H. Braam, M. Vandamme

A ROBUST AND VERSATILE SOLUTION FOR THE DIGITAL PUBLICATION OF MANUSCRIPT MATERIALS

I. Introduction

The rich text tradition of Rabghūzī's "Stories of the Prophets" is of great importance to turkology and islamology. Besides comprising very old and unique language monuments, the corpus represents also a unique cross-section of the development of the Chaghatay language and religious culture from the 14th century up to the end of the 19th century. The London MS of Rabghūzī's "Stories" was published in transcription with a translation, glossary, and indices in 1995 [1]. The RMC (The Rabghūzī Manuscript Corpus) project is the natural complement to this edition by taking into account all other existing major MSS. Efficiency reasons force us, however, to publish the materials in a digitalized photographical way on CD-ROMs and not in the form of a critical edition like the London MS.

This corpus will greatly enhance the study of the linguistic developments of the 14th century pre-Chaghatay dialects all through the Classic Chaghatay period up to the modern languages of Central Asia at the beginning of the 20th century.

It is of no less importance that the corpus will present a rich source for the study of the development of Islam in Central Asia from the 14th up to the 20th century.

II. Criteria

As the solution should be open to future developments (at present information technology is changing very fast) and should operate not only locally but also in wide areas, it is important to draw up first a list of criteria to be met:

1. Robustness (usage, media-conservation, distribution)
2. Openness (technology, contents, organisation)
3. Scalability
4. Standardisation
5. Versatility (usage, security)

1. *Robustness*

The media used should be simple and cheap to conserve; the production of the CDs must be simple, without need of extra programming or engineering.

2. *Openness*

a) Concerning Technology: solutions must be platform independent (*i. e.* run without problems on DOS-Windows, Macintosh and Unix systems); solutions must be easy to adapt to future technological developments.

b) Concerning Contents: solutions must support both addition of new data and adaptation to new standards.

3. *Scalability*

Solutions must support the growing process of turning a small project into a larger one.

4. *Standardisation*

This issue concerns the central problem of creating information systems. Information cannot be easily retrieved or shared if it is not represented when using a set of common standards.

There exist industrial standard compression formats (GIF, JPEG) which are being used now. The software solution must accept such standards.

5. *Versatility*

Various access strategies must be possible when using the *same* software; use of the software should be easy, without need of a printed manual, as well as of difficult codes. It must be possible to secure the information by means of passwords in order to discourage illegal copying [2].

