers, PIDs) allow us to unambiguously identify and name entities, such as people, even if the knowledge about them may be fragmentary. By opening up, sharing and exchanging information about collection agents, the data can be complemented, corrected, enriched and connected as Linked Open Data (LOD) in the semantic web (Fig. 1).

Material and methods

Data source and other resources

Data stored in the MfN collector wiki, which contains information on different collection agents and a number of significant research expeditions, were the basis for this project. In addition to basic biographical information, the pages for every person contain information about collecting areas and itineraries, information on objects in specific sub-collections of the Museum, archival material in the MfN archive and sometimes important publications, photographs or references. When this new "collector" project started in July

2022, a complete dataset was exported from the internal Semantic Media wiki instance (Fig. 2).

By creating new or updating and enriching existing Wikidata items, it was ensured that Wikidata identifiers exist for most of the collectors or other collection agents mentioned in the MfN collector wiki. For many of the people in the internal wiki, the GND number, an identifier from the *Gemeinsame Normdatei* or GND (Integrated Authority File) of the German National Library (German: *Deutsche Nationalbibliothek*, DNB), was recorded. The GND identifier is widely used by libraries and increasingly by museums and archives in German-speaking countries to document and to catalogue their holdings. At the MfN, the GND number is stored in several internal systems (library system, archival system, collection management system). For all people from the wiki dataset, the GND numbers were checked, missing ones added and some updated where necessary (older GND numbers were deprecated).

Information from the MfN collection databases and the MfN archive plus a number of other resources were checked to verify, enrich and, if necessary, amend data from the MfN collector wiki. Important sources included the Biodiversity Heritage Library (BHL), and specialised databases, such as “Biographies of the Entomologists of the World” (Groll 2017). A collection of publications on the history of the MfN and its collection served as additional information sources (e.g. Hackethal 1985, Hoppe 2003, Damaschun et al. 2010).

For all collection agents, basic biographical information was verified and corrected or complemented where necessary. At the beginning of the project, a set of central Wikidata properties was agreed on and were used for the complete dataset if applicable. In addition, a number of secondary properties were used frequently for many, but not all records. The focus was on the following three Wikidata properties:

1. *employer* ([P108](#)),
2. *archives at* ([P485](#)) and
3. *collection items at* ([P11146](#)).

The last property was newly created within the project duration following community discussion and a [property proposal](#) by SvM. In a few cases, the property *affiliation* ([P1416](#)) was used if there existed no employment at the MfN, but another kind of relationship without a formal contract, for example, as a doctoral student, a freelancer, a longtime volunteer or if significant collections made by a person (specimens or other objects) are held at the Museum. Another central property for the project was *occupation* ([P106](#)); some information on occupations was stored in the MfN collector wiki, but this information was considerably enriched during the project and specified for most collection agents, using both data from the wiki, but also from further sources. Additional Wikidata properties used for a part of the dataset were the following: *work location* ([P937](#)), *participant in* ([P1344](#), for example, for research expeditions), *educated at* ([P69](#)), *doctoral advisor* ([P184](#)), *doctoral student* ([P185](#)), *academic degree* ([P512](#)), significant person ([P3342](#)) with the qualifier *object has role* ([P3831](#), for example, *correspondent*, *co-collector*,

co-author, colleague, friend). The Wikidata properties used by the project are also documented in the [WikiProject](#). In addition, [Bionomia](#) profiles were newly created or enriched by attributing additional specimens for all agents that had collected or identified specimens. Bionomia links natural history specimens to collectors; it uses specimen data that are already available via the Global Biodiversity Information Facility ([GBIF](#); Shorthouse 2020a, Shorthouse 2020b). New Bionomia identifiers ([P6944](#)) were added to the respective Wikidata items.

Project communication and outreach

To promote the project and its events, a number of pages were created on the MfN institutional website. Additionally, a [WikiProject](#) was started to reach more people within the Wikidata community, to document progress, approaches, properties, queries and sources used, as well as events. A number of mailing lists (e.g. Sammlungsnetzwerk, Netzwerk Koloniale Kontexte), the [Wikidata 10th birthday Events calendar](#), direct emails to national and international colleagues and posts in social media channels (especially X, formerly Twitter) were used to provide information about the project and to invite prospective participants to the edit-a-thons.

During and after the second edit-a-thon, participants communicated via a dedicated [Slack](#) channel. Later, communication was moved to another already existing Slack channel (used by Bionomia scribes) to exchange information, ask questions or share best practice, workflows or experiences. To track progress made during the open Wikidata edit-a-thon, a Wikimedia dashboard was created by one participant after the introductory session for beginners. Project outreach included several presentations to different audiences, for example, at a meeting of [Salon KOSMOS](#) in Berlin, a call of the international [LD4 Wikidata Affinity Group](#) and at the [TDWG 2022](#) conference in Sofia (von Mering et al. 2022). Another publication about the project is forthcoming (Kaiser and von Mering, in press).

Edit-a-thons

Within the project, an open participatory approach was used to involve MfN staff members, but also other user groups, such as researchers worldwide, volunteers and citizen scientists interested in the collection or history of the Museum. The project concept included a number of workshops for editing Wikidata, known as edit-a-thons, similar to hackathons for coding or transcribathons for transcribing historical texts, respectively (Fig. 3). Edit-a-thons or similar event formats have successfully been applied in different contexts, for example, in the [Paleo Data Working Group](#) (Little et al. 2022, Bauer et al. 2022) or in organisations such as the [Smithsonian institution](#).

The edit-a-thons aimed:

1. to educate MfN staff, colleagues from other GLAM institutions (galleries, libraries, archives, museums), volunteers and citizen scientists about the potential of Wikidata,

2. to train them how to edit and create Wikidata items, reference statements and link information,
3. to test the format of edit-a-thons for future applications at the MfN.

