

The International Dunhuang Project Interactive Web Database of Manuscripts: An Introduction

Susan Whitfield, British Library, UK

In October 1998 the International Dunhuang Project Interactive web database went online, providing details on over 20,000 manuscripts from Dunhuang and other Central Asian sites in the British Library Stein collection, along with over 1,000 high quality colour images. This paper introduces the web database, concentrating on its interactive features. The following paper by Colin Chinnery focusses on one future plan for database development, namely the digitisation and incorporation into the database of the archive of Stein photographs.

Details of the International Dunhuang Project (IDP) can be found on the web site (<http://idp.bl.uk>) and only a brief introduction will be given here. IDP was founded in late 1993 following a meeting of the major holders of Central Asian manuscripts and documents: the National Library of China, Beijing; the Institute of Oriental Studies, St. Petersburg branch of the Russia Academy of Sciences; Staatsbibliothek, Berlin; Bibliothèque nationale de France, Paris; and the British Library, London. It initially received a three-year grant from the Chiang Ching-Kuo Foundation for International Scholarly Exchange and we must thank the foundation for their continued support. One of the major sources for manuscripts from the pre-eleventh century in Chinese Central Asia was a Buddhist cave complex near the town of Dunhuang, Gansu Province, hence the Project's name. The Project set its sights on the achieving two broad objectives through international co-operation and collaboration, namely to improve and standardize manuscript preservation and conservation methods, and to increase accessibility to the manuscripts and documents.

Since its establishment the Project has held several conferences for conservators to discuss their work and to exchange ideas. From the beginning, scientists outside the collections have been part of this discussion and the work of Professor Seddon and postgraduates in the Chemistry Department of the Queen's University of Belfast is an excellent example of how scientists can aid conservators by both developing new specialised techniques and adapting existing scientific methods for the analysis of ancient papers, dyes and pigments. IDP decided that there was need for a larger forum for discussion between the two disciplines of conservation and science and a new series, *British Library Studies in Conservation and Science*, was launched in 1996 with the publication of papers from the first conference. Further details are on the web sites, but the second volume provides a detailed discussion of the techniques used by Peter Gibbs and Professor Seddon to analyse the yellow dye used on the Dunhuang manuscript sutra paper (*Berberine and Huangbo: Ancient Colorants and Dyes*). The Project also produces a newsletter with regular news on advances in conservation, different techniques in use at the various institutions, and relevant

publications, as well as reports on the Dunhuang collections themselves and general Project news. It is available in hard copy (free mailing, please contact IDP) and online via the web page.

This paper is concerned, however, with the second objective of the project, that of making the documents more accessible to the scholarly community. The first step to accessibility is the production some sort of list of the documents, yet none of the major institutions has yet completed this work, let alone full cataloguing. IDP has set itself the basic goal of compiling an inventory of the manuscripts. This task should be easiest in the case of the documents from Dunhuang itself because they form a finite collection whereas documents are still being found at other Silk Road sites, yet there are several major problems even with this group. Firstly, for a long time many scholars assumed that the Dunhuang documents came from a single cave, now identified as cave 17, probably a storage room for ex-library copies and sealed in the eleventh century. However, reading the early expedition reports it is clear that Sir Aurel Stein, for one, collected material from caves and sites at the Dunhuang cave complex other than cave 17. For example, a large group of Tibetan documents are listed by him as from another cave, and S.++++, listed by Giles with the reference number +++, clearly came from a local tomb as in his catalogue Giles writes '+++'. Other expeditions also collected material from several places, such as stupas, yet did not always make this clear. The first problem then is to ensure that those manuscripts which are listed as from Dunhuang actually came from cave 17 itself.

The second problem is with those manuscripts whose provenance is uncertain or has been lost. We know from contemporary records that members of Count Otani's expeditions visited Dunhuang and Japanese collections now contain scrolls which are recorded in the catalogues as being from Dunhuang, but in most cases the manuscripts in question have no marks on them to prove this provenance nor is there supporting paperwork. Many of Count Otani's finds were sold and have since passed through the hands of one or more private collectors and in many cases the scrolls have been mounted, so that it is no longer possible to see if there is any information on the verso, nor to assess the quality of the paper. Identification of a scroll as coming from Dunhuang rather than another Silk Road site has to rely all too often on stylistic features and general characteristics, none of which are foolproof. It is easier with the Stein and Pelliot collections, as both have remained in the public institutions where they were first sent. But although Stein marked most of his finds with a site identification code and number, he did not do this with most of the thousands of scrolls he acquired at Dunhuang on his second expedition. The exceptions, the scrolls he did mark (which seem to include all those acquired on his third expedition), are interesting for another reason.