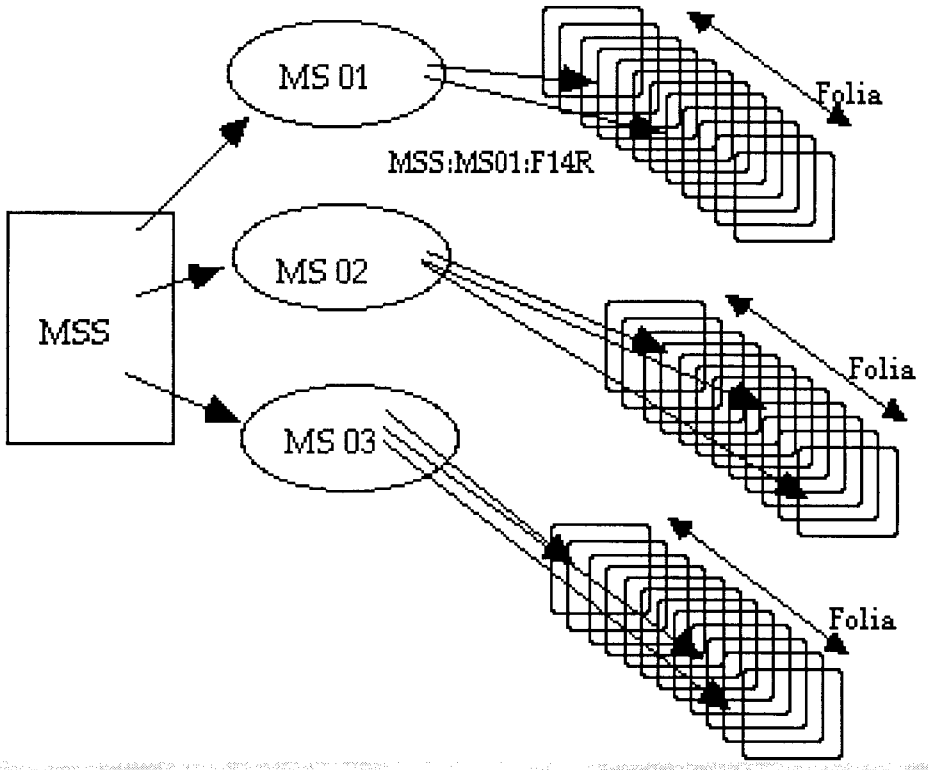


Fig. 1

III. Project Analysis

A) *Information and access structures*

We deal with a number of MSS of the same text. The texttree can be considered as the most basic representation of the MSS on which all other text descriptions depend. The hierarchy of the texttree corresponds to the hierarchy of the written textual materials of the MSS (Text, MS, Leaf, Line, Word, Character), not to the genetic stemmatological hierarchy.

A net can cut up a text in portions on all different levels of the texttree — we can impose a net on the level of text-lines (collation of poetry), on the level of leaves (different copyists), or on the narrative level. Every net is essentially a paradigm imposed on a text [3].

Our text consists of a series of cycles about the pre-Islamic prophets (like Adam, Solomon, Moses, and Jesus). These cycles are make up of the autonomous stories. Each story can be divided into substories.

So, the solution sought for must support at least the same representation of the texttree and must allow the application of different networks superposed on the texttree (*fig. 1*).

B) *Requested functionality of the system*

To put all MSS data on a disk, when using CD-ROM techniques, we need tools for the next functions:

1) creation of digital images of the MSS

2) digital manipulation of the images, including compression

3) graphical screen presentation of the images, including digital magnification

4) complex linking of multimedia materials (text, images, sound, animation) [4], according to the texttree and the networks

5) easy navigation by the users.

IV. Solution

For the creation of digital images of the MSS and their subsequent manipulation any industrial standard software can be used [5].

Although a CD-ROM can contain a lot of data, there is a limit: in these days about 600 Mbytes. The size of an uncompressed image file is a function of: the size of the image; the resolution (the number of pixels caught in an inch of the original); the colour quality. High resolution high quality colour images can be very large, tens of Mbytes.

So in order to maximise the number of images on a single CD some decisions are to be made. And here we face a few trade-offs:

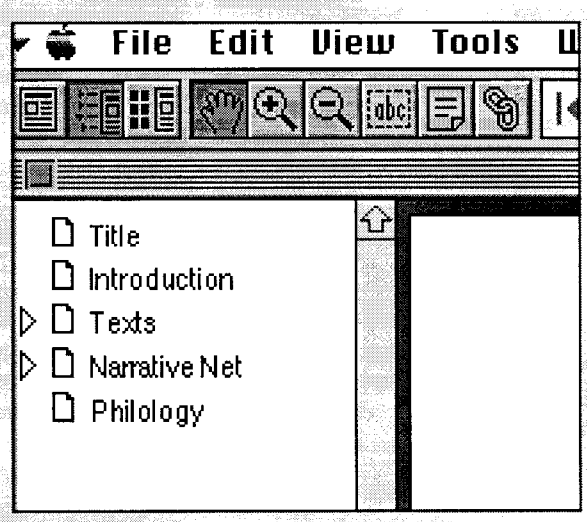


Fig. 2

— high quality colour takes much more space than 256 grey scale images;

— low resolution (resulting in small files) leads to unclear images or images that cannot be enlarged to become readable;

— very high compression may lead to a loss of data (resulting in unusable images).

