Novel, cost-effective method of archiving manuscripts

Physical preservation of manuscripts is a difficult task, even under the best of conditions. Indian paper manuscripts may last four hundred years while palm leaf manuscripts may, under the best of conditions, last seven hundred years. And many of India's ancient manuscripts are now in a bad condition. In the past, scribes would have been called in and would have meticulously copied them (perhaps adding scribal mistakes to those that might already be there).

Photography, especially microfilming, and photocopying (xerography) have sometimes damaged the originals but have almost always been costly—and offer only a few decades of preservation. Manuscripts could also get lost during microfilming. Scanners have been used by a few centers but are relatively slow and can damage the manuscripts, especially ones that are deteriorating. A far more expensive approach has been first to microfilm the manuscript, and then to use a medical or high-definition film scanner to digitize the manuscript as images. This expensive method can be used only by the best-subsidized archives in the world.

Another option that was rejected prior to 1998 was digital still cameras which were both expensive and could copy only a few pages before their images had to be downloaded into a computer. Their flash-card memory was much too expensive and limiting for travel to an archive and copying for days at a time. Hence, an experiment was done in June 1998, when the first manuscripts were digitized directly with a DV format camcorder. But, in 1999 a new generation of digital still cameras was introduced which met the needs for in-house digital copying. The procedure is even simpler and can be taught easily. Two new digitizing procedures have been successfully tested at archives in South India by the National Institute of Advanced Studies (NIAS). The technology is far cheaper than conventional methods and the learning time is minimal (Figure 1). But there are some expensive mistakes that can be avoided.

A few archives in South India were visited for our tests. The Bhagavad Gita (a critical edition, two CD ROMs, myriad translations into the languages of the world) was chosen as the text (Figures 2, 3). The main reason for the selection of Gita arose from its very innocence. It was the most likely manuscript that archivists and librarians would allow to be copied with a new technology (note 1).

The initiative began with two goals: (1) to test digitization of manuscripts; and (2) to digitally copy manuscripts of the Bhagavad Gita for inclusion in a computer database open to the nation and the world. The equipment and digital capture procedures worked superbly. Where access was permitted a greater range of textual variants and variables were discovered than hypothesized (note 2). There was a great cause for alarm that the rare cultural treasures of India were being rapidly lost due to

**Figure 1.** Cost, time and process comparison.

*Cost:* The digital process requires an initial equipment investment of 1/20 to 1/30 of the conventional microfilm lab. Few libraries and archives can afford the Rs 10 lakh investment, and so have the work done outside.

*Secondary cost:* There are no film or processing costs in digital copying.

*Time:* Since there is one step less, the work is done in-house and the process is simpler. It requires less photographic skills. The digital process is estimated to take a time of 1/4 to 1/3 that of the conventional method.

*Once the data reaches the computer, the time and cost are relatively the same.*

The major disadvantage of less resolution is rapidly disappearing with new megapixel cameras. However, most of the high-definition, colour information is not needed and is even discarded in most manuscript usage.
SCIENTIFIC CORRESPONDENCE

Figure 2. Selection from the paper manuscript on Bhagavad Gita in Devanagari, from Manuscripts Division, Osmania University Central Library, Hyderabad.

Figure 3. Selection from the paper manuscript on Bhagavad Gita (in Sarada script), from Adyar Library, Chennai.

neglect, damage, etc. But there is enough hope that this can be reversed with this initiative from NIAS.

What was needed was a method that was less expensive, more permanent, capable of full colour, faster, in-house, and extensible (capable of being used with master books, CDs, databases and share over the internet). NIAS has tested one such method. The NiDAC procedure should be used with a notion of sharing these rare manuscripts with others as freely and widely as possible. The archive or library can still secure copyright for the manuscripts that they own, so that others cannot publish or exploit them commercially. Their use can be licensed (especially manuscript graphics), and they can be shared via the Internet or by CDs. The philosophy of sharing information is as important as any archiving and computerization technique. Low cost educational and research materials should be the result of the NiDAC procedure.