The edit-a-thons were held with larger vs. smaller numbers of participants, only internal participants from the MfN vs. mostly external participants, in person or hybrid vs. fully virtual events, as a two-day event (2 x 3 hours plus pre-workshop Wikidata introduction for beginners) vs. a short version (1 x 2 hours).

A first in-house edit-a-thon was held on 29 and 30 September 2022. To involve staff members from different departments and with different backgrounds, specific colleagues were actively invited to join the workshop. Further colleagues registered after a general invitation was sent to all MfN employees, resulting in a total number of 19 registered participants, of which 13 actually participated in the event. Only three attendees had their own Wikidata account prior to the event, while the majority of the participants were absolutely new to the topic. A second open Wikidata edit-a-thon was organised on 17 and 18 November 2022, in close collaboration with the centre for the *Humanities of Nature*. The focus of this event was on colonial collections from the former German colony Kamerun (today Cameroon). Within the workshop, a list of people who were active in Kamerun and linked to colonial collections in the MfN was used as a starting point for editing and discussions. This list included people from the collector wiki as well as additional agents.

The workshop was open to anyone interested in the topic and held as a virtual event to enable participation from anywhere in the world. A total of 38 people from nine different countries eventually participated in the workshop; they had advanced knowledge in at least ten languages. Many participants of the workshop were staff members of other GLAM institutions including smaller museums with a regional focus and from Berlin universities. In addition, several international Wikimedians attended, including a number of participants from Cameroon and Nigeria. While most participants were experienced in editing wikis in general and Wikipedia in particular, only eight of them had previous experience with Wikidata. Therefore, an introductory and training session was organised prior to the edit-a-thon. It was attended by 13 people who learned the basics of Wikidata editing.

Results

Data transfer

The main result of the described project is an open-linked dataset comprising the information from the MfN collector wiki and data enrichments (von Mering and Stolze 2023). All internal data were cleaned, enriched and further linked to provide better access to this information on collectors and other collection agents. The original data export from the internal wiki comprised a total of 609 data records. Data cleaning

involved removal of a few duplicates and obsolete test entries, the remaining records being cross-checked for existing Wikidata items. After data cleaning and deduplication, the dataset comprised 596 distinct collection agents (Table 1). For 78.4% (467) of these distinct people, Wikidata items existed already. These items were verified and subsequently enriched with information from the collector wiki and other sources. Of the remaining entries, 72 people (12.1%) were disambiguated within the project and new Wikidata items created for them, while available information was insufficient for the disambiguation of 57 further agents (9.6%). The latter group requires further study, including targeted investigations and search for archival records or non-digitised sources. In total, 539 Wikidata items for all disambiguated collection agents from the original dataset were newly created and/or enriched. The dataset is deposited as a static snapshot on Zenodo (von Mering and Stolze 2023), but a dynamic dataset, regularly updated and further expanded, is accessible in Wikidata via direct search or SPARQL queries (for example queries, see the [WikiProject](#)).

Data enrichment and analysis

Some effort was made to increase accessibility in Wikidata and beyond. For example, the *labels* in Wikidata (i.e. the names of the people) were added in four languages (English, French, German, Spanish) and the *descriptions* of the Wikidata items provided or updated at least in English and German, often also in French and Spanish.

A focus in the project was to collect data for three central properties, i.e. *employer* ([P108](#)) to link to the MfN (and other institutions), *archives at* ([P485](#)) to state that records are held at the MfN archive and *collection items at* ([P11146](#)) to state that specimens or objects gathered by these people are housed in the MfN. Table 1 summarises information on these and additional statistics related to the dataset. Less than a quarter of the collection agents were employed by or directly affiliated with the MfN. For more than 40% of the people, it is currently known that specimens or other collection items are housed in the MfN (this number is potentially significantly higher). Only eleven collection agents in the dataset were female (1.8%) and only one name (0.2%) was identified as an indigenous collector, namely [Thomas David Aubinn](#).

By using the property *archives at* ([P485](#)) and linking to the newly-created item for the MfN archive ([Q113678597](#)), it was stated in Wikidata that the MfN archive is holding records connected to a person from the original dataset. The inventory number (German: *Signatur*) was added using the property inventory number ([P217](#)) as a qualifier. Recently, digital object identifiers (DOIs) were assigned to the finding aids (in German: *Findbücher*) of larger estates (in German: *Nachlässe*). These DOIs were also referenced in Wikidata, using the qualifier *described at URL* ([P973](#)). At the end of the project, the data showed that, for 55% of the collection agents in the dataset (294), archival records exist in the MfN archive and this was stated in Wikidata. For about 84% of the collection agents (452) exists a GND number and for 56% (304), a Bionomia profile. Of the 72 Wikidata items on collection agents newly created within the project, 83.3% (60) were enriched by other Wikidata users.

The dataset comprises collection agents with a variety of specialisations and occupations. More than half of the people (55.9%, 337) were active in the field of zoology, while a total of 58.8% (317) are collectors of natural history specimens or ethnographic objects. An overview of the different occupations and the respective numbers are given in Table 2 and Fig. 4 .

Edit-a-thons

From the experience gained during the limited number of editing events, the format allowed the organisers to quickly train diverse groups of people and to empower them to use, query and edit Wikidata. A considerable amount of content can be created during well-prepared edit-a-thons. One difficulty is, however, that participants often have distinctly different levels of previous knowledge about Wikidata, ranging from complete beginners to well advanced users. At the end of the first edit-a-thon, participants asked for a decoupling of the technical training part at the beginning from the actual editing event, in order to focus more on the creation of content. As a result, the second edit-a-thon was preceded by an introductory workshop held a few days before the event to provide basic training in editing Wikidata for new users of the platform.

