

Christoph Rauch, Staatsbibliothek zu Berlin

## Perspectives of Arabic-script cataloguing in German libraries

The information and documentation system is contemporarily in the middle of significant changes. German libraries and libraries in the whole world have become part of a global information network. This bears on the one hand the possibility for universal bibliographical control and on the other hand the possibility for offering library-services worldwide. The reader in a library of the 21st century expects access to books, databases, and media in any shape and language via one single access.

The „homogeneous retrieval of heterogeneous data is the manifested objective of the German scientific library-service.” This sounds quite simple, but in fact, it is not easy to realize. There are some problems one has to face and solve.

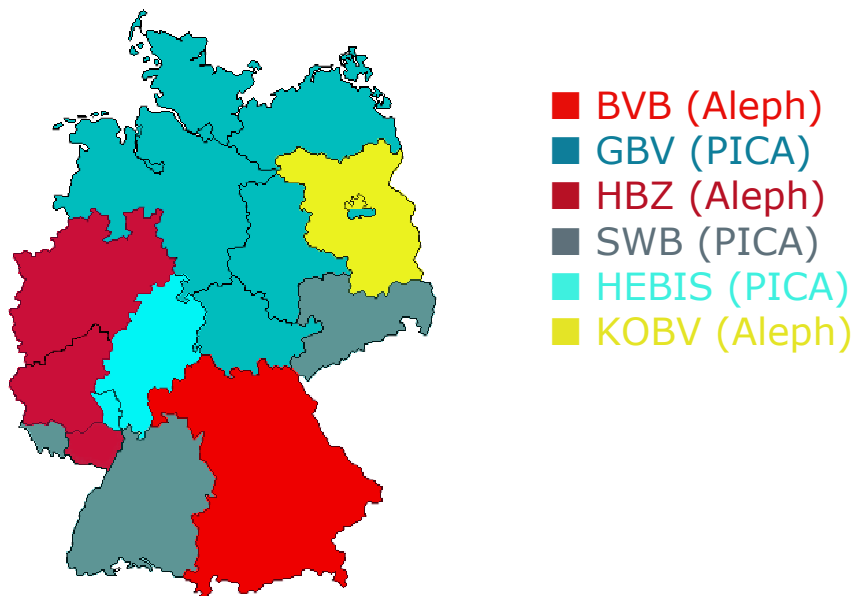
One question is the presenting of bibliographical data for Arabic-script materials in an automated library-environment.

In this paper I intend to give an overview of the situation and of the perspectives of cataloguing books in non-Latin script in German libraries. Doing that I focus on the collection at Berlin State library, where I started to work some months ago. To say it before, a multiscript-cataloguing is at the moment not practised in our library, but at least it is finally within the realms of possibility. The need of the complementary cataloguing in the original script has often been ascertained in the last years: For the Arabic literature in our catalogues we can argue: as long as different rules of transliteration – without integrating the original script – exist together, the retrieval of all literature from one point of access is not possible.

Saying this I have to admit that I do think that in general different rules of transliteration should be kept because they are inseparably part of the scholarly tradition of a country. And, not only that – more important they are connected phonetically to the respective language in which one transliterates. Let me give you an example: for instance some different pronounced consonants in the Spanish language. Here a transliteration following the Library of Congress norms does not seem to be suitable at all. The original script therefore should be the connecting link here.

First of all, before I am going to talk about different activities and problems, I want to give you an – unfortunately only tiny – idea of the specifics of the German scientific library-system.

Scientific libraries in Germany are – parallel to the federal political system – divided into six library networks, which – as a result of co-operative cataloguing – present their data via union catalogues. Using these union catalogues it is possible to search the data of many scientific libraries all at the same time.



(Library-Networks in Germany)

The coexistence of different systems implicates advantages like an innovative ability, but it might also prevent from a fast implementation of innovation for the entire country because the resources are divided.

Among the difficulties we are facing is the application of different cataloguing formats in German library networks. The German National Library (the former The German Library), which stands beyond the mentioned library-networks will use MARC21 as an interchange format beginning 2008 and therefore will stop developing its concordance to the German format MAB2.

While the Aleph-based networks will continue to use MAB2, the National Library and the PICA-based networks will keep on using their own MARC-based format. Beside this it is not clear when the new cataloguing rules RDA will be introduced. This was originally planned for 2007. What I want to say – and this is a result of my experience with the Common library network – is that the library-networks in Germany in general have completely different problems and therefore will not invest much effort in the concerns of small and scattered special-collection-libraries.