After some experimenting we came to the conclusion that in this case an optimal solution is reached by scan in 256 grey scale images. The material concerned is not illuminated at all. Some parts of the text is written in colour. These small parts can be scanned in full colour and added (using hyperlinks) to the grey version. The space profit is high because of the quadratic effect of using high quality colour.

It makes no sense to use very high resolution, (for instance, 1200 dpi) images if one does not intend to have a high quality reproduction for conservation purposes. In this case, usage by scholars reading the texts is envisaged, and a medium resolution of 100—150 dpi (depending on the handwriting of the MS) suffices. Besides, it is quite possible to use images in different resolutions on the same CD. So one can add a very high resolution picture of some small intricate portion of a leaf. Finally, Medium-high to High JPEG compression proved to give the best results in sharpness, possibility of enlargement and minimalisation of image size [6].

The **Adobe Acrobat 2.0** software gives us an industrial standard solution for all what we wanted above, while meeting the criteria mentioned in **II**. Sequential (browsing) and hierarchical entry methods (texttree and networks) are supported using the Acrobat feature of *bookmarks*: pointers to a specific location in a text or an image.

Adding small coloured fragments to a grey image is very simple and transparent in use.

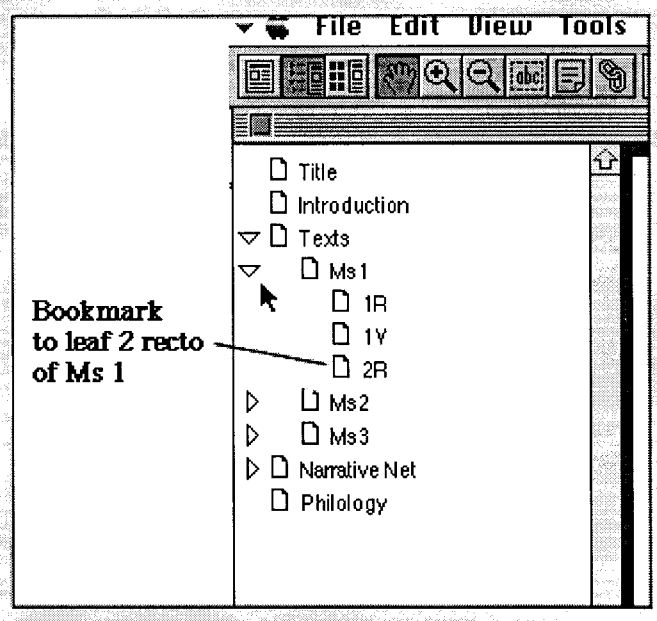


Fig. 3

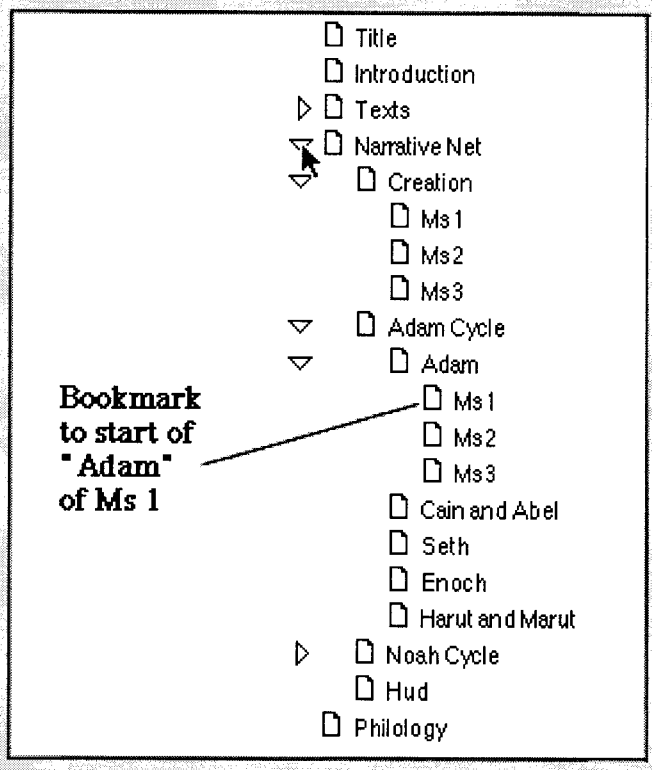


Fig. 4

Although not occurring in the analysis, the software supports hypertext and hyperlinks (from a location on a page (image or text) to another location (on a different page)), too. An important advantage of this software is that it allows the publication of electronic documents using fonts not present at the computers of the readers; this is an important fact in the publishing of oriental materials (*fig. 2*).