The first edit-a-thons had a focus on training staff members working in different departments of the Museum, therefore only creating and enriching a limited number of Wikidata items. In contrast, the second edit-a-thon focused on creating content. By working through a prepared worksheet and table, participants generated and enriched Wikidata items.

In addition to creating and enriching items on collection agents, a number of Wikidata items for relevant publications (used to reference certain statements) were newly created, as well as other items for related entities (linked to the collection agents, for example, colonial outposts). These were at least partially created, based on internal (even handwritten) sources, such as finding aids (*Findbücher*), entry books (*Eingangsbücher*) and catalogues. These items are now easily findable by querying Wikidata and are sometimes directly accessible via DOI links. During the two-day workshop, 13 active editors had added themselves to the workshop dashboard and their editing resulted in a total of 147 newly-created Wikidata items (not only for collection agents, but also for other entities related to people), 116 descriptions added and 94 changed, 133 Aliases added and seven changed, as well as 500+ references added. For more details, see the [dashboard](#) of the edit-a-thon.

Bringing together people from different backgrounds and with a wide range of previous knowledge and different insights was conceived to be particularly beneficial for capturing information on all aspects of the life and work of a person and, especially, their contribution to science. By researching, discovering and recording relationships with other people (co-collector, co-author, correspondent, friend, travel companion etc.), unknown additional sources of information, such as correspondence or other archival material could be identified and located.

Concepts and terminologies were discussed, including terms related to “collecting” in the context of violence and unethical collecting practices in former colonies or terms used in a specific historical context (e.g. locality names used by colonial powers) or titles of publications that include, for example, racist language. Another discussion centred on how to avoid re-creating and highlighting “colonial networks” by mostly adding data on agents active in former German colonies, but to enrich these data with information requested by and relevant for communities of origin. Participants also stressed the importance of engagement and exchange with communities of origin to learn about their needs and requirements linked to data accessibility and to enrich data accordingly. Together with the participants from Cameroon and neighbouring countries, an attempt was made to consider non-European perspectives and to record contributions by local people in the former colonies.

Due to internal demand from MfN colleagues, smaller versions of training and editing sessions (so-called “Mini Edit-a-thons”) were organised. These helped to develop more routine in editing Wikidata and overcome remaining obstacles or barriers.

Discussion

To put the natural history collections at best use for scientific studies on – for example – biodiversity (loss), climate change, provenance research and collection practices, it is important to know:

1. *what* is in them (which taxa, which bio- and geodiversity it holds),
2. *where* does the material come from (geography) and
3. *who* is linked to it (which collection agents or other actors).

The huge task of disambiguating people in natural history and other collections can be accelerated by a collaborative effort of sharing internal institutional information and opening up closed silos. Wikidata provides a valuable discovery tool or “finding aid” for improved access to cultural heritage data and supports better data linkage. The MfN supports the open data movement and the use of open licences. For media and data from the digitisation and collection disclosure process, CC0 is the default licence used (MfN 2022).

Data transfer

The transferred dataset comprising about 600 collection agents affiliated with or linked to the Museum für Naturkunde Berlin is important for collection disclosure and further data integration. In December 2023, i.e. about a year after the end of the project, 640 distinct collection agents are linked to the MfN. This means that 100 people in addition to the 539 from the MfN collector wiki are connected to the Museum and its collection. However, these data are only the “tip of the iceberg” in comparison to the total number of agents and other entities connected to the large collection of approximately 30 million objects; thus, more data collection and research are needed.

Data enrichment and analysis

The statistics provided for the dataset transferred to Wikidata show that there was some kind of imbalance when the collection agents were “selected” and added during the previous project creating the internal wiki. Collectors were not added systematically for certain subcollections or fields of work, but upon availability (e.g. from accession books) or related to the interest of the participating editors. Therefore, the data reflect some biases and have gaps in certain areas and further editing, enriching and creating open data are needed.

From a curatorial perspective, Wikidata is an external tool supporting internal data management. It facilitates data quality checks and, if identifiers are integrated into collection management systems, community-curated data can be used to verify information in the database.

Several websites and tools such as [Scholia](#) or Bionomia use data from Wikidata, connect it with other data and visualise it. Attributions made in Bionomia link natural history specimens to the respective collectors and/or determiners; these can be downloaded, for example, as a Frictionless Data package, thus allowing for roundtripping of enriched data into institutional databases (<https://en.bionomia.net/collection-data-managers>). This information is stored in Darwin Core properties recordedByID and identifiedByID.

Edit-a-thons

The format of edit-a-thons proved useful and valuable for empowering and training staff members and other interested parties, such as volunteers and citizen scientists, in using Wikidata, as well as in creating and enriching data. As a research tool, Wikidata helps to find and to extract information on collections in general and more specifically on collection agents. By providing identifiers for different entities and linking to external identifiers, scattered information can be connected, queried and analysed. Edit-a-thons are, thus, helping in assembling scattered puzzle pieces.

The edit-a-thons also provided a platform for sharing and validating knowledge, for discussing and reflecting on collections and historical contexts. Bringing together participants from a variety of backgrounds, some with specialist knowledge, helped to include different perspectives and to identify bias and gaps. Collaborative formats such as edit-a-thons can help to form a community and, with the options, to organise the events virtually allows them to engage with people from anywhere. By collaborating with project partners and by reaching out to Wiki communities in the respective countries, the perspective of actors in the countries of origin can be better incorporated and their needs identified. Possibly, some participants act as multipliers and spread the word about Wikidata.

