



Submitted on: 01.10.2023

Vocabulary Work in the Field of the Network Colonial Contexts - A workshop report

Moritz Strickert

Specialised Information Service for Social and Cultural Anthropology
University Library of Humboldt-Universität zu Berlin, Berlin, Germany
moritz.strickert@ub.hu-berlin.de



Copyright © 2023 by Moritz Strickert. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:
<http://creativecommons.org/licenses/by/4.0>

Abstract:

The Germany-centred network Colonial Contexts brings together experts from different areas of the GLAM spectrum. It deals with questions of digital consolidation, processing and visibility, as well as possibilities of use of digital materials and data from colonial contexts. To ensure interconnectedness and accessibility, the use of controlled vocabulary is central. To address these issues, a separate working group has been established. The Working Group Thesauri aims to gain an overview regarding existing thesauri/vocabularies, tries to network them and discuss ethical questions when it comes to questions of naming especially regarding cultural groups.

Keywords: Authority Files, Colonialism, Controlled Vocabularies, Decolonization, Library Ethics

In recent years, public debates about adequate forms of dealing with the colonial past and measures of decolonization in Germany have strongly increased. This includes politics of remembrance, questions of the restitution of collections of German museums and research institutions¹, and renaming of problematic street names² as a naming practice in public space.

The discussion of colonial structures takes place in many academic institutions simultaneously and affects the work of libraries.³ In the German-speaking world, the topic is being addressed in various forms, for example, by the open network *Decolonize the Library*⁴; thematic overlaps and discussions have also taken place during the digital *Scholar's Lab Critical Library Perspectives*⁵. These initiatives have in common that they increasingly situate libraries as embedded in social structures and discourses and do not perceive them as neutral institutions.

Colonialism/Decolonization in the work of the Specialised Information Service for Social and Cultural Anthropology

The Specialised Information Service⁶ for Social and Cultural Anthropology (FID SKA) addresses its services to the scientific communities and related institutions of the ethnological disciplines⁷. Historically, the genesis of these disciplines was often closely interwoven with colonial structures. The FID SKA takes up the topic in its work program regarding the engagement with relevant resources originating from libraries, archives as well as collections and the handling of these. Significant here is not so much the provenance of the material, e.g. in the form of library holdings. The issue is of greater importance in the context of museum collections. In the case of library collections, it is rather the terminology and content - especially when digitized - that are to be viewed critically. Moreover, the FID SKA also considers forms of indexing, for example, regarding the terminologies used, and the control of data collections, especially in the digital space. This entails how (sensitive) data from research can be archived, be adequately described and made available for potential reuse. In the process, the needs of the broader public meet the sometimes different needs of science. On the one hand, this includes scientific language(s) based on debates

¹ The most prominent example in this context is the return of the so-called Benin bronzes. The debate about this was even held in the German parliament: <https://www.bundestag.de/dokumente/textarchiv/2023/kw19-de-aktuelle-stunde-benin-bronzen-947704>.

² An overview of colonial street names and their renaming in the Mitte district (Berlin) can be found here: <https://www.berlin.de/kunst-und-kultur-mitte/geschichte/erinnerungskultur/strassenbenennungen/artikel.1066742.php>.

³ Demands for the integration of voices from the Global South in the form of texts used in teaching, affect the development of library collections. Regarding the presentation of holdings and discoverability, questions about the revision of classification and other indexing systems also play a role. In terms of scholarly publishing (open access), attention must also be paid to the complex of publication fees and how these can be paid by researchers with fewer financial resources.

⁴ The wiki of the group can be found at: <https://decolonizethelibrary.miraheze.org/wiki/Hauptseite>.

⁵ This brought together researchers and library staff in May-June 2022 for cross-domain discussions on diversity and ethics, among other topics. More information about the event: <https://lab.sbb.berlin/events/critical-libraryperspectives/#programm>. The results of the work can be found at: 027.7 Zeitschrift für Bibliothekskultur. 2022. Critical Library Perspectives. Volume 9. Issue 4. DOI: <https://doi.org/10.21428/1bfadeb6.2d09bc20>.

