University of Plymouth School of Architecture and Design

TQEF The Creation of a Database of West Country Buildings from MA/ PgDip.Architectural Conservation Student Coursework.

Bruce Induni,Associate Lecturer; Jean Manco,Associate Lecturer;Linda Watson,

MA/PgDip.Architectural Conservation Programme Leader.

Aim of the project

The aim of the project has been to create an accessible teaching and learning resource from the body of research contained in completed **ARCH 523X 'Evaluation of Historic Buildings'** assignments.

This has been successfully fulfilled with information on 150 buildings added to the database. Although the project has not set out to duplicate the material contained in the student assignments : it is a gateway into that material, not a substitute for it.

Nature of the resource

Introduction

A core part of MA/PgDip. Architectural Conservation at Plymouth is the module 'Evaluation of Historic Buildings'. The module is also taken as an option by MArch architecture students. The module is assessed by a single written and illustrated assignment.

Since 1993 a substantial number of these assignments have been completed. This library of student work contains the results of much assiduous research and no small amount of synthesis and insight. Very considerable research efforts have been made my many of the students. It is a valuable resource to any student of historic buildings. However, it is not easily accessible. There has been no catalogue or index for either the general reader or the specialist researcher.

The website and database produced by this project is a major step towards making the Arch 526 resource easily accessible.

Task set to students

The assignments catalogued and index by the project all addressed a common title :"Write an evaluation of a historic building". Complete freedom was given to students in the choice of building, with no stipulations of age, size or type. The format of the assignment was also largely a matter of student choice, with only a target word count being set. In practice, the great majority of

students have produced assignments well above the word target that are copiously and excellently illustrated.

It has been an expressly stated objective of the course team that this assignment should have both formative and summative utility. The open-ended nature of the brief has been intended to stimulate students to explore the process of research as well as to produce a work for assessment and grading. This objective may well have made marking more difficult, and it has certainly not eased the development of this project. A more restricted task would have produced a body of student work that would have been much easier to catalogue. On the other hand, the variety of buildings chosen, and to a lesser extent, the variety of approaches taken, has yielded a diverse and valuable resource.

Varied ability and experience of students

Students enrolling onto the ARCH 523X module come with a wide variety of backgrounds and abilities. The following groups are represented in the body of assignments:

- Full time architecture students, taking the 'Evaluation of Historic Buildings' module as an option, typically have developed observational and design skills and a significant technical vocabulary, but may lack any experience of older buildings
- Mature architects in full time practice seeking to obtain development of, and formal recognition for, historic building skills that they already possess.
- Conservation Officers and Planners from Local Authorities
- Owners of historic buildings who want to learn to repair and maintain their own buildings more effectively.
- Building Surveyors, contractors and builders who are looking for better academic and commercial understanding of their own specialisms.

The age profile of the course intake is generally mature, and the students are generally very highly motivated. However, the language, research and synthesis skills are not always well developed. Since the aim of this project has not been to grade the student work in any way - it has not been a remarking exercise - every effort has been made to extract all useful content regardless of individual student strengths and weaknesses.

Problems encountered by students

Students encountered all the usual difficulties found in researching historic buildings:

- Lack of documents.
- Inaccessible documents.

- Uncooperative owners.
- Over cooperative owners.
- Physical access problems.
- Unreliable oral history.

In addition, lack of experience caused difficulties for many students:
- Inappropriate choice of building on which to base the study. Many students chose buildings that were far too large and complex to be adequately discussed within the time and resource constraints. Such large and complex buildings also posed data presentation issues that were difficult to resolve within an 'academic essay' format. Conversely, some students selected buildings that were dull and limiting and could only produced dull and limited evaluations.
- Lack of understanding of the decay processes that had influenced the current condition of the chosen building. Tutorial briefing had warned students not to get involved in detailed evaluation of the condition of their chosen building as this would be dealt with later in the course. However, discussion of the developmental history of a building and of its possible future would often have been enhanced by greater understanding of decay processes.
- Failure to accurately identify materials.

Description versus interpretation

A general fault affecting many of the assignments was failure to interpret the evidence that was observed during survey. Many of the assignments are excellent descriptions of a building, and these descriptions are usually supported by excellent documentary research. Evaluation and synthesis based on the findings of the survey is generally much less strong. This is probably explained by the positioning of the assignment towards the start of the course where student powers of evaluation and synthesis are yet to be fully developed.