For future edit-a-thons, it is recommended to separate specific training events from dedicated editing events. This would mean that the true Wikidata edit-a-thons would

invite somewhat advanced users and include only a short introductory session at the beginning. Training workshops focusing on Wikidata and on other relevant tools such as OpenRefine are already organised by national initiatives and projects focusing on research data management, such as [WiNoDa](#) (*Wissenslabor für naturwissenschaftliche Sammlungen und objektzentrierte Daten*), [SODa](#) (*SODa – Sammlungen Objekte Datenkompetenzen*) or the National Research Data Infrastructure Germany ([NFDI](#)), local initiatives (e.g. Forschungs- und Kompetenzzentrum Digitalisierung Berlin [digiS](#)) or regional networks focusing on digitisation and research data management (e.g. Netzwerk Forschungsdaten Berlin-Brandenburg, [NFDBB](#)). Such workshops could be co-organised or combined with edit-a-thons in collaboration with the MfN.

Conclusions

Open and freely reusable data available in Wikidata will benefit everyone including many initiatives from local and national projects and activities (e.g. NFDI) to European and international efforts (e.g. DiSSCo, GBIF). However, collaborations need to build on common understanding of data and the challenges of domain-specific standards (e.g. [Darwin Core](#) and [ABCD](#) for biodiversity data or [LIDO](#) as part of [CIDOC-CRM](#) in museum contexts) and limiting interoperability. Although each transdisciplinary research project contributes to the liberation of data from domain-specific silos, they are asked to make their efforts transparent and findings available through knowledge infrastructure, such as Wikidata or more formal bodies driving the standard development (e.g. [TDWG](#), [RDA](#), [NFDI](#)) for a general gain of information and improved data quality due to input from multiple domains. For example, in the context of transdisciplinary provenance research, opening up internal information on collection agents facilitates cross-domain studies (e.g. on collection practices) as was successfully shown during the second project edit-a-thon.

Project results have shown that, by creating Wikidata items and identifiers for collection agents, it is possible to unambiguously identify people and to link them to other relevant entities such as collections, archives and other institutions. Wikidata items, newly created or enriched during the project, were further expanded by other Wikidata users and reused elsewhere. Therefore, a major recommendation from the project is to use Wikidata as an open and collaborative platform and central community-curated hub for data about collection agents, thus making collection-related data as FAIR as possible, i.e. findable, accessible, interoperable and reusable (Wilkinson et al. 2016), available in both human- and machine-readable formats, as well as several languages and linked to other external identifiers. The main advantages of Wikidata in the context of data management in collections can be summarised as follows:

1. free and open, i.e. no costs apart from internet access, accessible without authentication (but creating an account for editing is recommended) and data are published under the Creative Commons [CC0](#) licence, allowing reuse of the data;
2. multilingual, i.e. the user interface and data are available in many languages, property labels and descriptions are translated by the community (via manual editing and bots);

3. human- and machine-readable data can be read, edited and queried by anyone, humans and machines alike;
4. collaborative, an active global user community (e.g. WikiProjects for many topics);
5. easy to learn, easily explained to and understood by new users;
6. revision history allows us to retrace edits and if necessary revert changes;
7. by adding several referenced values, alternative perspectives can be recorded (e.g. agent was collector and looter);
8. connected, as part of the LOD cloud linked to many other databases or aggregators; and
9. a stable and sustainable platform.

Increased data linkage and integration are central for advancing research in biological sciences and natural history and this includes timely integration of Wikidata identifiers for collection agents in the collection management systems and the archival system. Collaborative workshops such as edit-a-thons can help to enrich Wikidata items and provide LOD that can be reused by anyone. Overall, the testing of the format was successful and showed that edit-a-thons could be a valuable addition for future collection disclosure activities and other projects at the MfN.

Outlook

The effort to provide openly accessible data on collection agents will be continued within the framework of the Museum's *Future plan*. Further Wikidata training will be offered to improve digital literacy of museum staff and more specifically to develop their Wikidata skills and their understanding of the potential of LOD for collection and data management. Future edit-a-thons are planned that could have overarching or specific topics, to generate open data linked to the MfN collection and other natural history collections. Such editing workshops could focus, for example, on certain taxonomic groups and connected collection agents or specific regions and people that were active there, thus fostering engagement with communities of experts or communities of origin. A key requirement for unlocking the full potential of LOD is to integrate Wikidata identifiers into the MfN data management systems. The benefits of adding and enriching further collectors or other collection agents to Wikidata will then be directly visible and data become (re)usable by collection staff for verification and quality control. When data about further collection agents are disclosed, a focus should also be on adding under-represented groups such as women and indigenous collectors. By linking to other entities such as publications, described taxa, collections and specimens, other significant people or places, the evolving linked open dataset could be used for network analyses. Tracking data (re)uses would help to show benefits for other actors, partners or projects, therefore justifying resources and investments that went into opening up, linking and integrating data. Another future step is to expand from historical collection agents to living people actively contributing to science, for example, by authoring publications, collecting or identifying specimens. They should use identifiers (e.g. ORCID identifiers) to be unambiguously linked to their research outputs (Groom et al. 2022, Meeus et al. 2023).

The MfN collector wiki also contains information about a number of historical research expeditions. Within the framework of a proposed TDWG working group on Research expeditions, more work will be done to model and contextualise data on such expeditions (von Mering et al. 2023).