Instead of using a scanner to digitize each page of a manuscript as a computer graphic, the NiDAC procedure begins with the new DV video format. The DV camcorder that was chosen was the third generation of a technology that is advancing in about three months. The DV video format simply records everything as binary code onto a mini DV tape. The camcorder weighs less than two kg even with an eight-hour battery. The camcorder connects to higher end computers via an IEEE 1394 cable and card. Thus, there is no composite video to digital conversion. The digital image can be manipulated as a graphic or converted into alphanumerics (text). Images before compression are quite large but become small files when compressed into image formats like JPEG (depending on quality and detail required). Computerization is completed with various forms of storage (re-writable media like hard drives are quite fragile but CDs have at least a half-century life rather than a few decades like the microfilm).

The second method is to use a megapixel digital still camera with extra large memory cards. The memory cards are then removed from the camera, put into a disk-like device, and then downloaded via the floppy drive port into the computer. This method is at least ten times faster than downloading via a parallel or serial cable. Again, if the camera is well-chosen, the image enters as a JPEG file and is instantly manipulable for storage and distribution. This method is far superior to the DV digitization when the work is in-house. It will also work for extended field trips to archives if a laptop or a computer with adequate storage is available. The DV digitizing method can however be utilized for work in remote archives for extended times with no computer access and uncertain or fluctuating electrical power supply, etc.

The NiDAC procedures allow in-house copying of acid paper books (most of those published in the last fifty years in India). Such books are characteristically yellowed and crumble in hand as one reads or photocopies them and may disappear without a trace unless there is inexpensive digitization (note 3). NiDAC methods also include the direct copying of microfilm and microfiche into the computer, reducing the expense of current methods.

India should adopt a two-fold step to preserve its ancient manuscript treasures: (1) immediately begin a ten-year project of digitally copying all ancient manuscripts with new and relatively inexpensive technology, and (2) develop and use speech recognition technology (note 4).

Notes

1. Pre-information age philosophies still govern most archives and libraries in India (they fear losing control of their manuscripts – otherwise, what else would explain meticulous procedures to limit access and copying of manuscripts). Consequently many manuscripts have no backup at a given library or archive. Some cannot be discovered by a published list, record or catalogue. (Current computer access of an archive’s holdings via the Internet seems decades away.)

2. Most of the archives and libraries were poorly funded and supported. The procedures at more than 75% of the institutions
visited were highly protective, bureaucratic (requiring long forms and longer waiting periods), and often disappointing: manuscripts catalogued could not be found, manuscripts listed in the catalogue as 'in good condition' were highly damaged by worms or silverfish, even microfilms were often missing or found in a deteriorating condition.

3. Digitizing will come sooner or later to every library and archive in the world. It is the future. Librarians and archivists should know that these methods might pose a threat to their holdings. Rather than attempting to stop digital copying, a far more positive approach would be to institute procedures where scholars and scholarly organizations, or even commercial companies, will enter into agreement with the archive. Licenses and agreements can be predetermine use, profit sharing with private companies and procedures that encourage mutual cooperation.

4. Sanskrit, like any language, is both spoken and written. Sanskrit as an oral language is one of the most regular in grammar and pronunciation. It is in fact, so rigid that it has been called 'artificial' - great for finished literature but poor for conservation. Oral Sanskrit is perfect for computerization. Written Sanskrit has a clear and precise relationship with oral Sanskrit and consequently is also a perfect candidate for computerization. Already projects around India have begun this process. But other priorities have actually thwarted Sanskrit's computerization. Some are clearly political. No living language, even those derived from Sanskrit, meet these criteria - their irregularity, flexibility and colloquial inventiveness are precisely at fault. Written Sanskrit can be represented with any script but since the Roman script or alphabet has only 26 letters, this representation is incomplete and less than adequate. (Alphabetic, written languages also have important differences from syllabic, written languages. This need not be a concern at this stage but might suggest the need for another type of keyboard or data entry system instead of just copying the Qwerty keyboard.) Diacritical marks are added to make up for this deficiency in a Roman alphabet representing Sanskrit syllables. This should be standardized in 2-bit code, also recognizing the differences between alphabetic and syllabic languages. Using a standardized, universal 2-bit code would allow instant switching by computer from any world script (Roman to Devanagari to Kannada) to any other script.

