Basler Africa Portal

Alice Spinnler, Martin Reisacher, Andreas Ledl, and David Tréfás

Basel University Library
Email: <firstname.lastname>@unibas.ch

Abstract

The article aims to present the project "Africa Portal" in which five institutions located in Basel create an integrated search interface in order to make their data accessible. Following a description of the process concerning normalizing different data structures and formats, the article also focuses on the outcomes and the possible developments of the "Africa Portal" in the future.

1 Introduction

The aim of the project "Basler Africa Portal" is to make visible research about Africa and specifically the research about Africa which has been carried out in Basel. It is important to include works by researchers based in Africa because in most cases its visibility is scarce both in printed and in online publications.

For centuries, Africa has been a focus of researchers in Basel. Economic, missionary, and medical research, as well as interests in healthcare and science, have brought Baslers to Africa; for either short or longer visits. Their research data and publications are brought together in the Africa-portal and as a result have become accessible.

2 Background

Several institutes with a strong interest in Africa are located in Basel:

- Mission 21
- Swiss Tropical and Public Health Institute
- Museum der Kulturen
- Basler Afrika Bibliographien
- Universität Basel: Centre for African Studies Basel, University Library

An Africa Portal already exists on the homepage of the Centre for African Studies Basel. It contains links to the catalogues, databases and archives of the above mentioned institutes. However, up until now, this data has not been brought together.

The idea to make this possible stems from a workshop about Linked Open Data, held in Basel 2016. During the workshop, the project ZHART (Zurich Art) was presented to participants. ZHART contains diverse information sources which have been brought together and made
accessible through one search engine. This was the starting point for the integrated search engine which is expected to replace the already-existing Africa Portal.

Representatives of five Basel Africa-related institutes met for the first time on 8 December, 2016 and agreed to tackle the problem by developing a new Africa Portal. The concept phase began in March 2017. Now, one year and eight project meetings later, the project has ripened. Soon there will be a beta-version of the "Research Data Viewer".

3 The Project Africa Portal

As mentioned above, the aim of the project was to bring together data and publication collections from all participating institutes, which are:

- Basler Afrika-Bibliographien (BAB): Library and Archive
- Mission 21 (M21): Library
- Museum der Kulturen (MDK): Library, ethnographic and photo collection
- Swiss Tropical and Public Health Institute (TPH): Library
- Universitätsbibliothek (UB): Library and EDOC

The different scopes of the institutions are accompanied by a wide variety of collected material:

- Books, journals, articles
- Films, videos, DVDs
- Photographies
- Audio sources
- Posters
- Manuscripts
- Maps
- Ethnographic objects
Archives and ethnological collections each have their own indexing systems. The Basler Afrika-Bibliographien even has its own library system and also developed a singular subject indexing system to include hierarchies. The Museum der Kulturen uses free keywords. All other partners’ library connection are brought together in the catalogue of IDS Basel Bern – Swissbib Basel Bern. Ever since 2011, indexing has followed the so-called "rules for the catalogue of subject headings" (RSWK) and the Integrated Authority File GND, which is collectively operated by the German National Library and a network of libraries in German-speaking countries. In 2014, non indexed title records were enriched by GND-data, wherever possible. Subject headings stemming from earlier periods aren’t adopted to GND-standards yet, but will be if feasible.

4 Different data, different formats

Cataloguing data provided by Swissbib Basel Bern is based on international standards such as RDA and MARC 21. Library catalogues operated by BAB, archives, and the ethnographical collection follow other standards.

Due to the different indexing traditions in libraries, museums, and archives, as well as different software solutions, it would require a great effort to deliver all data in a uniform exchange format. Therefore, the Africa portal will use raw data the way it has been entered onto the various databases. The minimal requirement is to define core elements which are common for all participants which are automatically recognisable and can be assigned to a certain institution.

There is a plan to enable mapping by using Google spreadsheets and to adjust the data in a decentralised way, using OpenRefine. This could mean that more institutes could participate in the future without the use of IT resources, so long as they did not use data models which were too complex.

Due to the different indexing traditions, a quite simple core element set has been defined:

- Person (author, editor, photographer, producer, director, collector, vendor...)
- Description (title, name of the object...)
- Content (person, institution, event, topic, culture, time, location)
- Type (text, film (also TV), photography, audio (also radio), graphics, project, map...)
- (ethnographical) object, calendar, computer media...