Ethical considerations should shape the publication of data. This involves recognising power structures, hierarchies and inequalities that were in place when information and collections were gathered and might be still in place when data is produced and published (Collins 2000, D'Ignazio and Klein 2020, Kaiser et al. 2023). For example, initiatives like the [CARE principles](#) for Indigenous Data Governance (Carroll et al. 2020) highlight the importance to evaluate whether indigenous actors might have reasons to ask for either restricted or limited access to some collection data and media linked to collection agents, especially objects and archival material (e.g. Kaiser et al. 2023, Möhrle 2023). Therefore, consultations with the communities of origin and careful reflection are necessary. Wikidata provides access to information and helps to bring together different stakeholders. However, knowledge contextualisation especially in historical contexts and how best to model information about, for example, agents and collections from colonial contexts is still under discussion in the community. This debate needs to be continued (e.g. Schwarz et al. 2023, Möhrle 2023, Kaiser and von Mering, in press).

With its large and historically important collection, the MfN has a special responsibility to open up internal collection information to enable research into a wide range of topics, including provenance and colonial history. Emerging technologies, such as artificial intelligence and machine-learning, can help to accelerate this process, but need to be critically considered (Thiel and Bernhardt 2023, Belot et al. 2023).

Acknowledgements

Funding for the project was provided by the “MfN Innovation fund” and a COST Mobilise grant for SvM (TDWG 2022; E-COST-GRANT-CA17106-6341d734). This project was only possible due to the support by colleagues from many different departments and teams at the MfN, including different collections, data management, archive, library and coordination office for scientific publishing. We thank the project team and contributors to the MfN “Sammler-Wiki”, especially Eva Patzschke, Alvaro Ortiz Troncoso, Andreas Abele-Rassuly, Anja Friederichs and others who have contributed to the Wiki in the past. Sabine Hackethal, Ralf-Thomas Schmitt and Ferdinand Damaschun (all MfN) are thanked for providing important information and literature about the history of the MfN. Alvaro Ortiz Troncoso and Falko Glöckler (both MfN) are thanked for data exports. In addition, we would like to thank all participants of the Wikidata edit-a-thons and mini edit-a-thons organised during and after the project. Siobhan Leachman (Wikimedia Aotearoa - New Zealand), David Shorthouse (Bionomia) and Quentin Groom (Botanic Garden Meise) are thanked for their support and contributions during the edit-a-thons. Holly Little and Erica Krimmel from the Paleo Data Working Group are thanked for supporting this work. We extend our thanks to the wider community of Wiki editors active in different Wikimedia projects. Frederik Berger (MfN) is thanked for valuable feedback that greatly

improved the manuscript. We would also like to thank the reviewers Rebecca Dikow, Erick Lopes Filho and Mateusz Zmudzinski for their valuable comments and suggestions.

Funding program

Museum für Naturkunde Berlin Innovationsfonds (internal institutional funding)

Grant title

„Sammler:innen Edit-a-thons am Museum für Naturkunde – innovative Formatentwicklung für partizipative Wissensvernetzung“

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Alberti SJMM (2018) Museum nature. In: Curry HA, et al. (Ed.) *Worlds of Natural History*. 1st ed. Cambridge University Press, 348–362 pp. <https://doi.org/10.1017/9781108225229.022>
- Ashby J, Machin R (2022) Legacies of Colonial Violence in Natural History Collections. *Journal of Natural Science Collections* 8: 44–54. URL: https://www.natsca.org/sites/default/files/publications/JoNSC-Vol8-Ashby_and_Machin_2021_0.pdf
- 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. <https://doi.org/10.5281/zenodo.6977243>
- Belot M, Preuss L, Tuberosa J, Claessen M, Svezhentseva O, Schuster F, Bölling C, Léger T (2023) High Throughput Information Extraction of Printed Specimen Labels from Large-Scale Digitization of Entomological Collections using a Semi-Automated Pipeline. *Biodiversity Information Science and Standards* 7: e112466. <https://doi.org/10.3897/biss.7.112466>
- Carroll SR, Garba I, Figueroa-Rodríguez O, Holbrook J, Lovett R, Materechera S, Parsons M, Raseroka K, Rodríguez-Lonebear D, Rowe R, Sara R, Walker J, Anderson J, Hudson M (2020) The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19 (43): 1–12. <https://doi.org/10.5334/dsj-2020-043>
- Centralblatt (1890) Anweisung betreffend die Behandlung der aus den Deutschen Schutzgebieten eingehenden wissenschaftlichen Sendungen. *Centralblatt für die gesamte Unterrichtsverwaltung in Preußen* 32 (10/11): 644–645.
- Cisneros JC, Raja N, Ghilardi A, Dunne E, Pinheiro F, Regalado Fernández OR, Sales MF, Rodríguez-de la Rosa R, Miranda-Martínez A, González-Mora S, Bantim RM, de Lima F, Pardo J (2022) Digging deeper into colonial palaeontological practices in modern day

Mexico and Brazil. Royal Society Open Science 9 (3): 210898. <https://doi.org/10.1098/rsos.210898>