Before I want to discuss the efforts of the library-networks towards multiscript-cataloguing, I want to speak about the situation at the Berlin State Library.

The Oriental Department of the Berlin State Library is primarily known for its large collection of manuscripts originating from the Middle East, Africa, Central and South-East-Asia. But the Oriental Department is also in charge of a large amount of printed material. The library holds approximately 600.000 monographs and an extensive collection of periodicals in European and Asian languages related to the Middle East, South-Asia and Africa.

The Oriental department was founded in 1919 as a separate department of the Prussian State Library. It took its present structure in 1992, when the Department of Africa and Asia of the “Deutsche Staatsbibliothek” of the former East Germany and the Oriental Department of the West-Berlin- State-Library were united.

Since the foundation of the Department in 1919, books in non-Latin languages have been catalogued in separate catalogues. The reason was not primarily the different alphabets – the data was entered in a transliterated form anyway. No, the reason was the different name-heading of the authors in different languages according to the cataloguing rules.

As much as the recent history of the library was affected by world-war II and the political situation afterwards, as much the holdings of the library were affected, too. A significant part of the books has been lost in the war, another part returned to the library and the rest became part of the latter founded West-Berlin State-Library.

A large amount of the old stock is already available in the Online-catalogue, because the West-Berlin-library decided after the war to put all data together in one catalogue. In contrast to that, the Non-Latin Books acquired by the State Library in East-Berlin,

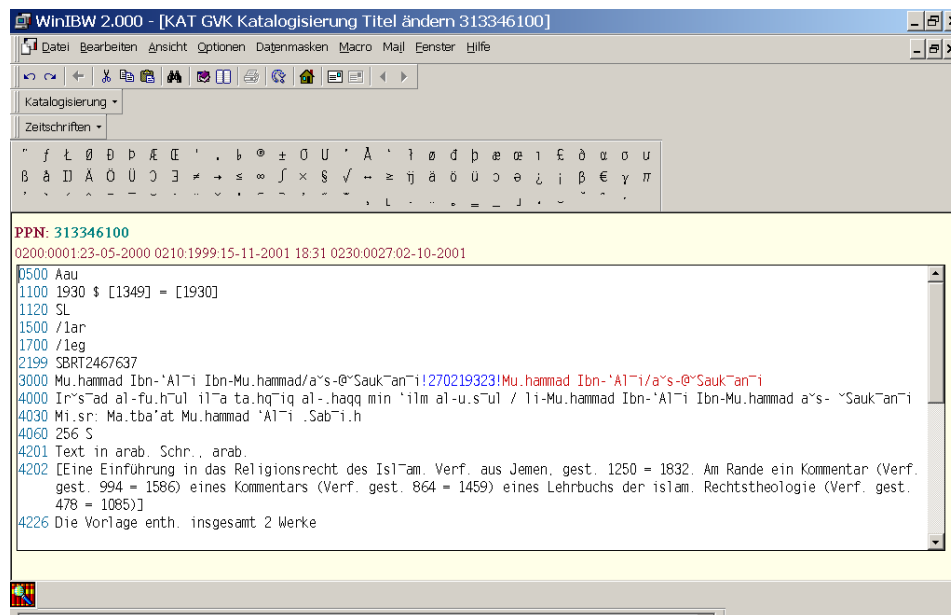
continued to be catalogued in separate card catalogues until 1992. These catalogues are still waiting for retrospective conversion.

The Online-Catalogue of the Berlin State Library is part of the Common Library Network (GBV). This catalogue, respectively the cataloguer's database, now contains material in three different rules of transliteration:

1. transliteration according to the *Prussian Cataloguing Rules* from 1909, which was used until 1985 (for example: Sajjid)
2. transliteration according to *DIN-standard 31635*, which was used from 1985 on (for example: Saiyid)
3. transliteration according to *LC Romanization-Tables*, for records retrieved from external databases (for example: Sayyid).

This situation implicates difficulties not only for the user but also for the cataloguer. At the moment the acquisition of a new title is to be conducted, the subject specialist feeds the catalogue with a new data record in the union-catalogue, if there is not already one in the system. The cataloguer is advised to use any available record for the title, so he has to check out twice (if there is no ISBN), within the German and Anglo-American transliteration system, to be on the safe side. If he finds a LoC-record he has to attend the local data and modify the transliteration into that one used in Germany. This workflow, in my opinion, holds the risk to enter errors in the record more then creating a complete new record, and furthermore takes more time.

The illustration here shown might give an impression, how time-expensive the input and correcting of the record is, - each diacritical is to enter via separate mouse-click.