A further challenge has been to ensure that each data set was earmarked by a common identifier which has to be migrated in the case of a change in the software system, for example, to ensure that data sets could also be clearly identified in the future and are citeable by future researchers.
This sounds easy. However, as shown in table 1 and followed by examples of title records by IDS Basel and BAB, it is far from trivial.

| Core Elements | UB / IDS Basel (Aleph, RDA/Marc) | BAB Library (Faust) | BAB Archive | MdK ethnographic and Photo Collection |
|---------------|----------------------------------|--------------------|-------------|--------------------------------------|
| Person, Institution | Author Editor | Author Editor | Author Editor | Photographer Collector Vendor Manufacturer |
| Title, Description | Title | Title | Title / Description Collection | Object (3D) Description |
| Geographic Term | Place of Publication | Place of Publication | Place of Publication Recording Location | Recording Location Place of Manufacture |
| Chronological Term | Publishing Year | Publishing Year | Publishing Year Recording Year | Recording Year Collection Year Year of Manufacture |
| Content | Person Corporate Body Event Uniform Title Topic Geographic Terms Chronologic Terms Genre |
| Type | Carrier Type |

Table 1: Core elements

As indicated in the table, the core elements common in all catalogues are not precisely circumscribed and thus heterogeneous. On one hand, this is due to the variety of materials used (text, photos, ethnographical objects, maps etc.) and the different requirements for a title record. However, on the other hand it is also due to the requirements of the various indexing systems which – with the exception of the catalogue IDS Basel – do not follow international standards and binding rulebooks. All core elements can have functions in terms of formality and content: A person can be an author, photograph, collector, or composer, but can also be the subject of a book, article, picture or photograph.
Figure 2: Catalogue record UB Basel (Aleph, RDA, Marc 21): With the help of Subject Category Code (072 $a af $2 SzZUIDS BS/BE) and local Codes (909A $f afbs-mono) the Africa collection in IDS Basel can be selected.
1. Periodical articles

| Author  | Betha, Jenny |
|---------|--------------|
| Title   | Developing an Understanding of Problems Being Experienced by Traditional Healers Living on the Western Border of the Kruger National Park: Foundations for an Integrated Conservation and Development Programme |
| Periodical or miscellanea: | Development: Southern Africa |
| Vol.: 15 | |
| No.: 4  | |
| Place of publication: | Halfway House |
| Publisher: | Development Bank of Southern Africa (DBSA) |
| Date: 1998 | |
| Physical description: | pp. 621 - 634, map, tables |
| Shelfmark: | P 544 |
| Genre: | Journals (Form) |
| Subjects: | Gathering; Hunt; Nature conservation; Traditional medicine |
| Geographical area: | Kruger National Park, South Africa |

2. Essays/books

| Author  | Labuschagne, R.J. |
|---------|-------------------|
| Title   | The Kruger Park and other National Parks |
| Place of publication: | Johannesburg |
| Publisher: | Da Gama |
| Date: 0000 | |
| Physical description: | 160p., ill., map |
| Language: | English |
| Shelfmark: | 15921 |
| Genre: | Books of plates |
| Subjects: | National parks |
| Geographical area: | Botswana, South Africa |

3. Map collection

| Title  | Durban; Leuanço Marques; Kruger National Park (Numbi) |
|--------|-----------------------------------------------------|
| Size (h x w) in mm: | 485 x 236 |
| Place of publication: | Johannesburg |
| Publisher: | Automobile Association of South Africa |
| Date: 0000 | |
| Series: | RR-11 |
| Language: | Afrikaans; English |
| Shelfmark: | L.A. 634 |
| Genre: | (Roadmaps; City Maps) |
| Geographical area: | Durban; Kruger National Park; Maputo City; Mozambique; Natal (historic Name) |

4. Map collection

| Title  | Kruger National Park and Lowveld |
|--------|----------------------------------|
| Size (h x w) in mm: | 800 x 525 |
| Coordinates: | E 30°00'–E 32°10'S 22°00'–S 25°00' |
| Scale: | 1 : 450 000 |
| Place of publication: | Rheinau [etc.] |
| Publisher: | Map Studio |
| Date: 0000 | |
| Edition: | 3. Auflage |
| Language: | English |
| Shelfmark: | L.B. 25 |

Figure 3: Catalogue record BAB library (Faust)
Figure 4: Catalogue record BAB archive (Faust)

Figure 5: Catalogue record Museum der Kulturen (here an example for an object from Australia)
The examples show that, despite a definition of core elements, their allocation is anything but trivial. It must also be assumed that, especially where there are no binding standards or authority files, the cataloguing staff have handled it differently over the years. The visualisation tool, Kibana, helps when it comes to revealing such inaccuracies.