- Collins PH (2000) Black Feminist Thought. Knowledge, Consciousness, and the Politics of Empowerment. 2nd. Routledge, New York and London. URL: <https://negrasoulblog.files.wordpress.com/2016/04/patricia-hill-collins-black-feminist-thought.pdf>
- Damaschun F, Hackethal S, Landsberg H, Leinfelder R (Eds) (2010) Klasse, Ordnung, Art. 200 Jahre Museum für Naturkunde. Basilisken-Press, Rangsdorf.
- D'Ignazio C, Klein LF (2020) Data Feminism. The MIT Press, Cambridge, MA. <https://doi.org/10.7551/mitpress/11805.001.0001>
- DMB (2021) Guidelines on dealing with collections from colonial contexts. Deutscher Museumsbund. URL: <https://www.museumsbund.de/publikationen/guidelines-on-dealing-with-collections-from-colonial-contexts-2/>
- Döring R, Dressler S (2018) Index Collectorum Herbarii Senckenbergiani (FR). URL: https://www.senckenberg.de/wp-content/uploads/2020/04/Index_Collectorum.pdf
- Fagerving A (2023) Wikidata for authority control: sharing museum knowledge with the world. Digital Humanities in the Nordic and Baltic Countries Publications 5 (1): 222-239. <https://doi.org/10.5617/dhnbpub.10665>
- Framework Principles (2019) Framework Principles for dealing with collections from colonial contexts agreed by the Federal Government Commissioner for Culture and the Media, the Federal Foreign Office Minister of State for International Cultural Policy, the Cultural Affairs Ministers of the Länder and the municipal umbrella organisations. URL: <https://www.auswaertiges-amt.de/blob/2210152/b2731f8b59210c77c68177cdcd3d03de/190412-stm-m-sammlungsgut-kolonial-kontext-en-data.pdf>
- Gladstone I, Pearl P (2022) Extinction Voices, Extinction Silences: Reflecting on a Decolonial Role for Natural History Exhibits in Promoting Thinking about Global Ecological Crisis, Using a Case Study from Bristol Museums. Museum & Society 20 (1): 50-70. <https://doi.org/10.29311/mas.v20i1.3806>
- Groll E (2017) Biographies of the Entomologists of the World – Online database, version 8, Senckenberg Deutsches Entomologisches Institut, Müncheberg. <https://sdei.senckenberg.de/biographies>. Accessed on: 2023-3-13.
- Groom Q, Güntsch A, Huybrechts P, Kearney N, Leachman S, Nicolson N, Page RDM, Shorthouse DP, Thessen AE, Haston E (2020) People are essential to linking biodiversity data. Database 2020: baaa072. <https://doi.org/10.1093/database/baaa072>
- 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>
- Hackethal S (1985) Kurzbiographien und Porträts Berliner Zoologen. Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin. Reihe Mathematik-Naturwissenschaften. Beiträge zur Entwicklung der Biologie in der Geschichte der Berliner Universität 3: 385-406.
- Heumann I, Stoecker H, Tamborini M, Vennen M (2018) Dinosaurierfragmente. Zur Geschichte der Tendaguru-Expedition und ihrer Objekte, 1906-2018. Wallstein, Göttingen.
- Hoffmann J, Quaisser C, Berger F, Blessing A, Glöckler F, Herrmann E, Hoffmann A, Petersen M, Schuster F, Tata N (2022a) Illustration des Projektes Sammlungserschließung und -entwicklung am Museum für Naturkunde Berlin. Museum

für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. <https://doi.org/10.7479/44ds-qd81>

- Hoffmann J, Quaiser C, Berger F, Blessing A, Glöckler F, Herrmann E, Hoffmann A, Petersen M, Schuster F, Tata N (2022b) Illustration des Prozesses der Sammlungserschließung am Museum für Naturkunde Berlin. Museum für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. <https://doi.org/10.7479/ryd4-p845>
- Hoppe G (2003) Zur Geschichte der Geowissenschaften im Museum für Naturkunde zu Berlin. Teil 5: Vom Mineralogischen Museum im Hauptgebäude der Universität zu den zwei geowissenschaftlichen Institutionen im Museum für Naturkunde - 1856 bis 1910. Mitteilungen aus dem Museum für Naturkunde in Berlin, Geowissenschaftliche Reihe 6 (1): 3-51. <https://doi.org/10.1002/mmng.20030060102>
- Kaiser K (2022) Duplicate networks: the Berlin botanical institutions as a 'clearing house' for colonial plant material, 1891-1920. The British Journal for the History of Science 55 (3): 279-296. <https://doi.org/10.1017/S0007087422000139>
- Kaiser K, Heumann I, Keysar H, Petersen M, Nadim T, Berger F, Korun M (2023) Promises of Mass Digitisation and the Colonial Realities of Natural History Collections. Journal of Natural Science Collections 11: 13-25. URL: <https://www.natsca.org/article/2796>
- Kücklich J (2020) Free Vector ICONS Vektorgrafiken aus der Open-Access-Strategie des Landes Brandenburg zum Thema Offenheit in Wissenschaft, Forschung und Kultur: Wissenschaftskommunikation, wissenschaftliches Publizieren, Open Access, Open Data, Open Science (1.0). Zenodo. <https://doi.org/10.5281/zenodo.3674561>
- Lentge-Maaß N, Berger F, et al. (2022) Praktischer Leitfaden für digitale Sammlungserschließungsprozesse am Museum für Naturkunde. Museum für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. <https://doi.org/10.7479/8j90-v917>
- Little H, Karim T, Krimmel E, Norton B, Utrup J, Walker L, Van Veldhuizen J (2022) Community Data Mobilization in Wikidata: A paleontology perspective. Biodiversity Information Science and Standards 6: e94416. <https://doi.org/10.3897/biss.6.94416>
- Meeus S, Ariño A, Bakken T, Braun P, Dillen M, Endresen D, Haston E, Lowe M, Meyke E, Santos J, von Mering S, Groom Q (2023) Who is Who in Natural History Collections? Zenodo. <https://doi.org/10.5281/zenodo.7781754>
- MfN (2022) Richtlinie zur Urheberrechtsauszeichnung für Medien und Daten aus der Sammlungserschließung. Museum für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. <https://doi.org/10.7479/td7j-7d69>
- Möhrle F (2023) Archive als Orte des kulturellen Gedächtnisses – Der Umgang mit Archivgut aus kolonialen Kontexten am Beispiel des Museums für Naturkunde Berlin. University of Potsdam, Master thesis.
- Page RM (2016) Towards a biodiversity knowledge graph. Research Ideas and Outcomes 2: e8767. <https://doi.org/10.3897/rio.2.e8767>
- Page RM (2022) Wikidata and the bibliography of life. PeerJ 10: e13712. <https://doi.org/10.7717/peerj.13712>
- Page RM (2023) Ten years and a million links: building a global taxonomic library connecting persistent identifiers for names, publications and people. Biodiversity Data Journal 11: e107914. <https://doi.org/10.3897/bdj.11.e107914>

