

## Workshop:

# "Digitalisierung chinesischer, tibetischer, syrischer und Sanskrit-Texte der Berliner Turfansammlung"

Berlin, 02.-03.06.2005

Veranstaltungsort: Berlin-Brandenburgische Akademie der Wissenschaften

Jägerstrasse 22-23 (Gendarmenmarkt)

10117 Berlin

Program m | [Fotogalerie](#)

### **Donnerstag, 02.06.2005**

*Ort: Einstein-Saal, 5. OG*

10:00 Uhr Eröffnung durch Prof. Dr. Christoph MARKSCHIES, Sekretar der Geisteswissenschaftlichen Klasse der BBAW und Vorsitzender der Kommission Turfanforschung

10:15 Uhr Peter ZIEME: Cooperative Studies between Japan and Berlin

10:45 Uhr Susan WHITFIELD: The International Dunhuang Project: Collaborating to Conserve, Digitise, Catalogue and Research Silk Road Manuscripts, Paintings and Artefacts

11:30 Uhr Hartmut-Ortwin FEISTEL: Union Catalogue of Oriental Manuscripts in German Collections

12:00 - 13:30 Uhr *Mittagspause*

13:30 Uhr [Kati BRAUCHMANN: Digitisation of the Turfan collection: Review of the first 6 years](#)

14:00 Uhr [Klaus Wille: Survey of the Sanskrit Manuscripts in the Turfan Collection](#)

14:30 Uhr [MITANI MAZUMI: NEW RESULTS OF RESEARCH ON THE CHINESE BUDDHIST TEXTS OF THE BERLIN TURFAN COLLECTION](#)

15:00 - 15:30 Uhr *Pause*

15:30 Uhr [Tsuneki NISHIWAKI: Zum Fomu-\*jing\* 佛母經 \("Sutra der Mutter des Buddha"\) und zu Problemen der Katalogisierung der chinesischen Fragmente vermischten Inhalts](#)

16:00 Uhr [Tatsushi TAMAI: Paläographische und linguistische Untersuchungen und C14-Datierung](#)

16:30 Uhr Gerd GROPP, Neues zur Arbeit an den Fragmenten der Francke-Körper-Sammlung, München

### **Freitag, 03.06.2005**

Ort: Seminarzentrum Taubenschlag, 5. OG

10:00 Uhr Präsentationen:

Digitales Turfan Archiv (DTA)  
Kati BRAUCHMANN

IDP: A Manuscript Cataloguing, Conservation and Image Management Database and Multilingual Web Server  
Susan WHITFIELD - Michael KAYE

R u n d t i s c h g e s p r ä c h e

11:00 - 12:00 Uhr Themenschwerpunkt:

Probleme der Identifizierung und Klassifikation der chinesischen, tibetischen, syrischen und Sanskrit-Turfantexte

12:00 - 13:30 Uhr *Mittagspause*

13:30 - 15:00 Uhr Themenschwerpunkt:

- Restaurierung und Reproduktion der Turfanfragmente

- Probleme der Aufbereitung der Primärdigitalisate der Fragmente und ihrer Einbindung in die IDP-database

Gesprächspartner: Kolleginnen und Kollegen der Staatsbibliothek zu Berlin, Orientabteilung & Abteilung Bestandspflege und Reprographie

15:00 - 15:30 Uhr *Pause*

15:30 - 17:00 Uhr Themenschwerpunkt:

- Die deutsche Version der IDP Webseite

- Probleme des Datentransfers aus bestehenden Datenbanken in die IDP database

Gesprächspartner: Kolleginnen und Kollegen der EDV-Gruppe der BBAW und der Orientabteilung der Staatsbibliothek zu Berlin

## **Fotogalerie Workshop**

02.06.2005





03.06.2005





# Digitisation of the Berlin Turfan Collection

—

## Review of the first six years

*Vortrag anlässlich des Workshops "Digitalisierung der chinesischen, tibetischen, syrischen und Sanskrit-Texte der Berliner Turfansammlung" Berlin, 02.06.2005*

*Kati Brauchmann*

Today I would like to give a summary review of the first six years of digitisation.