The earliest first-hand documentation on the cave following its discovery in 1900 is the papers of Sir Marc Aurel Stein and Paul Pelliot, the first foreign archaeologists to arrive at the site (in 1907 and 1908 respectively). These documents and photographs suggest that the manuscripts were stored in

some sort of order in the cave and not just stuffed in randomly as has been suggested previously. For example, a contemporary photograph by Stein (<http://idp.bl.uk/idphighlights.html>) shows a bundle of manuscripts from the cave wrapped in sutra wrappers, the traditional way of collating manuscripts. Sutra wrappers have small tags which stick out when the bundles were placed on the shelves of a monastery library and which provide details of which sutra the scrolls belonged to. Another picture of Paul Pelliot inside the cave itself, shows scrolls stacked to the ceiling but, again, seemingly in side wrappers and therefore in some sort of order. On those manuscripts which Stein annotated with a site mark, he used the format 'Ch.zz.nn', where 'Ch.' refers to the site (The Mogao Caves near Dunhuang were also called the Thousand Buddha Caves, in pinyin 'Qianfodong' but in the romanisation in use in Stein's day, 'Ch'ien-fo-dong', hence 'Ch.'), and where 'zz' is a roman numeral indicating the bundle number and 'nn' the number of the manuscript in that bundle, and so it is possible to reconstruct which manuscripts were found together inside one sutra wrapper. Since most of the information about the scrolls' original disposition inside the cave is irretrievably lost this information is invaluable.

A third problem with compiling a list of Dunhuang material is that a significant part of the material is now in private collections and we simply do not know the extent of this material. Moreover, Dunhuang manuscripts regularly come on to the market and then disappear again, usually purchased by private collectors. A related issue to this is the problem of forgeries: there are certainly Dunhuang manuscript forgeries around, but since many of them are in private hands and it is not in the interests of those collectors to have their works revealed as forgeries, it is difficult to estimate the extent of the problem. Even manuscripts in public collections obtained after the clearing of the cave in 1910, such as those obtained by Stein on his third expedition and by Otani and Oldenburg, are thought by some scholars to be forgeries. This also has to be taken into account when trying to produce a definitive list of the original contents of Cave 17.

The most flexible method to collate all this confusing information was to use a database. The International Dunhuang Project database started to be developed in early 1994, using 4th DimensionTM software (<http://www.aci.co.uk>), chosen for its relational power, off-the-shelf availability, affordability, cross-platform and multi-lingual functionality. Details of the database structure are on the web site and in other papers, and a technical page will shortly be added to the web site with full specifications.

The database was envisaged as having much more potential than simply providing an electronic list of the manuscripts. To exploit the technology to the full it was decided to create a database which would work on at least four levels. Firstly, it had to act as a tool for the internal use of the institutions in question, by listing the manuscripts, their location, their institutional pressmark -- a unique identifying number -- any restrictions concerning its availability, conservation history, current state of

preservation, and archival and alternate forms - microfilms, photographs, facsimiles etc. Second, it was to act as a general scholarly catalogue, concerned with a physical description of the manuscripts and a scholarly description of the texts written on them. Third, the database would offer the user surrogate manuscripts, by incorporating high quality, colour images of the manuscripts, from the tiniest fragment to the longest sutra scroll, in their entirety. Fourth, it was envisaged as a complete scholarly resource: not just the equivalent of a sitting at a desk with the manuscript and printed catalogue, but with a whole roomful of books, maps, photographs, and links to other vital electronic research tools.

Over the first three years of the Project care was spent in designing a database that would fulfill these roles, that was stable, easy to use and accessible, and data from the British Library Stein manuscript collection starting to be input. In 1998, having entered preliminary details on the bulk of the British Library documents (23,000 out of a total of about 28,000: the main groups still to be input are a large group of Tanguit/Xixia manuscripts and printed documents and Sanskrit fragments), money was raised to start the digitisation of the manuscripts. At the same time, work began on the design of a web interface so that this material should become accessible on the internet. The interactive web database went online in October 1998 and it is this which I want to introduce in this paper.