Photographic resource versus written description

The module has been running for almost 20 years and the technology of photography and desk top publishing has changed radically in that time. Early assignments contained little or no illustration. More recent ones are often lavishly illustrated. The project has tried hard concentrate upon content regardless of presentation. However, the development of the project as a web based resource has tended to discriminate against under illustrated assignments.

The quality of the student data and synthesis

It cannot be overemphasised that this project has not set out to assess the quality of the student work and has not systematically verified the quality of student data. Where suspect data has been observed this has been included in the database but flagged as doubtful or inaccurate. (See below under Data, and in the introduction to the database itself.)

Development

Initial plan: complete scanning

The initial design for data capture into the database was to scan all content from every assignment. This would be converted to text using optical character recognition and then edited for keywords and other relevant data. This would have had the considerable benefit of creating a full digital version of the student resource that could have been subsequently re-edited. Sadly, however, lack of time and funding made this option impractical. The time involved in creating and correcting the OCR process would have used all the project funding. A particular issue would have been that handwritten markers comments would have severely impeded the OCR process.

Implemented plan : limited scanning

To overcome the difficulties presented by complete scanning, keywords were selected 'manually' by reading every text. Scanning was limited to pictorial and bibliographic material. Whilst this brought the time commitment within the limits imposed by project funding it has also introduced an element of editorial inconsistency. Effort has been made to minimise this, it is an issue.

The picture content of each assignment was also edited before scanning. Only pictures containing unique information   for each building were scanned. In other words, where similar features from different parts of the building were pictured, only one exemplar was scanned. Hard to interpret and low quality images were also excluded. No systematic attempt was made to digitally improve the quality of the images.

Bibliographic material was completely scanned in '.jpg' format. Each bibliography has been captured as a viewable but not editable image. To make the bibliographies editable and searchable they will need to be converted to text by using 'optical character recognition' software. This should present few technical problems but will be very time consuming and is possibly the basis for another complete project.

Recording not marking

The aim of the project has never been to perform a remarking exercise. The editing process has consciously sought to avoid use of marking criteria in the selection of data for inclusion in the database. Synthesis, elegance of argument or quality of presentation have thus been ignored.

Not complete recasting into another format

As well as avoidance of 're-marking', the project has sought to avoid 'manually' recasting the student material into another format. Such an effort would be both pointless and time consuming. There is only one sensible option if complete digitisation is felt to be desirable, and that is to scan and OCR the result. As noted above this option would ultimately yield the most reliable and complete database, but would also be slow and costly, mainly due to the difficulties of achieving accurate OCR.

Developing the web site as an entry point to the data :

Within the project aim of creating an accessible teaching and learning resource, a particular objective has been to enable students to find personally interesting research areas before they have a technical vocabulary. For this reason the website uses images as the primary access gateway to the data.

Format

Introduction

This section discusses the detailed mechanisms adopted to achieve the aim of the project.

Data

- Choice of database programme: Microsoft Access 2007 was thus chosen as the data storage vehicle. It is recognised that there are significant compatibility issues between Access 2007 and previous versions of Microsoft databases but these are outweighed by the advantage of being able to include picture attachments directly into the database
- Keywords versus images : Images have been chosen as the primary access route to the data. The reasoning behind this is :
  - Traditional library keyword indexes require the user to have pre-existing knowledge of relevant keywords. Students embarking on the ARCH 523X assignment (the primary target audience of this project) often do not have this knowledge.
  - Many, if not most, library catalogues use a fairly arbitrary selection of keywords. Cataloguing staff rarely have time to read the material in depth and do not always

possess sufficient specialist knowledge to make sound editorial judgements on choice of keywords. Using images has not wholly solved this problem. Time pressure and 'editorial fragility' are still significant factors in this projects' index, but arbitrary adoption of cataloguing criteria has been minimised. The index images are the product of editorial selection, but none of the images has itself been edited. Thus the students of ARCH 523X are able to directly connect with the intentions and interests of the original authors.