ACKNOWLEDGEMENTS. We thank the Manuscripts Division of Osmania University Central Library and Adyar Library, Chennai. We also thank Prof. B.V. Sreekantan, Chief Advisor of this project, and Prof. Roodham Narasimha, the Director of NIAS, for their encouragement. The initial phase of this project is supported by INFOSYS Ltd. Bangalore.

SANGEETHA MENON
GEORGE M. WILLIAMS

National Institute of Advanced Studies,
Indian Institute of Science Campus,
Bangalore 560 012, India

Why does river Brahmaputra remain untamed?

Come May and the wild river of the east is seized with spells of unremitting spate, attendant deluge, sediment deposition and erosion. The flood hazard has been ravaging upper Assam unabatedly since time immemorial. Explorers of ancient times - those who established Praagjyotishpur in Kamrup province - must have discerned in this river something very different from all other rivers of the subcontinent, as evident from the singularly masculine name they gave it - 'Brahmaputra' or 'the son of Brahma'. Despite gigantic efforts and colossal expenditure (> Rs 15,000 million) in building 3647 km of embankments, 599 km of drainage channels and 431 km area of soil conservation, Brahmaputra continues to wreak havoc by uncontrollable floods year after year. Records show that catastrophic floods occurred in 1954, 1962, 1966, 1972, 1973, 1977, 1978, 1983, 1984, 1987, 1988, 1991, 1993, 1995, 1996 and 1998. Upwards of 9600 km² land, that is 12.21% of the geographic area of Assam, is annually affected by floods. In 1998, the flood which came in 4 frightening waves, deluged 38,200 km² or 48.65% geographic area of the state, putting in peril the lives and properties of 12.5 million people¹.

Why are we so helpless in containing the spates of Brahmaputra and coping with its flood hazard? Why do our efforts go awry and all civil engineering measures end up in shambles? The answer, in my opinion, lies in our failure to recognize the reality of active faults and continuing crustal movements in this geodynamically restless region. Understandably, the flood coping measures have never been designed to accord with this recognition.

Girdled as it is, by the arms of the Eastern Himalayan Syntaxis - the knee-bend of the mountain ranges - the Assam terrane is underthrusting northwards under the Arunachal Himalaya and, less energetically, eastwards beneath the Indo-Myanmar ranges². There is, therefore, very severe deformation and attendant faulting and thrusting in the terrane caught between the Himalaya-PatkalNaga ranges and the Meghalaya massif (Figure 1). The drastic reduction of the width of the alluvial plains from 350 to 300 km, respectively, in the flood plains of Sindh and Ganga to less than 100 km in the Brahmaputra basin is not without significance. Coming of the Meghalaya-Mikir blocks of the Peninsular Indian Shield closer to the Himalaya explains this attenuation of the alluvial domain of the Quaternary foreland basins.

The severe deformation of the Assam region is eloquently expressed in its much faulted framework³. The E-W trending faults (Dauki Fault, Brahmaputra-Mikir Fault), and the transverse tear faults (Kopili Lineament, Dhubri Fault, Duddhni Fault, Chidrang Fault, Um Nagot Lineament) and thrusts (Dapsi Thrust, Barapani Thrust) are among the many that dissect the terrane of the Meghalaya-Mikir blocks (Figure 2). The E-W Dauki and Brahmaputra-Mikir Faults roughly demarcate the southern and northern physiographic limits of the Meghalaya Plateau which is a horst of sorts. The plateau stands as a ~ 2000 m high physiographic escarpment against the sunken 3-4 m high (above sea level) Sylhet Plains in Bangladesh.

Most of these faults are seismically quite active⁴ as borne out by the distribution of epicentres in the fault zones (Figure 2), such as the locations of 7 earthquakes of M ≥ 4.5 along the 26.5°