PUBLIC ACCESS TO COLLECTION DATABASES:
THE BRITISH MUSEUM COLLECTION ONLINE (COL): A CASE STUDY

Tanya Szrajber
Collections Services Department
The British Museum
Great Russell Street:
London WC1B 3DG
England:
tszrajber@thebritishmuseum.ac.uk
http://www.britishmuseum.org/

In October 2007, the British Museum launched Collection online (COL), the Web version of its collection database. Beginning with two-dimensional works of art (mainly on paper), the Collection online project entails the phased release over three years of all the collection database (Merlin) records for the British Museum objects, as well as Conservation and Science records. By the end of 2009, the entire database of over 1,700,000 records for objects ranging from archaeology to ethnography, coins to prints, and spanning the broadest historical and cultural ranges, will be available to the public worldwide. The database is not yet complete, since the British Museum collection is extensive. However, the residue comprises mainly large holdings of sherds and documentary prints (e.g. bookplates and trade-cards), held in two departments. The primary collections in most departments are already recorded on Merlin, due to a long-standing and intensive programme of retrospective documentation and input.

Collection online is a dynamic system, as it is updated on a weekly basis to reflect new records and amendments to existing ones, including text and images. As a preliminary to releasing the data on the website, the COL project included two additional enhancements to the records: the addition of a large number of images (the target is 250,000) and text from published catalogues. A team of scanners was employed to digitise existing transparencies, scan two-dimensional objects directly (where considered safe for the objects) and scan the catalogues using OCR software.

In addition, and as a long-term commitment, curators themselves are encouraged to take new digital images and to facilitate the process, they are provided with cameras and training. Images of works still protected by artistic copyright (meaning works by an artist who is still alive or has died in the past seventy years) are on the Merlin database but will not be available on Collection online unless specific permission has been given by the copyright holder of that artist's work.

CONTENT AND AUDIENCE

A fundamental decision from the onset of the Collection online project was to make the records publicly available as they appear in the database, rather than attempt to re-write them for any particular audience. One reason, applicable to any such project, is the difficulty in determining a particular audience, since it was envisaged that the database would be consulted by a range of different users: academics, the general public and people with specialist knowledge of local history or culture. Another is the sheer size of the database which would make any re-writing totally impracticable. The database was originally intended for internal British Museum use, and its primary purpose was to provide an inventory of the collection, with academic information added over the years to facilitate, and record the results of, research. Users of COL will notice a large discrepancy in terms of quality and accuracy between the different records they come across in searches. In order to explain this, a history of how they were originally created is provided.

HISTORY OF THE COLLECTION DATABASE

Merlin is the fourth computer system used by the British Museum to record its collection.[1] The first computerisation project began in 1976 in the Department of Egyptian Antiquities. Over the years, and several systems later, it was extended to other curatorial departments. Around 1979, a section specifically dedicated to

---

1. For more detailed information on the project history, see Szrajber, T. (forthcoming in 2009), Collection Online: the British Museum Collection Database Goes Public. In Bentkowska-Kafel, A., Cashen, T., Gardiner, H. (eds), *Digital Archive Fever, Proceedings of the 23rd CHArt Conference held on 8th-9th November 2007, Birkbeck College, London*, Computers and the History of Art Online Proceedings, vol. 10, www.chart.ac.uk.

documenting the collections on computer was created - the Collections Data Management Section (CDMS). In the early 1980s, the purpose of the project was defined as the creation of an inventory of the Museum's collection, as a result of pressure from the Public Accounts Committee. In 1988, an online system (MAGUS) was at last introduced and this signalled a change in approach by the Museum to the project, since MAGUS was accessible to curators and much more sophisticated than its predecessors. The current system, Merlin, a product of System Simulation Ltd (SSL) was installed around 2000, and is used to record the collections in all the curatorial departments, as well as for Conservation records (Figure 1). It is SPECTRUM-compliant, supports images and UNICODE. The Documentation section works closely with curators and scientists on the terminologies (of which more later), and provides training, support and general advice.