Another project currently under development is "IN_CONTEXT: Colonial Histories and Digital Collections," which is based at the Staatsbibliothek zu Berlin. For more information see: https://blog.sbb.berlin/in_context_01/.

⁶ The German system of specialised information services envisages that the needs of the specialist communities are taken up by libraries and that specialised information services support researchers as infrastructure partners by means of research-related services. Despite their institutional affiliation, the specialised information services often deal with questions that are also relevant for other institutions, beyond their own institution and thus transregionally.

⁷ This includes university and non-university institutes, museums and other memory institutions. FID SKA is based at the University Library of the Humboldt University in Berlin and the Qualiservice research data center at the University of Bremen. Further information can be found at: <https://www.evifa.de/en/about>.

within scientific disciplines and, on the other hand, a more general language (in this case, German) aimed at broad comprehensibility, which is also shaped by a perspective of the Global North. Increasing global connectedness and availability of resources also requires reflection on the role of knowledge organization systems, their specific perspective, and how these can be made more inclusive.

Introduction of the Network Colonial Contexts

The developments and questions described above are also relevant to the Network for Sustainable Research Structures for the Processing of Collections from Colonial Contexts⁸ (Network Colonial Contexts). Founded in the autumn of 2020, it is an open association of researchers and research infrastructures focusing on provenance research or the (retro-)digitization of colonial collections, but also on the bundled presentation and description of these holdings. This includes specialised information services as well as museums, (ethnological) collections, archives, libraries and research institutions (including (provenance) researchers in social and cultural anthropology, history, political science and art studies) who deal with materials from colonial contexts in their work. Since its foundation, regular meetings and international workshops have been taking place. In addition, separate working groups, each with a specific work focus, have been established.⁹

At the centre of the network is the question of how it is possible to (digitally) bring together the data and materials from colonial contexts in a meaningful way, describe them, process them, and make them visible and usable in a sustainable way?

All this not only from a technical point of view but against the background of (research-)ethical considerations. The resources covered include not only objects but also writings related to the topic, for example, in the form of related museum/archive finding aids, correspondence, historical sources from the colonial period, or research on this topic.

Three overlapping areas are of particular importance in the practical work of the network: networking, technical development, and visibility. The former aims to improve the coordination and linking of thematically appropriate projects from different professional and institutional fields. It is essential not only to focus on German-speaking countries but also to promote international cooperation so that these perspectives, diverse experiences and approaches can be better reflected and integrated into practical work. Special attention is paid to the cooperation with colleagues from the societies of origin and the consideration of their needs.¹⁰

⁸ For further information on the network see: <https://www.evifa.de/en/netzwerk-koloniale-kontexte>.

The term "colonial contexts" is used here in accordance with the definition of the German Museums Association:

"The term "colonial context" describes much more than "only" formal colonial rule [...] Colonial context [is used for] circumstances and processes that have their roots either in formal colonial rule or in colonial structures outside formal colonial rule." In: German Museums Association: Guidelines for German Museums - Care of Collections from Colonial Contexts. Available under: <https://www.museumbund.de/wp-content/uploads/2021/03/mb-leitfaden-web.pdf>, p. 26.

⁹ The working groups include: WG 3-way strategy (critical monitoring of political processes), WG Third-Party Funding (funding of network activities and individual projects), WG International cooperation (organization of international workshops), WG Research tool (evaluation of (cooperative) research tools and environments), WG Thesauri (cooperative vocabulary development see below). For further information see: <https://www.evifa.de/en/netzwerk-koloniale-kontexte/join>.

¹⁰ One concern was to make clear that in regions of origin, Internet access is sometimes limited, so data volumes should be adjusted for databases and high-maintenance software should be used as little as possible.