(Cataloguing terminal WinIBW 2.000)

Considering the readers-perspective, so of course it is not an impertinence to expect from a visitor at a special-collection-library to have some efforts in finding a title and consult a card catalogue in addition to the OPAC. Moreover the librarian can give the library-visitor advices, personally or via the homepage, on how to retrieve arabic-script titles in the OPAC. But all this is not possible at the level of an union-catalogue. The most commonly used union catalogue, the Karlsruher Virtual Catalogue, gives access to all the important library-networks in the world. But in reality, doing research in this catalogue, even the Arabic holdings of the city of Berlin are not easy to find, because they are transliterated according to different rules. The holdings of the *Centre for Modern Oriental Studies* in Berlin, for example, are transliterated according to the Encyclopaedia of Islam, 2<sup>nd</sup> edition. Meanwhile the holdings of Berlin State Library are translated according to Prussian Instructions and DIN.

The implementation of Unicode worldwide in computer operating systems and in integrated library-systems paved the way for representing any script and to do the cataloguing using Unicode. The American library-networks, as well as other libraries in the world, are already cataloguing in original script and in transliterated script for

years. Usually they use MARC21-category 880 „Alternate Graphic Representation“. This repeatable field is connected with the standard-fields and allows retrieval in original and in latin script as well.

The last report of the „Task Force on Non-English Access“ of the Association for Library Collections and Technical Services, from April 2007, gave a very good impression of the current state of the art in the United States.

Why should this way of cataloguing, which is - as I mentioned - now common practice at American and other libraries worldwide, not also be possible in German libraries? All big network-systems are by the end of last year capable of Unicode. And both the ALEPH and PICA-based networks started some projects toward multiscript-cataloguing. Pioneering tasks have the libraries with CJK (Chinese, Japanese and Korean) materials, but also for Hebrew materials the first steps are done.

It was already mentioned, that in the German library-system different formats are used. This avoids a unique solution concerning the problem of multiple script cataloguing. A first step to approach the problem has been done by the German committee for standardization when it legislated a new field for the MAB2-format to be used for multiscripts in Unicode in 2004. This field is fairly parallel to the 880-field for *multiscript*-cataloguing used in MARC21 in its function and structure. It is called field 671 „multiscript-description“ and it is repeatable and linked with the main-fields in latin script. I'm not completely sure but I don't know even one library which is using this suggested model.

PICA and ALEPH networks are still developing their own models for multiscript-cataloguing and are not done yet in doing that. Here I want to discuss different models of two library-networks. Due to the different states of development – the projects are still in their beginnings –it is not possible to compare those two different models substantially. Therefore my presentation aims to give only a general overview.

PICA's cataloguing system WinIBW (Windows Intelligent Workstation) is since the end of last year capable of Unicode-representation. It is principally possible to

integrate multiscript, but most of the libraries have local systems where one has to enter the Arabic letters in a coded form. The new version, which allows entering the Arabic script for example directly via Windows script-control, is not working yet. Cataloguing of Arabic script is only theoretically possible, but there are some pilot-projects for other languages which I want to introduce to you briefly:

In the Netherlands OCLC-PICA developed a *prototype-transliteration-module* for automatic transliteration from original script to latin script. For the testing phase the Greek language was chosen. The main-field, for example „author“ or „title“, is repeatable. In the first field the title is entered in Greek language whereas in the second, the repeated field, the code for latin-script has to be entered only. After that the librarian can enter the button “transliteration” and the transliteration will automatically appear. This automatic transliteration is based on the script. For the Arabic language, I do not think it will be that easy to apply. Arabic script is simply not possible to converse from sign to sign, because of the lacking of the vowels.

Automatic transliteration would require the use of thesauri. The project I know about, is still in the pilot phase and it is not applied in the practice yet.

The first practical application of multiscript-cataloguing in a PICA-based network in Germany, was started in a project cataloguing old stocks of Hebrew books at the Herzog-August-library at Wolfenbüttel in Germany. This library is especially famous for its collection of 17<sup>th</sup> century books. The mentioned project was realized in cooperation with the Bodleian library in Oxford. However the data are not yet transferred to the union-catalogue of the Common library network. At least they are available in a special database on the library's homepage.

As the Common library network does not transfer any multiscript data to the union catalogue, we have to assume that the multiscript-cataloguing is still in the starting blocks here. This is a fact I am very sorry about since part of this network are many libraries with huge collections of Asian books. What I do want to mention are the Special Collection for the Middle East in Halle, a city south of Berlin, and the Special Collection for East and Southeast-Asia at the Berlin State Library. The East-Asia holdings in Berlin are catalogued in original-script and are kept in a separate database and excluded from the union-catalogue. The arabic-script periodical content database

of the Middle East Virtual Library in Halle is also still excluded from the union-catalogue whereas at least the books and periodicals are catalogued in a transliterated way.