DATA CONTENT ISSUES

There are various contributory factors to the lack of homogeneity in the records. First and foremost are the sources of paper documentation available to the documentation teams who over the years, created the majority of the records on the database: the accession (acquisition) ledgers known as 'Registers', record cards and in some departments, published catalogues, as well as the objects themselves. Since the Registers go back to 1753, and some of the catalogues provided were not the most recent publications, the information was often brief, inadequate or outdated. In addition, there was little curatorial involvement in the initial stages of the project, and since the information provided in the sources was at times ambiguous (e.g. a term might stand for a place or a period, an object or a ware) and at times archaic (e.g. colonial terminology for placenames or ethnic groups), such problems were reflected in the database records. The lack of direct access by non-documentation staff until the advent of MAGUS meant that curators could not access the database directly and so could not appreciate its potential. With MAGUS, they could search across a department's collection according to any criterion (e.g. material, type of object or provenance) rather than relying on the particular approach reflected in a published catalogue or set of record cards. As the database grew such searches became cross-departmental (e.g. by collector). Access to retrieval also encourages editing, since

once inadequate data is directly visible on the system, the need to amend it is more relevant. Since 1993, on-line registration of new acquisitions has been the responsibility of curatorial staff, encouraging more curators to use the database. The input of specialist and academic information to the records has made a significant difference to the database and its potential, although the process to update them all to the required standard will take many years, given its large size and the commitment of curators to other projects.

Another factor in the variation in the records is the number of data migrations from one computer system to the next. The inevitable progression to more sophisticated systems over the years meant that some data fields were added mid-way through the project, since the need for them became evident as the technology improved. The result was a large amount of post-conversion editing. This entailed having to identify specific data pertaining to a new field (e.g. Ware) when it was previously combined with other data in a single field (e.g. free-text Description). Post-conversion editing had to be done globally by the Documentation Section (often involving special programming), rather than manually, due to the large number of records involved, and could not deal with the more subtle or hidden data issues.

Finally, the database was created in stages, with departments joining the project in sequence, after the initial choice of candidate departments had been made. There are problems associated with this approach, as well as advantages. It is certainly preferable to complete an individual departmental database in a reasonable time-frame, before embarking on new ones, to enable users to make meaningful searches across that entire collection. The disadvantage is that the Documentation Section could not determine the overall requirements for cataloguing the Museum-wide collection, since until the actual work begins with any Department, it is difficult to predict all possible eventualities. The piecemeal approach affected terminology as well as data field choices.

TERMINOLOGY ISSUES

There are various forms of terminology control available on the database and hence in COL. They fall into three main categories: drop-down lists, thesauri and authorities. The drop-down lists are generally used for limited sets of codes and are not set up for candidate terms, although new ones can be added by the Documentation staff. Examples are methods of acquisition or production, dimension types and units, etc. In COL, the codes are available in their expanded forms only, to be meaningful to outsiders.

The thesauri are complex structures, with built-in relationships between the terms, and include Broader Terms (BTs), Use for Terms (UFs) which are also known as Narrow Terms (NTs) and Related Terms (RTs). They are polyhierarchical, meaning that a term can have more than one Broad Term, and several Use For or Related Terms. In addition, there are explanatory notes defining the terms (or sometimes their use in the British Museum) called Scope Notes.

The two Authorities are the Biographical Authority (for people and institutions) and the Bibliographic Authority (specifically for catalogues and equivalent publications of the British Museum collection). In particular, the Biographical Authority is fast becoming a major source of information on world-wide producer names (ie. artists and makers, but also publishers, printers and scribes) since many obscure people are recorded, as well as famous ones. Other names relate to acquisition (e.g. donors, vendors, etc.) as well as other associations to the object (e.g. people portrayed, authors, etc.). This Authority is extensive, with over 155,000 names to date.

The terminologies were developed in-house over many years, and are based on terms used in the British Museum paper documentation for 250 years and gradually entered in the database, as described above. Many are foreign terms, with very specific meanings and no exact English equivalents. The thesauri were set up in collaboration with curatorial and scientific staff. The decision to develop in-house terminologies rather than import them from established sources was made for several reasons: the British Museum collection is so varied, from so many cultures, that it was felt that no available resource could satisfy the requirements. The terms used in the documentation reflect a corpus of academic knowledge built up over 250 years of