At the technical level, the aim is to identify the needs of different institutions and the extent to which possible solutions already exist. Isolated solutions are to be avoided thus increasing sustainability. Transparency should be ensured for data that is at times sensitive, while at the same time always paying attention to respectful handling of those materials. In sustainability and networking, the mutual understanding of possible standards in the form of authority files and controlled vocabularies plays a special role.

Beyond indexing and making holdings accessible, visibility includes the public articulation and communication of the needs and interests of members and participating institutions.

Digitization as a driving force

The topic of digitization represents an important point of reference within the network and acts as a starting point for various questions. During digitization projects of historical literature, this included how to deal with the fact that journals and books are accessible detached from the physical placement and thus, no direct contextualization takes place. Material that may contain a strong colonial bias, including discriminatory and racist perspectives, becomes accessible to a wide audience online through digitization - without the scholarly framework of the institution, critical reflections on the collection, and similar framings being necessarily perceived by readers (unlike a physical visit to an academic library). At the same time, this includes as a subfield how to deal with the fact that photographic material from the colonial era can degrade people and/or was created in unclear power relationships/without consent.¹¹ Also important in this context is the preoccupation with sacred or secret objects, rituals, and activities and the extent to which these can be shown and described. All these topics have in common that there is a need to reflect increasingly on questions of contextualization and interpretive authority during expanded accessibility. At the same time, individual research and case examination are hardly logistically possible in mass digitization projects.¹²

At the same time, the digitization of resources offers numerous opportunities. It makes resources accessible on a larger scale. In the best case, an online public sphere is created, together with greater participation and transparency. In the ideal scenario, not only researchers but also societies of origin can benefit. In concrete terms, this online access can serve as a basis for return claims and restitution. Without access to data, e.g. in the form of inventory data or object images, no dialogue about the provenance and possible restitution of objects can develop.

Questions that arise in this context are: Who determines what may be shown and how? Which designation is adequate? What language(s) are necessary? Which region of origin is correct? Which standardized designations for forms of acquisition (exchange, gift, purchase, robbery) are appropriate? Who may have a say and who decides on all this?

¹¹ For an outline of the problem from an ethnological-infrastructure perspective see: Matthias Harbeck und Moritz Strickert, Zeigen/Nichtzeigen, in: Visual History, 21.09.2020, <https://doi.org/10.14765/zzf.dok-1927>.

Matthias Harbeck und Moritz Strickert, Freiwilligkeit und Zwang. Eine Diskussion im Kontext der frühen ethnologischen Fotografie, in: Visual History, 28.09.2020, <https://doi.org/10.14765/zzf.dok-1928>.

¹² These questions are discussed further here: Matthias Harbeck, "Die Ethik des Digitalisierens: Fragen zum Umgang mit Materialien aus kolonialen Kontexten in der Massendigitalisierung". LIBREAS. Library Ideas, 40 (2021). <https://libreas.eu/ausgabe40/harbeck/>.

Perspectivity and cultural embeddedness of controlled vocabularies

Many of these questions are also important when dealing with controlled vocabularies. In many cases, these are embedded in hegemonic knowledge systems of the Global North and therefore reflect specific (historical) perspectives and are not neutral.¹³ In this discussion, questions of foreign and self-designations, for example, of groups, speaker positions and the historicity of the terms, are of particular importance. Relevant questions in this context include: What terms are being used? How did the definition process work out, and to what extent is it possible to integrate other perspectives and to work together on an equal footing while maintaining standardization? To what extent is it possible to implement multilingualism?¹⁴

This complex of questions is of particular importance concerning resources from colonial contexts. Those materials often emerged in historical constellations in which processes of definition and categorization were often unilaterally initiated by the Global North and defined the Global South as an object of knowledge. The management and indexing of these collections cannot do without controlled vocabularies and authority files. During this, authority files are leaving their narrower library focus and are increasingly being used in museums, other institutions, and research projects. These systems are particularly important for bringing together and linking previously separate collections. In the best case, the newly created online presences enable resources to be available to a larger circle of users (especially the societies of origin), thus lowering thresholds and achieving a higher degree of participation. One example is the inclusion of holdings in a separate colonialism-related portal of the German Digital Library.¹⁵