The little triangle shows that the hierarchical *bookmarks* are hidden (*fig. 3*).

Double clicking the icon to the left, for example, of 2R results in the presentation of the image of leaf 2 recto of Ms1 on the screen. It is possible to open simultaneously two or more images in order to compare them.

The same mechanism to describe the texttree (MSS, MS, Leaf) is used to implement the narrative network (*fig. 4*).

Double clicking the icon to the left of Adam will result in the presentation of the image of the leaf of Ms1 on which the story of Adam of the Adam cycle begins. In this way one can open two different versions of a story (if they are present on the same CD).

The Acrobat Reader software supports sequential browsing: **First** (presents the first page), **Last** (presents the last page), **Next** (presents a following page), **Back** (presents a previous page). It also supports to retreat on your steps if you have used any *bookmarks* or hyperlinks, since these are of course none-sequential (*fig. 5*).

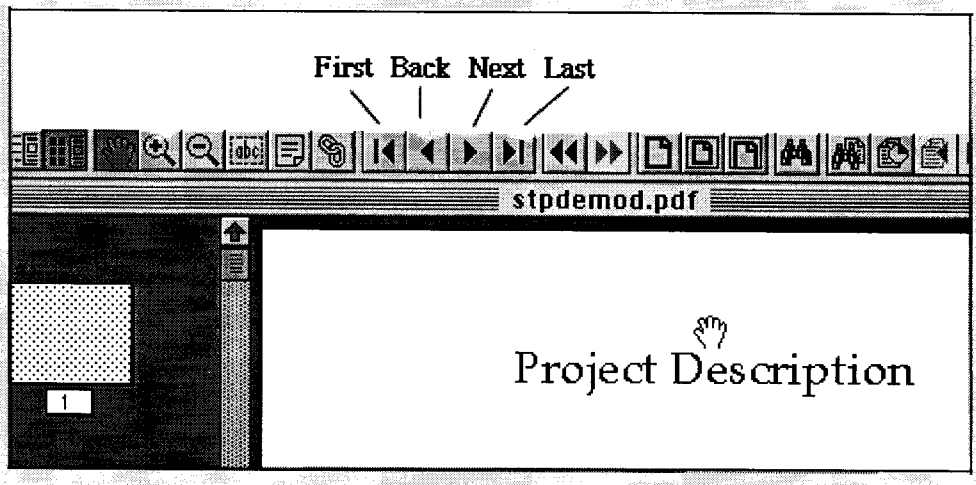


Fig. 5

V. Summary

We describe a way to develop a robust system for the digital publication of Chaghatay manuscript materials. An analysis of the information structure involved is given, and criteria for possible solutions are discussed. Finally, a concrete solution using standard software is presented and discussed.

Notes

1. Al-Rabghūzī, *The Stories of the Prophets, Qiṣaṣ al-Anbiyā'. An Eastern Turkish version*, critically edited by H. E. Boeschoten, M. Vandamme, S. Tezcan, with the assistance of H. Braam and B. Radtke (Leiden, 1995), i; Al-Rabghūzī, *The Stories of the Prophets, Qiṣaṣ al-Anbiyā'. An Eastern Turkish version*, translated into English by H. E. Boeschoten, J. O'Kane and M. Vandamme (Leiden, 1995), ii.