documenting the Museum's collection and are standard in scholarly discourse. The terminologies are therefore sufficiently developed for any documentation and cataloguing requirements in the British Museum. Candidate terms can be incorporated far more easily into an in-house system, after vetting by the Documentation staff, rather than waiting for an external source to do so. The decision has proved to be correct, since not only are British Museum staff using and adding to the terminologies, but other institutions have asked permission to adopt (and adapt for their own purposes) two British Museum thesauri which were published jointly by the Museum and the Museum Documentation Association (recently renamed Collections Trust): the Object Names and Materials Thesauri. Other thesauri are for standard terminologies, such as Cultures and Periods, Techniques, Wares, Schools and Styles, Ethnic Names and Subject, as well as a more specialised one called Escapement (applicable to the collection of clocks and watches). The most complex terminology control is that for Places, which is structured like a thesaurus but functions slightly differently.

There are challenges in creating terminologies for such a diverse collection. In the case of names, rules for entering them had to be written, dealing with western and non-western forms, honorary names (titles), those of rulers, companies and partnerships. The provision of a repeating field for AKAs to cover spelling variations, different transliteration conventions and name changes (titles, married names, pseudonyms) means that all the names recorded in the documentation so far fit the structure. There is little possibility of ambiguity, since homonyms can be distinguished by their biographical details, e.g. Voltaire the watchmaker as well as the better-known philosopher of the same name!

In the case of thesauri, different problems arise. For placenames this relates to the changes in the boundaries of named place over time, the need to incorporate archaic as well as modern regions and countries, and the inability to identify certain places if they are now too obscure and no longer exist. In other thesauri, there are problems of interpretation and meaning depending on the culture the term relates to. Thus 'Bronze Age' has different meanings in different disciplines, and in this case and similar instances, the solution devised is the have elaborate Scope Notes explaining the use of

such terms in different contexts, rather than to create several 'Bronze Age' entries. Since the thesauri do not cater for homonyms, unlike the Biographical Authority, certain terms had to be redefined in order to avoid ambiguity. Thus a tailpiece in the context of prints is a type of book-illustration, yet in the ethnographic collection it refers to a personal ornament. When Subject was extended to the Asian and ethnographic collections, categories originally devised for western art had to be modified. Thus the term 'religion' had to be expanded to include 'belief', 'shepherd' had to broaden to 'herdsman', etc.

FROM MERLIN TO COL

It is not a simple decision to make an internal database available online, since there are advantages and as well as risks, financial considerations and staff implications. The idea was first discussed as early as 2001[2] although it was only seriously followed up four years later after the current Project Manager, Antony Griffiths, Keeper of the Department of Prints and Drawings, raised the idea with the other Keepers in June 2005. The project is run on existing staff resources since the production of COL was effectively carried out by the Web team, Documentation and IS as part of their workload   In addition, two contracts funded from Museum resources were drawn up for special technical expertise from an external company and for the scanning team. Decisions on screen design, schedule of data release, field inclusion, presentation of the data, and the use of images are made at Working Party meetings by representatives from the main parties mentioned above and a curatorial member of staff.

One of the initial questions was how to make the data available, ie. whether to release the whole database in one go, or samples (finite collections) from each department, or more complete sets of data such as an entire departmental database at a time. The decision was to begin with two-dimensional objects across the collection, then in a staged process, three-dimensional objects, followed by Conservation and Science

---

[2] Szrajber, T. (2002), Merlin – Opening up New Horizons in Collections Documentation at the British Museum. *Scanning the Horizon, Conference Proceedings 12-14 September 2001, University of East Anglia, Norwich*, *mda information*, volume 6, number 1, August, pp.109-112.

data. The last stage involves the notable collection of photographs primarily from the Department of Africa, Oceania and Americas, but including some from Asia and Middle East. This is because they are not 'registered' (ie. accessioned) as part of the main collection, but were historically classified as archival material.

The choice of two-dimensional works as the first release was decision was made for several reasons. The first is that although most of the two-dimensional art is in the Department of Prints and Drawings, it includes works in all the other departments and is therefore representative of the collection as a whole. Such examples include: Indian and Chinese paintings and Japanese prints (Asia), Mughal and Persian paintings (Middle East), icons and the Vindolanda tablets (Prehistory and Europe); banknote designs (Coins and Medals) Mexican codices and Indigenous Australian paintings (the Department of Africa, Oceania and Americas), Egyptian wall-paintings (Ancient Egypt and Sudan) and drawings of Greek sites and antiquities (Greek and Roman). Secondly, since the Department of Prints and Drawings receives a very large number of visitors to its study room (Print Room) as well as a continuous stream of written public enquiries, the practical advantage of COL is that many questions could be answered by the public directly from the website. Another reason is the very high quality of records for that Department. Although the initial data entry was by Documentation staff, nowadays a team of print cataloguers is employed to enter the data, funded externally, and curators have a high input as well.