Introduction of the WG Thesauri

During the network Colonial Contexts, the idea of establishing a working group on thesauri arose, which intends to work on these topics collaboratively. This should provide the basis for improved discoverability and (international) accessibility for previously scattered holdings. The WG aims to bring together people from different fields (GLAM, scientists and staff of infrastructural institutions) for a common dialogue on controlled vocabularies. For this reason, the WG aims to link, harmonize, and critically reflect on an obtained overview of existing controlled vocabularies/thesauri. However, many of the vocabularies used are decentralised or only available for a limited time. The aim is to bring together other perspectives and expertise, and work on ethical questions together, as this would be too big a task for individuals in its entirety. In doing so, it connects to numerous vocabulary projects that have already revised or created entirely new vocabularies with a power-critical perspective.¹⁶ However, an unchanged transfer of work

¹³ For example, Gartner emphasizes that the seemingly neutral designation or categorization of facts is based on specific decisions, which always have a historical-ideological embedding. Gartner, Richard (2016): *Metadata. Shaping knowledge from antiquity to the semantic web*. Basel: Springer. Gartner (2016), S. 42.

Regarding ethical responsibility in the cataloging process, see: Cataloging Ethics Steering Committee: *Cataloging Code of Ethics*. (2021). <http://hdl.handle.net/11213/16716>.

¹⁴ These questions are also of importance in the GND project of the FID SKA. This project aims at processing the ethnological data holdings of the central German-language authority file, the so-called Gemeinsame Normdatei (GND). More information on this project at: <https://www.evifa.de/en/about/fid-projects/gnd>.

¹⁵ The portal "Collections from Colonial Contexts" makes digitized and indexed collections from so far 25 pilot institutions searchable and available online in the German Digital Library (DDB). Further information can be found here: <https://ccc.deutsche-digitale-bibliothek.de/?lang=en>.

¹⁶ Some projects are for example: "Indigenous and Northern Affairs Canada Subject Thesaurus" <https://www.sacisc.gc.ca/eng/1484595265832/1618943350587>, "Ngā Upoko Tukutuku / Māori subject headings"

processes or concrete work results is not possible due to historical differences and a different work focus. The WG Thesauri strives for cooperation with researchers from regions of origin and representatives of the societies of origin to discuss appropriate terms, e.g. for cultural groups, places and objects. Unlike many projects that have a geographically restricted focus on one country/region of the world and have cooperated directly with more clearly defined groups/communities of origin, this is difficult to do on a global scale with limited resources. Thus, it remains to be established how key recommendations such as "the use of "ethical outreach" when dealing with still-living identities, items, or groups" and "consultation with described subjects" can be appropriately integrated into the work of the WG Thesauri.¹⁷

To structure these tasks, the working group has been divided into three subgroups, which pursue different work goals but cooperate with each other and meet at regular cross-group work meetings. These include:

- Survey
- Evaluation
- Vocabulary processing with digiCULT.xTree

The first group conducted a survey to find out what vocabularies exist in other institutions, in what form, and how they have been used and maintained in practical work. Following the ongoing survey a detailed evaluation will be done. Using the software digiCULT.xTree, a merging of existing vocabularies should finally occur, and these should then be edited collectively in the step.

Current work status

Basis of data

In the course of the WG activities, the idea arose to develop a thesaurus for cultural groups (this includes terms for societies and communities) in an exemplary test run. The development should be accompanied by the creation of an editorial guideline and lead to the development of an editorial workflow. In doing so, the sustainable data stocks of the Gemeinsame Normdatei (GND), as the central German-language authority file, and other vocabularies from other institutions, will be used. Approximately 4000 GND-records serve as a working basis. These include preferred designations and variant designations for cultural groups, a unique identifier, a source citation, and sometimes definitions.