2. Cf. H. Braam, M. Vandamme, "Using Robust CD-ROM techniques in the preservation and disclosure of Oriental Manuscripts", Paper read at the 1st *IOICA* conference (Moscow 1995), in print.

3. For a more detailed analysis see M. Vandamme, H. Braam, "In search of formal identity and difference: considerations based on collating some Turkic manuscripts", *Manuscripts of the Middle East*, 3 (1988), pp. 32—40.

4. Sound and animation, of course, not applicable here.

5. In our case, **Adobe PhotoShop 3.0**.

6. For example, an image of 14 × 24 cm, 300 dpi, 256 grey takes in eps format 4660 Kb, in the Acrobat pdf format, using JPEG high compression, takes 353 Kb. But this resolution is high in comparison with the goal. It can be magnified 8 times, and this is in most cases unnecessary. A normal resolution will be 150 dpi, a 14 × 24 cm picture then takes 1343 Kb in eps, and only 153 Kb in pdf, high JPEG compression. So, a single CD could take about 3500 of such images. The pdf files were created using the **Acrobat Distiller 2.0** software, the PDFWriter extension lead to less good results.

Illustrations

Fig. 1. Texttree showing the hyperlink access connections from the MS-descriptors to the folia and the simple next/previous leaf access orderings.

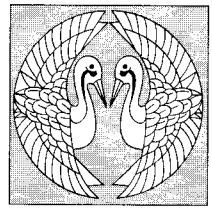
Fig. 2. A table of contents.

Fig. 3. A part of the "Texts" *bookmarks* uncovered (by clicking on the triangle).

Fig 4. The narrative *bookmarks* shown for the Adam Cycle.

Fig. 5. Acrobat Toolbar: sequential functionality, and other tools like magnification.

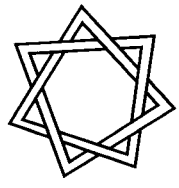
RUSSIAN ACADEMY OF SCIENCES
THE INSTITUTE OF ORIENTAL STUDIES
ST.PETERSBURG BRANCH



Manuscripta Orientalia

International Journal for Oriental Manuscript Research

Vol. 2 No. 2 June 1996



THESA
St.Petersburg-Helsinki

CONTENTS

Colour plates:

Front cover:

The inside of the manuscript's front cover (on the left): Čudabandaka (Skt. Cūḍāpanthaka; Tib. Lam-phran-bstan), "The Great Yum", MS, vol. 5 (call number K 24), 15.0 × 16.0 cm.

Back cover:

Plate 1. The inside of the back cover (from left to right): 1. Esru-a (Skt. Brahmā, Tib. Tshangs-pa); 2. Bigar (Skt. Śiva, Tib. ?); 3. Qormusta (Skt. Indra, Tib. brGya-byin), "The Great Yum", MS, vol. 5 (call number K 24), 52.0 × 15.5 cm.
Plate 2. The inside of the front cover (on the left): Inggida (Skt. Aṅgaja, Tib. Yan-lag-'byung); (on the right) Bagula (Skt. Bakula, Tib. Ba-ku-la), "The Great Yum", MS, vol. 4 (call number K 24), 53.0 × 15.5 cm.
Plate 3. The inside of the back cover (from left to right): 1. Qayanggiru-a (Skt. Lohakhaḍga Hayagrīva, Tib. Rta-mgrin lcags-ral-can); 2. Beiji Maq-a-kala (Skt. Aghora Mahākāla, Tib. Beg-tse); 3. Čoytu Ökin tngri (Skt. Ekamātā Shrī Devī, Tib. Ma-cig dpal-ldan lha-mo), "The Great Yum", MS, vol. 4 (call number K 24), 52.0 × 15.5 cm.