Sanskrit Manuscripts on the Internet

Dominik Wujastyk

Cambridge, 28 June 1996

1 The demand for digitized information

Everyone today is mad with digitization-fever. What does it mean for Sanskritists? Where will it end? Can we afford it? Can we afford not to do it?

The talk opened by addressing the question "What is the audience for digital archives?"

The following categories of possible users were considered:

- The general public, especially people of South Asian ethnic background;

- Scholars;

- Creating a new audience: New affordances offered by the new technologies and information-delivery systems. Reference was made to Donald A. Norman's important books on information and system design, *The Psychology of Everyday Things* (1988), and *Things that make us smart* (1992).

- It was noted that Chadwyck-Healey sells large digital archives to libraries and cooperatives;

- It was also pointed out that students today increasingly come from homes where PCs are taken for granted, and these students demand new technology at their colleges and universities. Institutions of higher education – including libraries – which remain unable to meet this demand will certainly be marginalized.

2 Economic aspects of the problem

This part of the talk addressed questions such as:

- Commercial or free distribution? Or cheap?

- If commercial, how much? (The example of Chadwyck-Healey's pricing was used.)

- Digital water-marks as a means of protecting rights to digital data;

- Copyright;

- What are the most appropriate delivery methods: Internet? CDROM? Future media?

3   The scholarly uses of a digital MS archive

The issues addressed here included brief discussion of:

- the uses of digital MS images in palaeographical studies;

- the creation of a bounded corpus (what is culture if not a shared body of knowledge, a shared literary or artistic corpus?);

- critical editions;

- reading texts (MS and printed text presented together simultaneously);

- study of miniature paintings (better to have a photo-disk);

- digitization and/or cataloguing: the crucial importance of retaining funding for cataloguing as part of any digitization project. A digital archive without a finding aid is useless;

- the importance of the MARC 850 linking field for integrating SGML/HTML encoded resources as well as image data with existing library finding-aids.

4   Problems and Technical issues

Some of the decisions to be made regarding the creation of digital resources depend very much on technical issues. Some of these were briefly explored:

- Technical obsolescence: Attention was drawn in particular to the work of The Commission for Preservation and Access, and it's recent *Report of the Committee for the preservation of digital archives.*

- It is vital to take the advice of conservators when contemplating putting a large number of manuscripts through the physical process of digitization. If digitization is done from existing microfilm, this too needs to be handled carefully, and the existence of a digital image must not be allowed to make other kinds of archival copy redundant.

- The quality of images is one of the first issues that has to be decided. It depends on such features as:

    - colour;
    - resolution;
    - size of image files;
    - size of the scanner window (which becomes a significant expense, for sizes of object greater than A4);
    - file format (GIF proprietary); JPEG limitations; etc.

- Although the worldwide web seems to be the very latest thing, it is important to plan ahead to a time when it is possibly obsolescent in its present form. HTML is a moving target (whereas SGML is an ISO standard).

5 What is everyone else doing?

In this part of the talk, I briefly surveyed some of the major projects of manuscript digitization that have come to my attention recently.

Indic projects:

- The Columbia/Brown project to digitize Skt. MSS in USA.

- IGNCA, New Delhi

Non-Indic projects:

- Columbia/Berkeley project for western MSS;

- (Perseus; TLG (Theodor Brunner); Rutgers center for electronic texts in the humanities; Oxford Text Archive);

- Chadwyck-Healey: 100,000 books on 18 CDROMs; All English poetry and drama < 1900; 11 editions of Shakespeare; All Latin literature from AD 200 - 12th cent. Nice to use: good affordances;

- Wellcome Video-disk: 50,000 images (including hundreds of Indian paintings). Extremely successful; in constant use by researchers. New affordance, opens up whole new way of studying;

- The Vatican / Rio de Janeiro University + IBM;

6 Conclusion

Can a modern library afford to be left out? What will happen if Cambridge, say, doesn't digitize its world-class Buddhist manuscript collection? The IGNCA is already planning to undertake the large-scale digitization of Sanskrit manuscripts. How can such projects be paid for, and managed in such a way that they will be successful and meet real needs in the scholarly world as well for the public and indeed all users of libraries with Indic holdings? These are crucial questions, and if we ignore them, we risk marginalizing the collections for which we are responsible.