Since I personally have taken part only in the last two years of this project I have to thank Ramin Shaghghi for his excellent documentation of the digitisation project. Thanks to this detailed documentation I was able to continue with the digitisation without much loss of time.

### Introduction

The Berlin Turfan Collection comprises over 40.000 fragments of Oriental manuscripts from several expeditions to the area around the Turfan oasis on the northern Silk Road. The Oriental Department of the State Library in Berlin is responsible for the administration and conservation of these fragments. The art objects and the illuminated or calligraphically outstanding manuscripts are kept in the Museum of Indian Art in Berlin-Dahlem.

### Digitisation and its aims

- *Preservation of the Content:* to save and protect the content of the manuscripts
- *Protection of the original manuscripts:* to reduce the work with the original fragments in order to avoid damage to the fragments
- *Facilitating international access:* to attend to the frequent requests from many researchers for reproductions of the originals; making the fragments available to a larger number of scholars and scientists may accelerate the publication of the as yet unpublished texts
- *Digital reconstruction:* to prevent damage the joining of fragments should be carried out only digitally in the future.

### Digitisation project "Digital Turfan Archive"

For all these reasons a project was created, the Digital Turfan Archive. The project, sponsored by the German Research Foundation DFG, has been running since 1997 and the present phase will be finished next month (July 2005).

### Organisation

The Oriental Department of the State Library supports the digitisation project by restoring the fragments as well as producing the colour slides.

The Digital Turfan Archive Project itself employs two people, one full-time and one half-time. The staff of the Berlin Turfan Research Group and the Turfan Group of the Union Catalogue of Oriental Manuscripts in German Collections undertake the management of the project and provide scientific support.

## **Characteristics**

Before I illustrate the actual process of digitisation I want to explain some of the decisions which had to be made before the implementation of this project.

Selection of material to be digitized:

The Turfan fragments are divided into groups of texts that can be classified together on the basis of the language in which they are written and for other specific reasons. The first group of texts to be digitized were those in the Tocharian language. In this case the digitisation was privately financed. As the Turfan Research Groups currently edit and catalogue the Old Turkish and Middle Iranian manuscripts, these groups of fragments within the Berlin Turfan Collection were selected for the first phase of the digitisation project sponsored by the German Research Foundation DFG.

Why "double preservation":

The choice of "double preservation" on analogue material, in this case slides, and digital media, was mainly the result of consultations with the TITUS-Project. Purely digital preservation carries some risks, since the life span of the new digital storage media has not yet been finally proven while high quality slides can reach a life span of more than a 100 years. For example, when we copied the TIFF-master images from the CD ROM-archive to new external hard discs last year we had to assess damage on some CD ROMs dating from the beginning of the digitisation project in 1997. Some files recorded on these CD ROMs were no longer useable, so we had to digitise these destroyed images once again because the TIFF-master images existed only on CD ROMs at the time. Slides enable us to re-digitise them at any time, maybe also with enhanced image scanners. Another argument against a purely digital preservation was the low-quality of digital cameras at the time. Therefore a double analogue-digital preservation provided the required high-quality images and guaranteed long-term storage.

Why the photography work was outsourced:

Because of the age of the material it is essential that they be photographed in a professional laboratory but with limited personnel and financial means there was no possibility of setting up the technical equipment inhouse. Therefore the project is undertaken with the

cooperation of the State Library Berlin, where professional pictures are taken in the form of slides.