Another key issue in any such project is the public user interface, including search and results screens. The Working Party agreed on a choice of four search options, to cater for different needs: a Basic search (the default), an Advanced search, and for more specialist enquiries, a Museum Number or Provenance search, and a Publication reference search. The Basic search is mainly free-text but includes date and 'image only' refinements. Since it does not access the terminologies, it is therefore of limited value, but easy to use. The aim is to encourage users to opt for the Advanced search since this allows users to select terms and hierarchies from the Authorities and thesauri and search for the terms in the correct fields, thus providing more accurate results (Figure 2). Help text is provided to assist users for each search option, and to explain the various terminologies.

Having determined a search mechanism, the next question is how much of the record the public should have access to when viewing the results: all data fields, or only some? The Working Party decided that all fields would be made available apart from: price, valuation, the Museum's storage locations (when Departments regarded this as a security risk), NGR references, a 'General Notes' field for internal queries, Label text (text from exhibition or gallery labels, etc). The terminologies are similarly available in their entirety (apart from the addresses of private individuals in the Biographical Authority). This decision was based on the fact that it is British Museum public policy that all the work generated by staff is in a sense public property and should therefore be made publicly available.

Since images play a significant role in the project, the question of whether to charge for them also had to be addressed. COL provides a free digital image service at two levels. On the ordinary Web level, a high quality image, deliberately selected as the best achievable given delivery constraints, can be downloaded. These images are good enough to use in, say, a Powerpoint presentation. At a higher level, COL provides a free overnight online delivery service of a much higher quality image (a larger file) suitable for a printed publication. In this service, terms and conditions are stipulated for non-commercial, usually academic, purposes. Users are reminded that if a work is still in the artist's copyright, they need to obtain the separate permission of the artist or his or her estate or successor in title in order to reproduce it. The Museum is not responsible for helping users in doing this research, but suggests DACS as one source of information. The images can be published, provided that the publication is non-commercial, and of an educational, scholarly or academic nature and has a print-run of not more than 4,000. The images must be credited: © Trustees of the British Museum. Users have to register details and are sent the image as an e-mail attachment. For very high resolution images and for commercial purposes, the users are referred to www.bmimages.com.

Finally, there is the important matter of providing a feedback mechanism for public comment, and how to deal with the comments when they are received. The screens

displaying the detailed records retrieved from searches invite public comment: 'Noticed a mistake? Have some extra information about this object? Please contact us' (Figure 3). The comments are collated and dealt with by a member of the Web Liaison Officer (a member of the Documentation Section). They are either answered directly, or forwarded to curatorial colleagues, depending on the nature of the comment.

ADVANTAGES AND PROBLEMS IN GOING ONLINE

The main advantage of COL is the fact that the public, nationally and internally, can search the British Museum collection in detail, and so access to the collection is now much greater than ever before. This is particularly significant since the greater part of the collection is not on public display, with about 40,000 objects in the galleries and several million in storage. COL thus enables users to gain a much more profound knowledge of the collection than could otherwise be achieved. In addition, visitors to the study rooms can now select the objects they wish to see in advance, and this process saves staff time and results in a reduction in the handling of the objects themselves. The presence of images plays an important part in enabling academics from across the world to draw comparisons or make discoveries about their own collections in relation to that of the British Museum. Comments received are incorporated into the Merlin records where deemed suitable by curators, and such improvements fed back to COL. Were more institutions to offer the same access, the overall effect would be a sharing of information about comparable objects held in various collections across the world, and this can only be beneficial to the public as well as to the individual institutions.

There are also problems and challenges associated with such a project. The records, as stated earlier, are inconsistent, and there are large numbers which require improving by curatorial staff. This process is increasing, although the improvements could not be made to all the records in time for their release on COL, given the size of the database. It was considered more important to release the entire database and face criticism of some of the records than to withhold any data. Nevertheless, the possibility of negative comment remains, relating mainly to paucity of information or

inaccuracies. For this reason, many curators were initially reluctant for records which had not been improved to be included in COL.