At the same time, they are linked in parts to other vocabularies and foreign language authority files. However, some of the institutions interviewed characterized these datasets in the current status as inadequate regarding the coverage of cultural groups. Despite this limitation, the working group pursues close cooperation with the GND and wants to take advantage of the GND in terms of sustainability and reusability. The goal is to productively combine the advantages of the GND's large and structured data pool with additional review and expertise from the scientific community.

<https://natlib.govt.nz/librarians/nga-upoko-tukutuku>, "Pathways Thesauri" <https://www1.aiatsis.gov.au/>, „Respectful Indigenous Terminology Platform Project" <https://www.nikla-ancla.com/projects>.

¹⁷ Watson, B. M. (2021): Advancing Equitable Cataloging. In: Proceedings from North American Symposium on Knowledge Organization 8. <https://doi.org/10.7152/nasko.v8i1.15887>, S. 9-10.

For an example of the concept of epistemic partnership, see: Duarte, Marisa Elena; Belarde-Lewis, Miranda (2015): Imagining: Creating Spaces for Indigenous Ontologies. In: Cataloging & Classification Quarterly 53 (5-6), S. 677–702. <https://doi.org/10.1080/01639374.2015.1018396>.


| [Sachbegriff (GND)] | | Verwendung: s |
|------------------------|--|---|
| Sachbegriff: | Sikaner | |
| K10plusPPN: | 700426221 |  |
| GND-Nummer: | 4495244-2 | |
| | Link zu diesem Datensatz in der GND: http://d-nb.info/gnd/4495244-2 | |
| Typ: | Ethnografikum (sie) | |
| Definition: | Ureinwohner Siziliens | |
| Quelle: | B 1986 | |
| Ländercode: | XA-IT [Italien] | |
| GND-Systematik: | 17.1 [Volkskunde, Völkerkunde (Allgemeines)] | |

Figure 1 Link: <https://swb.bsz-bw.de/DB=2.104/PPNSET?PPN=700426221&INDEXSET=21>

Working tools

digiCULT.xTree as a tool for collaborative thesaurus work

The online platform digiCULT.xTree serves as a tool for editing the vocabularies. It provides partially free access to other vocabularies and has an interface to many object documentation systems, which is relevant because many members of the working group are affiliated with museums. The program complies with current standards for documentation and thesauri. DigiCULT.xTree currently manages more than 200 vocabularies and provides a forum for exchange regarding terminology work. The project's current work status is that a first import of the GND data sets regarding cultural groups has been done. A check has just been made to see to what extent all relevant fields information has been copied, and there is a need for reworking in some areas. The results produced during the WG's work are to be made available to the public for further use via digiCULT.xTree in the long run.

Other working tools

In addition to working with digiCULT.xTree, the working group also uses other tools to structure its work. A wiki has been established for the collection of relevant resources. In addition to the meeting protocols of the individual working sessions, the wiki contains: Draft editorial guidelines, and a collection of relevant experts we plan to contact for contested issues. In addition, a growing collection of vocabularies already being edited, which serve as a source and inspiration.

As an important tool for structured preparation, a worksheet has been developed for personal planning of the joint discussion and input into digiCULT.xTree.

| Term definition in the local museum database: Term, synonyms etc. | Preferred designation / self designation | Researched in reference work | Record in GND or Wikidata: Term, link | Record in AAT: Term, link | Exact match? (Y/N) | Record in Glottolog (Language): Term, link | Discussion point(s) | Comments |
|---|--|------------------------------|--|---|--------------------|--|--|-----------------------------------|
| Sikaner Synonym: Sikanen | | n.a. Wikipedia | Sikaner http://d-nb.info/gnd/4495244-2 Sicani (Q952123) https://www.wikidata.org/wiki/Q952123 | Sicani http://vocab.getty.edu/page/aat/300312337 | Y | n.a. | Term / synonym checked: non-critical, biased, etc. | Example: ancient people in Sicily |