## Digitisation process

- *Preparation:* The fragments have to be checked and prepared for transport to the Reprography Department of the State Library. Also labels containing the fragment's signature sometimes have to be renewed.
- *Restoration:* In cases of badly glazed fragments, fragments which are soiled by Turfan dust or crumpled fragments, restoration is necessary. In that case we send them first to the Restoration Department of the State Library in Berlin. This department has achieved excellent results regarding the restoration of the manuscripts. For instance, two sheets of one single fragment which had been stuck together were separated. A different script appeared on the inside of the two sheets which is sensational.
- *Making Photographs:* The Reprography Department of the State Library Berlin produces colour slides of the originals.
- *Scanning:* The colour slides are then brought to the Digitisation Project where they first have to be identified and marked with the signature of the fragment in accordance with its signature in the administrative database. The signature is the basis for the file naming system in the Digital Turfan Archive. Now the slides are scanned with a colour slide scanner with a resolution of 2700 dpi.
- *Image editing:* Some scans have to be rotated in order to bring them to the right position in accordance with the direction of the script. Also it is necessary to enhance the lightness and contrast of the digital images. These operations are easily done with the help of the colour slide scanner software.
- *File saving:* Now the digital image has to be stored as an image file with an appropriate name. It was decided to use long file names according to the signature of the fragment. If the sides of the fragment have been identified as recto and verso, "recto" and "verso" is added to the file name. Otherwise the file name contains the neutral additions "side 1" or "side 2". The word "detail" is added if a fragment has extra detail images. These master image files are saved in uncompressed TIFF format.
- *File storing:* The cheapest and simplest storage for such large files (20-30 MB for each TIFF image) was to burn them on CD-R's at the time (1997) when the project first started. For security reasons two identical series of CDs were created and stored in two different places. Because of doubts about CDs as a medium for long-term storage we decided last year in addition to store all the master images also on two external hard discs. These two hard discs are stored at two different storage locations.

- *Quality management:* The progress of digitisation is noted in an administrative database where signature, slide number, CD number and also other details are documented. For example, we note if a fragment is kept in the Museum of Indian Art or not in its place for other reasons. In such cases we provide a scanned black and white photograph of the fragment if one exists.
- *File compression:* The TIFF master images are too large for daily use and for use on the Web the file size has to be reduced. For daily use the master image files are converted to the compressed JPEG format and stored separately. For use on the Web the image files have to be much smaller in size because of the need to reduce the loading time for the images. Therefore we first reduce the resolution of the master image. Depending upon the dimensions of the image and the script size, the resolution is reduced then to a half (1350 dpi) or a quarter (675 dpi) of the TIFF image. These files are stored separately as JPEGs. These operations are easily done by batch conversion tools provided by image editing programmes.
- *Presentation in the Internet:* Now the low-sized JPEGs can be used for presentation in the Internet. The text groups are divided into categories according to their content, language and script. In our Website the user is first guided to these categories. After selecting a category the user is linked to a table with a numerical order. By clicking on a page with the number he is looking for the user is linked to a thumbnail review including the signature. The last step is just to click on the thumbnail image in order to view the larger image of the manuscript fragment. These images are intended simply to give an overview of the Turfan collection. The user can contact us anytime to get the high-resolution TIFF-images of the fragments for working on and publishing.

Due to the way the Digital Turfan Archive is organised the user has to know the signature of the fragment he is searching for. For lack of financial means, staff and catalogues the Digital Turfan Archive provides the digital images of the fragments only, but no meta-information concerning the content, publication etc. To get this information, the user has to consult the editions, published concordances or catalogues, if available.

## **User Guides**

At the third stage of digitisation we were able to improve the usability of the Digital Turfan Archive.