PUBLIC RESPONSE AND CONCLUSION

The Museum's commitment to provide as much access to the collection as possible, virtually as well as physically, is clearly demonstrated by making the database publicly available on COL. At the time of writing[3], the overall response to the database has been extremely favourable, with colleagues in other institutions congratulating the Project Manager on this achievement by the Museum. COL was praised for the quality of text and image, the sheer size of the database, and the indexing (terminologies) which enable high-level research. Comments received by the Web Liaison Officer include the provision of more precise or additional information about objects, suggested corrections (regarding, e.g. identifications) and enquiries relating more to the collection itself (e.g. other works by an artist represented). Although the project has not yet been widely publicised, as the first phase was to some extent experimental and certain problems became apparent which had to be sorted out in the second phase, the percentage of visitors to COL is nevertheless increasing and is already 7% of the total British Museum website use.

A new initiative in line with, and made possible by COL, is the publication of Online Research Catalogues on the British Museum website. These consist of introductory text and essays published through the web content management system, with links to the COL records for the relevant objects. The first of these, on Icons in the British Museum is due to be launched in June or July 2008. While similar in structure to traditional printed Museum object catalogues, they offer several advantages: they are free-of-charge and immediately accessible to the public, much easier and faster to update and cost the Museum far less to produce.

In conclusion, a project of this size and complexity requires careful planning, good communication between all relevant parties, and a sound knowledge of the database

---

[3] May 2008

and the collection. In the case of the British Museum and COL these conditions were achieved with internal, specialising in different disciplines but working as a team. As more institutions make their collections publicly available on the Web, the aim to make world-wide cultural heritage more openly available to the public will finally evolve from theory to practice.
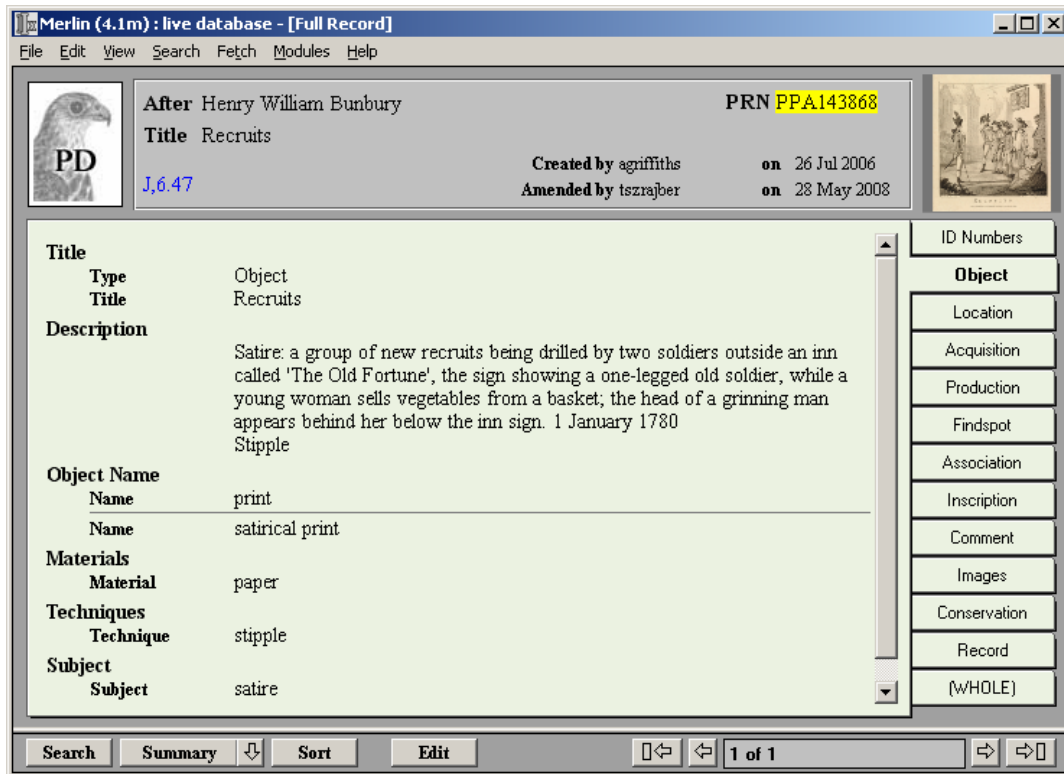