The Project database contains twenty interrelated files and several hundred fields, but in order to ensure quick access the first stage of the web database only makes core information accessible. The database will be upgraded at regular intervals both to supplement the information offered and to make changes in accordance with users' requests and suggestions. The database is accessed through a log-in procedure. This is in no way intended to restrict users: anyone can log in as a new user, choosing their own username and password, but because of this procedure it will be possible, at a later development, to enable users to recall searches they made during previous sessions on the database. So if a scholar is interested in a particular group of manuscripts this groups can be recalled immediately, without the scholar having to re-enter the search string (which might be very complex).

Once the user is logged in they can access their scholars record from the database and make any emendations to this information, such as change of address, new research interests, recent publications etc. It is only possible to enter and change your own record and a note is made that the record has been changed via the internet so that the administrator can keep a check on the changes. The user can also send general comments at any time using a comments button. There is also a help button which links to information about use of the database and will be made more comprehensive as the database becomes more complicated. At the moment only the buttons are in Chinese and English: other commands and information are in English only. It is planned to create Chinese pages in the near future.

The first screen after log-in is a search screen, enabling either a general search using a combination of one to four search terms (language of document, physical form of document (e.g. scroll or fragment), site where document was found, and general subject), or a specific search where the user enters the library pressmark or manuscript number. The latter, of course, is directed at those already familiar with the material. At the next stage of development concordance lists will be added and the possibility of doing more complex and refined searches with more help functions. The search facility is essential to the success of the database and we welcome any comments about what users would like.

The results of the search are shown in a list format with thumbnail images, ten items appearing on one page. If the document has not yet been scanned, then a default image is displayed. By clicking on the image or the manuscript number the physical details of that items are shown in a full screen display, with the data on the left and a medium-sized image on the right. If, as in the case of *The Diamond Sutra* (ms. no. Or.8210/P.2), the document is too large to fit into one image, then a series of thumbnails of all the images are displayed in the bottom register. Clicking any of these brings the image to the medium display. Clicking the medium-sized image will bring up a full screen image. How many images are used to record a single document depends on the size and legibility of that document. A small fragment will be displayed in one image, and a large scroll on a series of images. If text is particularly small or illegible, then the image presented to the user will be much larger than the object itself. We aim always to present an image which is at least as clear as the original object and, in many cases, clearer. Further documentation on the procedures for capturing images and pilot project to reach the standards are available.

The user can reach the text details of the document either from the listing or from the physical details page. This is, in essence, the catalogue page, giving a title (in original language and translation), description of the text, date (where applicable) and transcription (in original script) and translation of the colophon. At a later stage it will also be possible to access available transcriptions of the whole text from this page, but our first priority is to digitise the manuscripts. Again, the user can see medium or full size images of the text.

The interactive feature of the database, seen in the scholar details function, is also available here. Any scholar can click 'Catalogue' to enter his or her own research. A new screen appears containing the user's name and a box for entering their research. This will then appear on the web when the database is updated with their name as author. Copyright resides with the author, who is free to send in emendations at any time. In this way, we hope to encourage scholarly debate. We will not edit the content of any research, nor try to make it agree with other catalogue entries. It is perfectly possible therefore, that the details on one text will include several scholars' catalogue entries and that they might be contradictory. The Project's aim is to interpret as little as possible: we leave that to the

people working on the manuscripts.

The broader research function of the database is seen in the sites and bibliography functions. The sites button brings up a screen giving details of the site where the object in question was found, along with pictures and maps. We hope to develop this further by digitising the Stein collection of photographs (see Colin Chinnery's paper). The bibliography offers a list of books and articles related to any particular manuscript or text. In the future we will scan in early articles which are out of copyright and so enable scholars to read them on screen.

Now that the database is online, we are keen to accelerate the digitisation process so as to enable scholars to get access to more manuscripts, but how quickly we can digitise depends entirely on our resources. Scholars and others can help us by sponsoring a sutra. You can chose any document in the Stein collection that is not yet digitised and, for a contribution, we will make the digitisation of that document a priority and add your dedication to the digitised image appearing on the web database, thus following in the tradition of Buddhists who paid scribes to copy sutras for the well bring of relatives or all sentient creatures, adding a colophon to this effect at the end of the manuscript (see Or.8210/P.2). Only 1,500 documents are digitised out of a total of 28,000 in the Stein collection alone and so any help is appreciated. Details are on the web page.

There are many other ways in which we hope to interact with the scholarly and other communities in the coming years and we hope that the interactive web database will be the start a long-term and rewarding dialogue between the International Dunhuang Project and all those people interested in the Dunhuang manuscripts.