#### Multiscript-projects in ALEPH-contexts:

The „hbz“ (university-library-service-center) in Düsseldorf in Western Germany, close to Cologne, is the head of the union-catalogue for scientific libraries in the federal state of Northrhine-Westfalia and parts of Rhineland-Palatine. In this library-network the ALEPH-system is used. This system is much more able to deal with multiscript records. One of the reasons ALEPH was developed in 1980 in Israel was that at this time there was no other system available which could cope with Arabic and Hebrew materials all at once. The ALEPH-integrated system is also used in a number of libraries in Bavaria and Berlin. The university-library-service-center in Düsseldorf offers multiscript-cataloguing for its customer-libraries. Currently, four libraries with materials in Chinese, Japanese and Korean are cataloguing multiscript-records into the union-catalogue. The service-center does not use the – for this purpose developed – field “671” in the MAB2-format. But it developed its own standard for multiscript presentation, based on MAB-format. At the moment it is the only practice-proved and tested model that has good possibilities for exchange with external data bases. The Bavarian Library-Network (BVB) recently decided to assume the model, which is currently only used for east-asian-materials.

Many libraries in Germany possess large historical collections of oriental books and periodicals. The catalogues of these highly specialised holdings are not only useful for the local readers, but rather important as bibliographical databases for scholars worldwide, which need to be better represented in Online-catalogues. The basics for the development in this direction are done: And I do hope that in some years multiscript cataloguing will be normality in German libraries.

Therefore I am hopeful that interloan-tickets, as can be seen here on the next picture, will be part of past times. At least I never saw catalogues in Arabic countries using a transliteration for titles written in European languages...



Best-Nr. 003101617	Bestellnummer 188 003101617	Friststempel	Friststempel 188
Aus: Zahl d. Bände	Bibliographischer Nachweis KOBV 00 (KOBV) AK:00 SK:00 <i>Klrot (Arbeits)</i>	Signatur	Signatur
Name und Anschrift des Entleihers Berlin Freie Univ. Nutzer	Bücher: Name, Vorname d. Verf., Titel - Unterit., Aufl., Ersch.-ort u. Jahr, ggf. Serie u. Bd., ISBN Zellschriften: Titel (ungekürzt), Bd., Jahr/Ort Rai, zii, al-Sifan>ai, ni, , Ahfmad ibn >Abd Allai, h, d. ca. 1068.: Kitai, b Tal, ril, kh madii, nat Sifan>ai, >M / allafahu Ahfmad ibn >Abd Allai, h al-Rai, zii, al-Sifan>ai, ni, ; >uniya bi-tahifqi, qih wa-wadif> fahai, risahu. Sifan>ai, >M : . . . Bei Aufsätzen: Verf., Titel, Seiten:	Benutzer wünscht: -- Lieferung nur bis 14.10.2007 Datum -- Auch andere Auflage Nein <input type="checkbox"/> Übersetzung Nein <input type="checkbox"/> -- Kopie falls Original nicht lieferbar Nein <input type="checkbox"/> Falls Kosten anfallen, bis max. Euro > ..... einverstanden	Zahl d. Bände
Benutzernummer 3677953			
188 Dieser Abschnitt muss bis zur Rücksendung im Buch bleiben			
Berlin Freie Univ. Garystr. 39, 14195 Berlin	Zahl d. Bände Berlin Freie Univ. Garystr. 39, 14195 Berlin	Entleiher: Name, Vorname: Berlin Freie Univ. Nutzer	Bestelldatum und Unterschrift des Sachbearbeiters 18.04.2007
			Berlin Freie Univ. Garystr. 39, 14195 Berlin

Web-adresses of some mentioned institutions:

- ➔ Berlin State Library: [www.staatsbibliothek-berlin.de](http://www.staatsbibliothek-berlin.de)
- ➔ GBV (Common Library Network, Göttingen): [www.gbv.de](http://www.gbv.de)
- ➔ Hebraica-database Wolfenbüttel: <http://dbs.hab.de/hebraica/index.php>
- ➔ HBZ (University Library Service Center, Düsseldorf): [www.hbz-nrw.de](http://www.hbz-nrw.de)
- ➔ Middle East Virtual Library (Halle): [www.menalib.de](http://www.menalib.de)