

Digitising and Documenting Endangered Material: A Tale of Three Projects

Purbasha Auddy

School of Cultural Texts and Records

Jadavpur University

This paper would like to talk about three projects that created digital archives of books and manuscripts, and were funded by the Endangered Archives Programme initiated by the British Library in London. The projects were executed and housed at the School of Cultural Texts and Records at Jadavpur University in Kolkata.

The three projects are:

1. Archiving texts in the Sylhet Nagri script
2. Archiving 'popular market' Bengali books
3. Digital archive of early Bengali drama

The policy of the projects was to have two or three digital copies. One with the funding agency, that is the British Library; one with the project executor, that is the School of Cultural Texts and Records; and a third copy, with the individual collector, if any, who was the custodian of the physical material.

As the theme of the conference aims to foreground the documenting of diversity, I hope the discussion regarding these three cases of creating digital archives would do justice to the theme and the background paper of the conference. At the beginning of this paper the focus will be on the diversity of the three projects I just mentioned and then I would shift to their secondary representation and the metadata.

There were several prominent issues identified which proved that the digitisation projects mentioned above should be undertaken as the material was considered as endangered.

- ❖ In the case of Sylhet Nagri script, it is almost a dead script.



This image is showing a page from the Sylhet Nagri primer named *Pabelā Ketāb o Dui Khurar Rāg*. This script emerged as an alternative script for Bangla language in north-east Bengal. Currently, only a small number of aged people can



read this script. In this photograph we can see an aged person, Manai Miya Sadiyal who gave a Sylhet Nagri book for digitisation. Also he recited verses from the book.

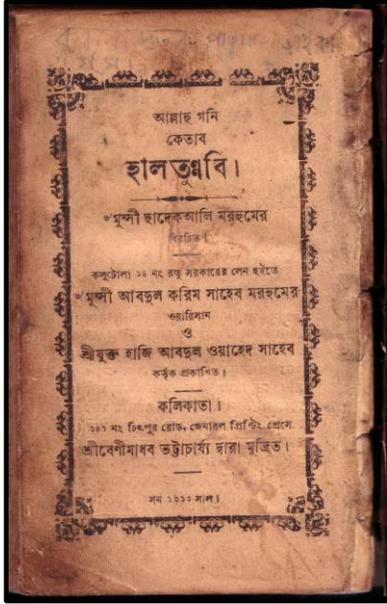


Manir Uddin singing from *Haltunnabi*; 14.11.06, Kachudaram, Sonai

In this picture another person named Maniruddin is singing from a Sylhet Nagri book. He gave two Sylhet Nagri books for digitisation.

When we tried to locate the texts we found only three institutions (Kendriya Muslim Sahitya Samsad in Sylhet, Bangladesh; Nehru College, in Cachar, Assam; and National Council of Education, Bengal in Kolkata) had Sylhet Nagri texts. Otherwise the texts were located with the individuals. To some individuals, speaking about these texts to the people of another community is really a very sensitive issue and there was a fear that people might take those texts away from them.

These texts formed a socio-religious identity for the custodians of these texts. This particular project involved extensive field work with a portable scanner and a laptop so that we could digitise the book in front of the custodians of the material and return the material to them. Our policy was to digitise every text we were getting, even though titles may be the same because during field work it was difficult to find out the differences, if any, between the volumes. Interestingly, we later found that there were indeed differences between some volumes of the same title. For example we collected 21 copies of *Halatunnabi*, a popular text which narrates the life of Prophet Mohammad.



This is the title page of *Halatunnabi*. The next image is a page from a Sylhet Nagri manuscript. These Sylhet Nagri texts cover subjects like metaphysical and spiritual; Islamic rituals and code of conduct, including lives of the Prophet and saints; love songs and love stories; social issues within Muslim society; and commentaries on natural disasters and social calamities.