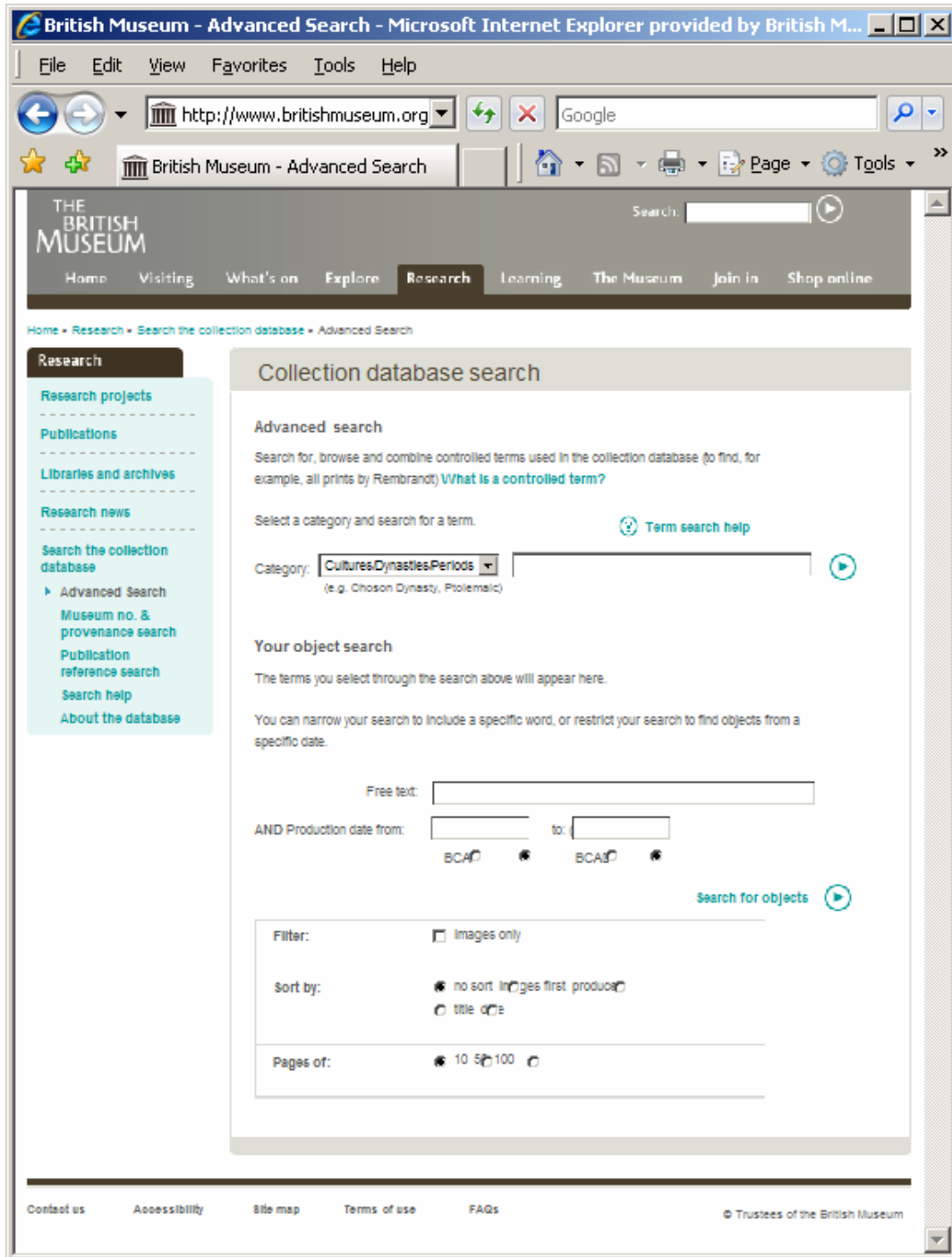Figure 1: Merlin database screen: Object 'tab'.

Figure 2. COL screen: Advanced search.

Figure 3   COL screen: full record.