- Patzschke E, Ortiz-Troncoso A, Abele A (2016) WIKI-Ansatz und kollaboratives Arbeiten im Forschungsmuseum. BMBF 2013 – 2016; Förderkennzeichen 03IO1315. Final project report. <https://doi.org/10.2314/GBV:872479447>
- Reimers Y (2022) Sammlungs- und Dokumentationsprofil Historische Bild- und Schriftgutsammlungen. Museum für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. <https://doi.org/10.7479/aeth-rz39>
- Savoy B (2022) Africa's Struggle for Its Art: History of a Postcolonial Defeat. Princeton University Press, Princeton, NJ. <https://doi.org/10.1515/9780691235912>
- Schwarz A, Möhrle F, von Mering S (2023) Collections from Colonial Australia in Berlin's Museum für Naturkunde and the Challenges of Data Accessibility. Biodiversity Information Science and Standards 7: e111980. <https://doi.org/10.3897/biss.7.111980>
- Shafee T, Mietchen D, Lubiana T, Jemielniak D, Waagmeester A (2023) Ten quick tips for editing Wikidata. PLOS Computational Biology 19 (7): e1011235. <https://doi.org/10.1371/journal.pcbi.1011235>
- Shorthouse D (2020a) Bionomia: Linking natural history specimens to the world's collectors. <https://bionomia.net>
- Shorthouse D (2020b) Slingshotting 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>
- SMB (2023) Acquisition and Accession Logs of the Ethnologisches Museum. Staatliche Museen zu Berlin. URL: <https://www.smb.museum/en/museums-institutions/ethnologisches-museum/collection-research/acquisition-and-accession-logs/>
- Thiel S, Bernhardt J (Eds) (2023) AI in Museums. Reflections, Perspectives and Applications. transcript, Bielefeld. <https://doi.org/10.14361/9783839467107>
- 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>
- von Mering S, Stolze E (2023) Collection agents affiliated with the Museum für Naturkunde Berlin. [Data set]. Museum für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. <https://doi.org/10.5281/zenodo.10425160>
- von Mering S, Braun P, Cubey R, Groom Q, Haston E, Hendriksen A, Johaadien R, Leachman S, Marsden L, Rainer H, Santos J, Endresen D (2023) Modelling Research Expeditions in Wikidata: Best Practice for Standardisation and Contextualisation. Biodiversity Information Science and Standards 7: e111427. <https://doi.org/10.3897/biss.7.111427>
- Vrandečić D, Krötzsch M (2014) Wikidata: a free collaborative knowledgebase. Communications of the ACM 57 (10): 78-85. <https://doi.org/10.1145/2629489>
- Waagmeester A, Stupp G, Burgstaller-Muehlbacher S, Good BM, Griffith M, Griffith OL, Hanspers K, Hermjakob H, Hudson TS, Hybiske K, Keating SM, Manske M, Mayers M, Mietchen D, Mittra E, Pico AR, Putman T, Riutta A, Queralt-Rosinach N, Schriml LM, Shafee T, Slenter D, Stephan R, Thornton K, Tsung G, Tu R, Ul-Hasan S, Willighagen E, Wu C, Su AI (2020) Wikidata as a knowledge graph for the life sciences. eLife 9: e52614. <https://doi.org/10.7554/elife.52614>
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P,

Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>

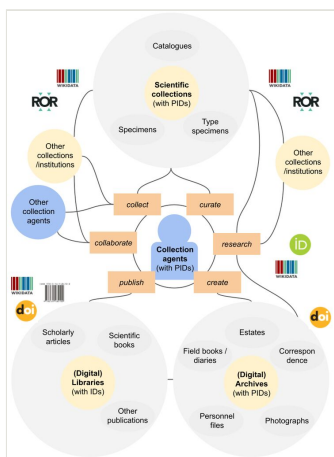


Figure 1.

Collection agents play a central role in all steps of the research process. Persistent identifiers (PIDs) enable linking people with collections, libraries and archives, as well as other institutions and agents ([Wikidata identifier](#) for people or other entities, [ROR identifier](#) for research organisations, [ORCID ID](#) for researcher or Digital Object Identifier ([DOI](#)) for objects, for example, publications or specimens).

Sammlerbiographien

Arbeitsunterstützung der Sammlungsmitarbeiterinnen und -mitarbeiter hinsichtlich Recherchen zu Sammlern. Informationen über Personendaten, Reiserouten, Sammelgebiete, Referenzen, Quellen, weiterführende Informationen im www. Vernetzung mit der Historischen Arbeitsstelle. Das Wissen der Vielen zusammenführen, ergänzen und teilen.

Navigation

- Hauptseite
- Sammlerbiographien
- Expeditionen

Intern

- AG Sammlerbiographien
- Personenlisten aus den Sammlungen
- Seiten mit Lizenzhinweis
- Vorlagen und Redaktionelles
- Empfohlene Aufgaben
- Hilfe
- Hilfe zum Sammler-Formular

Sammler des Tages

Küenthal, Willy Georg

- 1883 und 1885 Westküste Norwegens [1]
- 1886 Fahrt in die Arktis, um Wale zu studieren [1]
- 1889 eine zweite Arktisfahrt Ostküste Spitzbergens und nach König-Karl-Land [1]
- 1894 Malaisches Archipel und nördliche Molukkeninsel [1]

Figure 2.