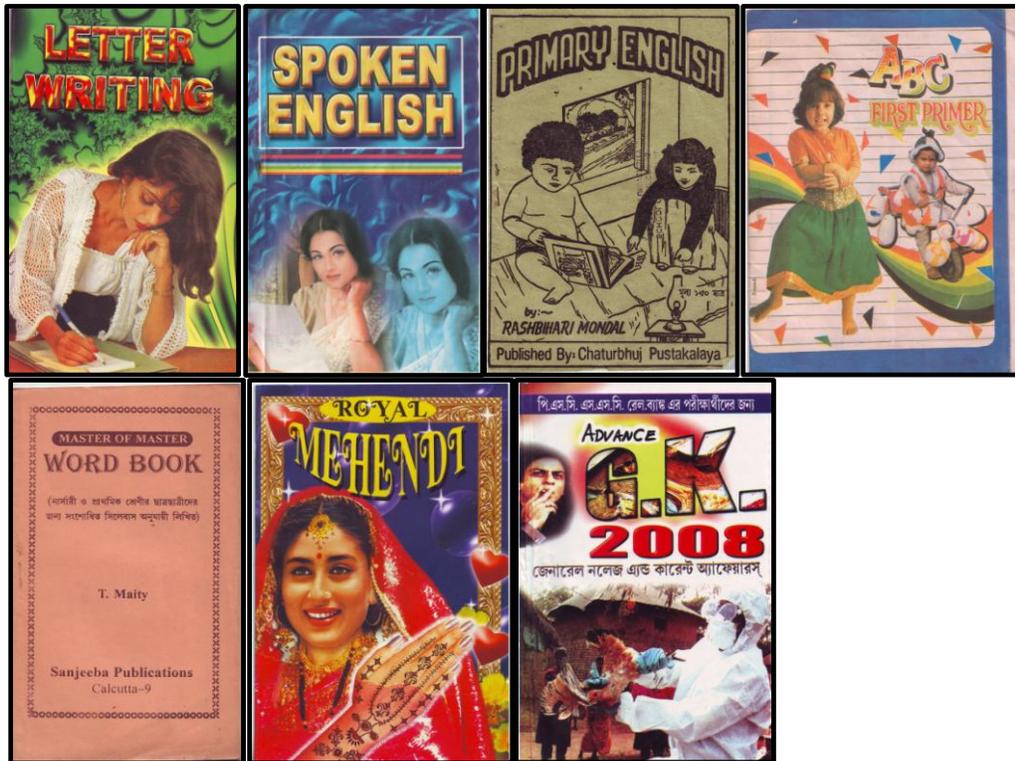
Manuscript of *Rādāpiāri*

- ❖ The second project was on ‘popular market Bengali books’. Apparently, it may seem that these books are not endangered and are still in production but no libraries, institutions or individuals have tried to build a collection of this kind of literature. Usually these books are not preserved for a long time but are rather thrown away or sold as scrap paper. These kinds of publications are ephemeral in nature. However these books vividly portray popular

printing practices. To elaborate the phrase ‘popular market Bengali books’ a few points come to mind:

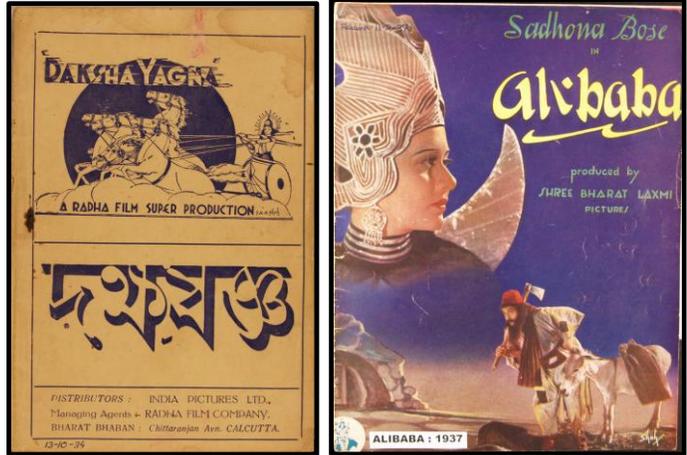
- Low cost of production in terms of paper, printing and content
- Sold in public transport, such as local buses and trains, and at local fairs
- Characterised as non-metropolitan

Several small publishing houses publish these books with an endless subject option or whatever is in vogue, like popular literature, religion, folk culture, local history, pornography and erotica, astrology, beauty tips, fashion and cookery, manuals of agriculture and animal farming, instruction on technical occupations such as repairing machinery and appliances along with serious topics like citizens’ rights, law, government procedure, public hygiene and social reform or sometimes one can find ambitious publications like *An Easy Way to Learn English: Make English Your Lap-Dog*. This book is so popular that it has at least 14 editions. We collected and digitally archived the 14th edition of this particular title. Here I would like to show some title pages.



I chose these pages because the titles are in English, so all of us can understand.

We also collected Bengali film and drama booklets under this collection. These were also ephemeral in nature but interestingly, just these booklets were collectors’ items as they were related to Indian film and drama history and due to digital preservation it came in to the public domain.



Here are the title pages of two film booklets.

These books have the information about the cast and crew, storyline and song lyrics.

After the completion of the project, both a physical as well as a digital archive was created.