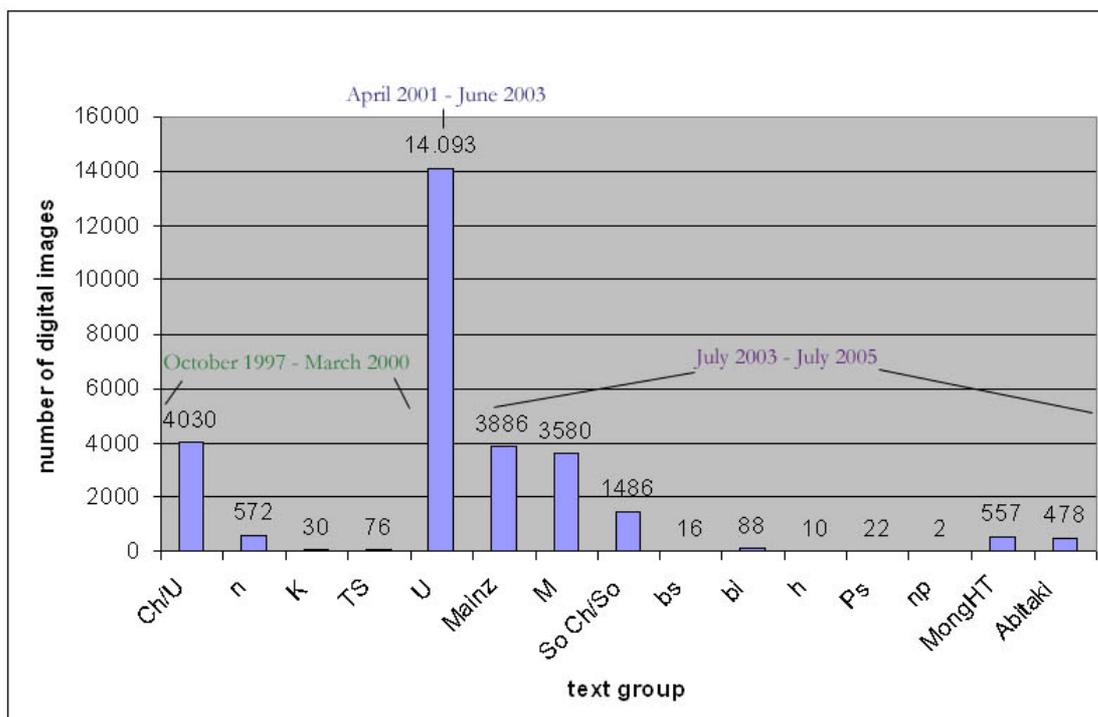
- We redesigned the home page for a better overview and faster access.
- We prepared a detailed information site about the project with up-to-date facts about the status quo of digitisation. The user can browse the text-groups directly from this site.

- The text groups M and So include plates where many fragments are combined under one glass plate. We have digitally labelled each single fragment in these collective plates so that users do not have to consult the catalogues or contact the Turfan Research Group to find a particular signature
- For the latest digitised text groups we have prepared a supplementary information site with the following remarks:
  - Information about further reading: catalogues, editions etc.
  - Information about different formats: a great number of fragments were placed under glass; but formats such as scrolls, bound booklets, folded books, blockprints etc. are mostly not protected by glass.
  - Information about location: some fragments are located at the State Library Berlin or the Museum of Indian Art.
  - Information about joined fragments or different fragments combined under one glass plate.
  - Information about catalogue numbers.

We would like to prepare such information sites for the text groups already digitised, too. For example, part of the Uighur text group was catalogued after digitisation so that we will be able to add secondary information about these fragments. However, the digitisation of the text groups as yet not digitised has priority.

## Status quo of Digitisation

Up to the present there have been three successive stages of digitisation:



All of these text groups are already digitised and at the end of this month all 14,480 fragments will be available on the Website of the Digital Turfan Archive.

### **Future perspectives**

We would like to go ahead with the digitisation of the Chinese and Sanskrit fragments. At the beginning of this year the Turfanforschung has applied to the German Research Foundation for financial support and is at present waiting for the decision. For this new project the staff of the Berlin research team will to collaborate with the IDP project to use the advantages of the IDP database to make these texts available to as wide a community of scholars as possible.

#### **Quellen:**

Ramin Shaghghi : Digitization of the Berlin Turfan Collection. In: Turfan Revisited. The first century of research into the arts and cultures of the Silk Roads. Berlin: Dietrich Reimer 2004. (Monographien zur indischen Archäologie, Kunst und Philologie. 17.), p. 318-320