Screenshot of the homepage of the internal MfN collector wiki (officially, the wiki “Sammlerbiographien”, i.e. on collector’s biographies) highlighting a “collector of the day”, in this case [Willy Küenthal](#).

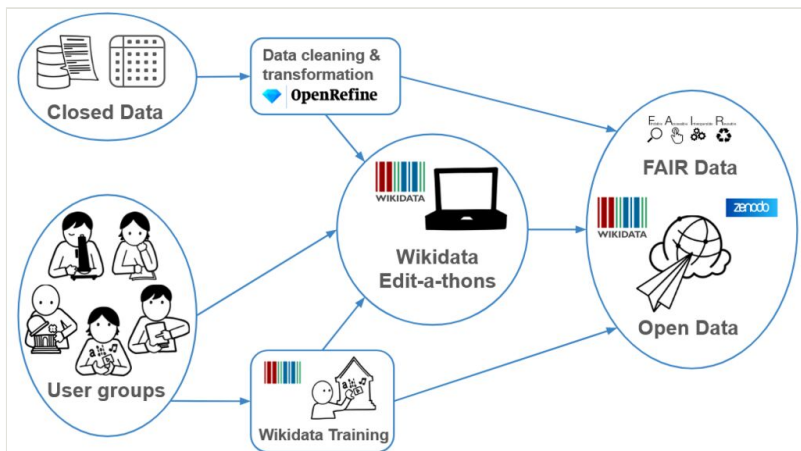


Figure 3.

Workflow from closed data to Open Data following the FAIR principles – using participatory formats such as Wikidata training and edit-a-thons. Edit-a-thons can be customised to fit the needs of different user groups, for example, the number of participants, the type of participation (in person, hybrid or virtual) and the length of the event. Different user groups included MfN staff members from different departments, researcher and colleagues from other GLAM institution, citizen scientists, volunteers and other interested parties. Icons used from Kücklich (2020) (CC0 via Zenodo/Wikimedia Commons), illustration of FAIR data principles from SangyaPundir (CC-BY via [Commons](#)).

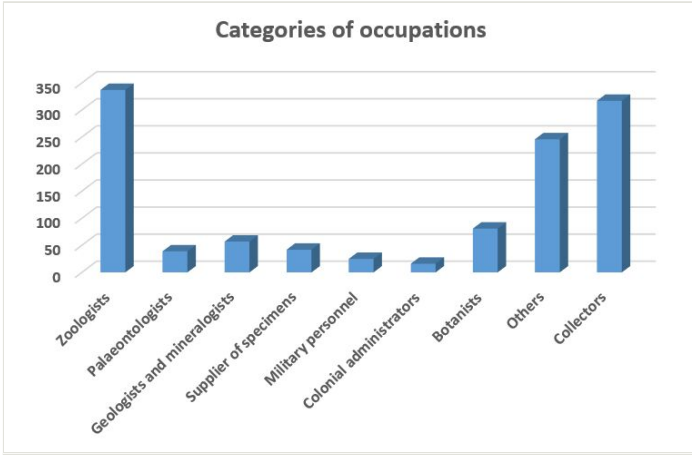


Figure 4.
Main categories of occupations within the dataset of collection agents (total number of disambiguated people = 539; multiple mentions possible).

Table 1.

Numbers and percentages related to the MfN collector wiki dataset (double entries possible).

Entries in the MfN collector wiki	Number	Percentage
<i>Data transfer and editing</i>		
Distinct collection agents in the dataset (after cleaning and deduplication)	596	100.0
Total disambiguated collection agents with Wikidata identifier after project	539	90.4
Newly created Wikidata items for collection agents from the dataset	72	12.1
Newly created Wikidata items that were enriched by other Wikidata users	60	10.1
Collection agents that could not be disambiguated due to a lack of information, i.e. no Wikidata item was created at this point	57	9.6
<i>Data analysis (percentages refer to 539 disambiguated collection agents)</i>		
Collection agents that were employed at the MfN	134	24.9
Collection agents with collection items at the MfN	255	47.3
Collection agents with archival material at the MfN	294	54.5
Collection agents with GND number	452	83.9
Collection agents with Bionomia profiles	304	56.4

Table 2.

Categories of occupation, included specialisations and occupations and number per category.

Category of occupation	includes	Number
Zoologists	ornithologist, mammalogist, herpetologist, entomologist, ichthyologist, arachnologist, carcinologist, malacologist, conchologist, bryozoologist, marine biologist, protozoologist, helmithologist etc.	337
Palaeontologists	palaeozoologist, vertebrate palaeontologist, palaeobotanist etc.	39
Geologists and mineralogists	petrologist, chrystallographer	57
Supplier of specimens	dealer of naturalia, insect dealer, owner of menagerie, animal trader, trader of minerals etc.	42
Military personnel	military officer, military physician, member of Schutztruppe etc.	25
Colonial administrators	government officials in former colonies (e.g. Resident, Stationsleiter, Bezirksleiter, governor)	16
Botanists	mycologist	81
Others	explorer, physician, (naval) surgeon, anatomist, physiologist, pharmacist, chemist, university teacher, teacher, veterinary, ethnologist, anthropologist, archaeologist, theologian, pastor, priest, missionary, translator, linguist, diplomat, politician, jurist, librarian, model maker, engraver, preparator, taxidermist, scientific illustrator, photographer, farmer, (big game) hunter, gardener, forester, geographer, meteorologist, writer, painter, poet etc.	246
Collectors	zoological collector, botanical collector, fossil collector, collector of ethnographica etc.	317