- ❖ The third project I want to discuss is a collection of periodicals and books (and just one manuscript) on drama and songs. This whole collection was collected by a noted book collector and was not available for public access. These books were considered ‘important books’ as opposed to the popular market books discussed in the example of the previous project. These books were considered valuable and hence worthy of being collected. But the individual agreed to digitise the collection due to the fragile condition of the collection. Moreover, people used to visit him to consult those books. Gradually it became a bit difficult for him to manage the steady stream of visitors as the condition of the books deteriorated. However, when the whole collection was digitised his problem was sorted as now he shows the books in digital format stored in CDs and DVDs or refers the user to the School of Cultural Text and Records which also holds a copy of the digital files.

Some pictures





A book or a document can be digitised with a flatbed scanner, cradle scanner or a high-end DSLR camera among others depending on the nature and the condition of the material.



Uncompressed TIFF is the file format which is considered as archival quality. When a material is digitised it becomes an image with a technical filename that a scanner or a camera generates. But those filenames need to be changed with a uniform file and folder name so that the digital archive can be built in an organised way. For example, for the Sylhet Nagri project the text *Hālatunnabi* was coded as HALATNB01. The code has eight characters. All titles under this collection were coded with eight characters. If we collected more than one volume of a title, we kept the provision of numeric characters to avoid duplicate naming. This means other volumes of *Hālatunnabi* would be coded as HALATNB02, HALATNB03 and so on.

For the other two projects we also created codes for publishers and incorporated them in the file and folder name. In the project on popular market Bengali books, there were an enormous number of publishers to be coded. These codes are the most important element of a metadata of a digital archive as they help to fetch digital data swiftly.

After digitising, the projects accumulated the following amount of material.

Archiving Texts in the Sylhet Nagri script

- Printed books + manuscripts: 103
- Copied into 341 CDs
- Number of images: 13,654

Archiving 'Popular Market' Bengali books

- Texts: 2980
- Copied into 284 DVDs
- Number of images: 96,973

Digital Archive of Early Bengali Drama

- Texts: 385 volumes covering 243 titles
- Copied into 600 DVDs
- Number of images: 112, 174

Spreadsheet played a very vital role in creating metadata, especially when we are not using repository software like DSpace or integrated library system like KOHA. A spreadsheet has very useful functions like search, find, sort, filter that helps to arrange and organise large datasets. Moreover, if needed, a spreadsheet can be exported to DSpace or KOHA.

While creating a metadata for a secondary representation of an item, it becomes very important to describe physical items in detail. Let us take a book from the collection of popular-market Bengali books for example:

The digitised book is described by Digital folder name, Title, ancillary title (if any), date(s), extent, dimensions, creator(s) - author(s), creator(s) - editor(s), publisher(s), place of publication, subject, language, such as in the following entry.

Reference Number: RM-1937-001

Digital Folder Name: 127_SriBP_Albb

Title: alibaba [in Roman script]

Ancillary Title (if any): ālibābā

Volume and Issue Number: Not Applicable

Date(s): CE: 1937

Extent: Covers + 14pp

Medium of copies: Photographed digital copies

Medium of original material: Printed in black and white. Front cover coloured.

Creator(s) - Author(s): Unknown

Creator(s) - Editor(s) / Copied By: Unknown

Date(s) of Author(s)/ Editor(s): Unknown

Publisher(s): Sri Bharatlaxmi Pictures

Place of Publication: Calcutta

Subject: A booklet on the film Alibaba with its synopsis, cast list, lyrics of the songs and other production details. A small strip of paper, mentioning in print the year of release, is pasted on the front cover, Image 001. The year of release is also handwritten on the title page, Image 003. Price not mentioned

Earlier history: None found

Physical characteristics: Printed in black and white. Front cover coloured.

Dimensions [in centimetres]: 17.03 cm by 13.17 cm

Price: Not mentioned

Languages of material: Bengali, English

Note on item/images: Images reflect the condition of the original.

Languages of material: Bengali, English

Creator(s) of digital copy: Purbasha Auddy

Date(s) of the digital copy [dd.mm.yyyy]: 09.01.2009

Hardware: OLYMPUS E-500

Software: Adobe Photoshop 7.0.1

All these elements are standard metadata-heads taken from Dublin Core Schema which is a standard practice of describing resources.

So far this paper has tried to highlight some facts and standard practices that are related to digitisation and documentation of endangered books and manuscripts. But I would like to end with a different aspect of digital archiving. When a digital archive is built, it is stored in CDs, DVDs, external hard disks or a server. But these media of digital storage are itself vulnerable. It needs continuous monitoring for detecting whether the medium and the data stored in it is getting corrupt. But unfortunately, when a project ends we hardly give time to them later on due to lack of funds and personnel. The three projects I have talked about are now over but their data need to be still looked after so that they are sustained for a longer time.