Figure 2: Worksheet

The worksheet lists the terms of the local museum databases/vocabularies (preferred designation and synonyms) with an additional source citation based on reference works. Here the list of the accepted subject-specific reference works of the GND is consulted, which often serves as a basis for the creation of new records. If the term could not be found there, "n.a." must be entered and an own source must be provided. Furthermore, a search determines whether a matching entry exists in the GND, Wikidata, and the Art & Architecture Thesaurus (AAT), and the term and link are added to the table. At the same time, we check whether the records are exact matches. A coupling with languages by means linking to Glottolog¹⁸ is also envisaged. The last two columns include discussion points regarding the term characteristics and whether these are to be described, e.g. as outdated, discriminating, uncritical. The point comments refers to suggestions for changes regarding the preferred designation of the GND or the necessary enrichment of the data sets with further synonyms.

Open questions and outlook

The project is still in the testing phase, which aims to determine which workflows can be implemented. The next steps are to organize a joint workshop in the fall, in which the working group would like to work collaboratively on the first terms and discuss possible sources that could serve as a basis for the processing. Furthermore, it will be explored how different perspectives (relevant experts and previous experiences) and information can be incorporated into the practical work process.

In practice, the WG must establish how different degrees of granularity in the various vocabularies are handled, e.g. how to deal with smaller subgroups and from what size onwards recording with a separate data set makes sense and is practicable in terms of workload. In the further course, an approach must be developed how to deal with terms that have been identified as problematic. In particular, the historical material and the integration of older vocabularies raise questions about the perspectivity of knowledge, the changing nature of language over time and how to deal with

¹⁸ Glottolog is a (bibliographic) database for indexing the world's languages. It was developed by the Max Planck Institute for Evolutionary Anthropology in Leipzig (Germany). Further information can be found at: <https://glottolog.org/>.

it.¹⁹ An important point is the focus on multilingualism to increase international accessibility, further increased by synonym enrichment and links to other vocabularies. A cooperation with the GND-Mul²⁰ project, aiming at multilingualism, links the project to the GND. The WG Thesauri strives to link back to the GND as a sustainable working tool to address the disadvantages of time-limited and institution-specific vocabularies.

It will be necessary to figure out how to deal with cultural groups that are not yet part of the GND. This term gap is to be closed through the recording in digiCULT.xTree. Subsequently, it will be clarified to what extent these new records can become part of the GND, since they can be made available to a larger circle of users by that. Similar issues are relevant for cultural groups with Wikidata records that are not yet part of the GND. In the further course, it will then also be necessary to consider which proposed changes regarding already existing GND-records could become part of the GND. The potential enrichment of the data sets with alternative terms can represent a further field of work in this context. Changes to the preferred name of a dataset require extensive coordination and may also entail labor-intensive postprocessing. It remains to be seen to what extent sufficient resources are available for this process.

Beyond possible changes of the data pool of the GND the entire thesaurus processed in digiCULT.xTree will be available for museums and other institutions for further use. Through the interface functions the vocabulary can, for example, become part of the local collection management systems.

By working on misrepresentations and biases, the goal is to achieve a higher level of inclusion. This is to be done by improving the search functions through a multitude of synonyms, at best multilingual, and by making holdings accessible with a more extensive vocabulary provided with identifiers. At the same time, cultural groups should be represented adequately and according to their own designations and, if possible, it should be avoided that foreign designations are uncritically adopted and visibly displayed on platforms.

This higher level of involvement also helps to support sustainability goals. In this way, the goal of education can be taken up and thereby lead to inclusive education and promote lifelong learning opportunities for all. In the long term, this might be a way to build global partnerships and promote an intensified knowledge exchange.

¹⁹ The intention is to mark/contextualize problematic terms that occur in vocabularies collected and not to delete them completely, as this would affect historically oriented research. At the same time, an integration of problematic synonyms is intended to enable search results that are as low-threshold as possible, since users may only be familiar with these terms.

²⁰ More information about this project at: https://gnd.network/Webs/gnd/DE/Projekte/GNDmul/GNDmul_node.html.