Therefore, a person’s name can not only have various spellings and represent someone different depending on the catalogue, but even within the same catalogue. As an example, below is a list of the different forms in which the name of missionary Fritz Ramseyer appears:

- Ramseyer Fritz
- Ramseyer Friedrich August
- Ramseyer Friedrich August Louis

Authority files with persistent identifiers (here: GND-numbers) are the precondition even more if enrichments should come from the Semantic Web.

The same is true in geographical terms because the same names for locations can appear in different regions and countries around the world. In order to provide a unique identification, they must contain geocodes. It is not possible due to lacking capacities and resources to attach (where necessary) authority files to the names of people and geographical locations by hand, especially because African names often lack authority files. Automatic mapping is a possible option when it can be verified whether the number of errors is still within the tolerated limit.
5 BARTOC and the visualisation of BAB subject headings

The Basel Register of Thesauri, Ontologies & Classifications (BARTOC\textsuperscript{1}) is the most comprehensive terminology registry for Knowledge Organization Systems such as thesauri, classification schemes, subject heading schemes etc. It contains the option to assign URIs for lists of subject headings so these can be clearly identified. Therefore, even subject indexing systems which have no authority file attached can be provided for the Semantic Web.

After modelling the BAB thesaurus in SKOS\textsuperscript{2}, it was uploaded to BARTOC’s RDF triple store, which returns a browsable visualisation.

---

\textbf{Thesaurus der Basler Afrika Bibliographien}

![Visualisation of BAB thesaurus in BARTOC](https://bartoc.org/en/node/2006/visual)

---

\textsuperscript{1}https://bartoc.org  
\textsuperscript{2}https://www.w3.org/2004/02/skos/
6 Kibana

Kibana, a visual tool, can make mapped data visible and makes it easier to discover inconsistencies. There are at least hints for headings which are possibly not correct. This has an added advantage for the participants since it offers them a possibility to analyse their own catalogues and to quickly detect problems in mapping and/or indexing.

7 Data processing for the Research Data Viewer

The data must be available in a machine-readable form, e.g. a CSV-file.
The files will be processed as follows:

Step 1: Normalisation

- Each file undergoes a mapping with the core elements via GoogleDocs Spreadsheet, which means that they are transformed into a defined standard format and ingested into an elasticsearch-index.
• Kibana visualises data. This is followed by an examination and – if necessary – adjustments to the original data or the Google Spreadsheets Mapping and is therefore a repeat of step 1.
• Developing of a prototype of the "Research Data Viewer" which hopefully can easily be adjusted to further technological advancements.

Step 2: Enrichment

• The core elements are normalised as far as possible in the standard format (e.g. time values, country names). The geographical terms are enriched with data from Geonames in order to match them with a geocode. However, because in most cases only the location names are available, automatic enrichment must be supervised in order that no wrong matchings are reported. A viable compromise must be found between recall and precision. How many location names can be enriched and how high the feasibility is for a correct enrichment remains to be seen. Location names can then be linked for instance to Wikipedia, as well as persons and subjects if they are unambiguously identified.

Step 3: Setting up an annotation tool

• A research platform must also enable the exchange among researchers as well as between data, pictures, objects and researchers. The annotation tool aims to encourage researchers, as well as all those who know Africa, to present their knowledge to the public. In many cases, only a few details are known e.g. to describe photos. Hints to persons, situations, and locations are most welcome.

8 Consolidation

There is a plan to install a beta-version of the "Data Research Viewer" in the first quarter of 2018, in order that data can be searched for and retrieved.

This would be the moment to bring the project into daily life. All participating institutes must have an agreement to keep the portal running.

At present, "Basler Africa Portal" is only a temporary working title. A more fitting name with a more appropriate acronym is yet to be found.

9 The vision for the future

As mentioned above, the project aims to incorporate research data about Africa in order to bring together knowledge about Africa which has been gathered by people in Switzerland, Europe, and all over the world and make it accessible through one integrated search.

We focus especially on the southern part of the continent as well as Tanzania, where the five founding institutes mentioned in this text have close partners with whom they collaborate and do research. It is an aim to include research carried out by our research colleagues in Africa in order that their scientific results become more visible and accessible in the Western world.
Viewpoints developed by African researchers can bring new and different input into Western research. At the same time, African researchers also receive knowledge and in many cases access to research about Africa which has been conducted in Basel and elsewhere.

Admittedly, the expansion is a very ambitious project. The realization isn’t ensured yet and depends above all on the financial and personnel capacities of the Basel University Library. Nevertheless, there are already contacts to interested parties for future collaborations.