- Structure of database : The data is entered into a single table using Microsoft Access 2007
  - Flat file versus relational: The nature of the data and the designed end use of the data by ARCH 256 students did not require the separation of data into multiple relational tables. However, one of the key design parameters has been to make the database easily expandable. Future expansion, and especially expansion by linking to other data sources can be accommodated by linking additional tables to form a relational structure.
  - Key field choice: A 'key' consideration in database design is selection of a a meaningful unique identifier for each record within a data table. In a commercial database this key field would usually be a product identification or invoice number. Rather than create any arbitrary identifier, it was decided to maintain simplicity and accessibility by using the building names as used by the assignment authors.
  - Temporary library numbers: Solely for the purposes physical handling and storage during data entry each assignment was given an identification number. These numbers are included in the database but have no real significance and can be deleted if desired.
- Access to the database : Direct links to the data from the images in the web site index give access to a modified version of the database, which has the following differences from the full version :
  - It is translated into HTML to give faster and more effective browser based access, but has removed the full searching, reporting and querying functionality of the native Access 2007 format.
  - Confidentiality issues have been resolved by removal of potentially sensitive data
- The full database : This has been produced as a standalone resource for staff, and at staff discretion, for selected students. This restriction addresses issues generated by giving students direct access to potentially sensitive data such as names of building owners and building content.

However some sensitivities do remain :

- ○ Confidentiality : Assessing the Data Protection Act responsibilities generated   by publishing material contained in student assignments was outside the scope of this project and remains to be resolved at University level.
- ○ IPR : It is understood that students sign all their Intellectual Property Rights to the University as a condition of enrolment. However, where students have used material given by third parties, the situation is far from clear. This is especially true of images, drawings and documents copied from County Record Office archives. This is a significant issue and needs clarification before the full database can be released to students.

Pictures

- ● As a gateway : As explained above, images are used to index the data so that searching is not contingent on specialist knowledge or vocabulary. Such 'index images' do not need to be of particularly high quality or high definition. Small file size is the key factor.
- ● Selection : No additional images have been added to those contained in the work submitted by students. This is significant because some student work was poorly illustrated both in terms of quality and quantity of images. A good example is the assignment on Harry Hems' House in Exeter, which contained no illustrations of scannable quality. As a general rule, the older assignments are less well illustrated than more recent submissions. Where assignments were copiously illustrated, not all the images have been used. The editing process has used the following criteria :
  - ○ Generality : preference has been given to images that make a general point about a type of building or provide a context for other images that record building details
  - ○ Relevance of detail : images showing details of buildings are included where they add to the understanding of a particular building or type of buildings, but repetition of similar details (eg. window cills) has been excluded.
  - ○ Clarity : poor quality pictures have not been included.

- ● Size : No attempt has been made to standardise the physical dimensions of images, though many have been either cropped to remove irrelevant areas or generally reduced to fit the page layout of the database.
- ● Format : The JPEG format has been used for all images since compression and small file size were more significant than absolute image quality.

Maps, plans and site drawings

- Area location maps have not generally been scanned into the database because of copyright issues. Exceptionally, buildings in very remote areas, where location is relevant to the content of the assignment, do have a location map.
- Site plans and drawings have not been included. Again copyright is a significant issue as many of the site drawings used by students have been donated by outside professionals. The large formats used by much of this material make effective scanning prohibitively difficult and time consuming.

Bibliographies

- Introduction: The ARCH 523X assignments have generally been very well researched. Many are based on exceptional research effort involving a wide variety of primary and secondary material. It is this research effort that makes this project so worthwhile. However, the bibliographic skills of the student authors are far from perfect. The limited nature of the bibliographic resource included in the database reflects the imperfections of the student's own efforts.
- As pictures: Student bibliographic material has been included without editing. Each bibliography is captured as an image (or images) with original formatting and markers comments and questions. It was felt that the markers comments generally form an essential part of the 'complete' text. However, markers comments were occasionally so extensive as to occupy page space well outside that of the student material. Here the user of the database must refer to the original text for the full marker comment. In other words, extensive comments by the marker which do not directly contain student material, have not been scanned.
- OCR : To make a searchable database of student bibliographic material would require :
  - Processing of the 'raw' scanned image using Optical Character Recognition software
  - Removal of the markers comments, which are handwritten and cannot be processed by OCR software
  - Standardisation : students have not used a consistent format for their bibliographic material and abbreviations and acronyms would need to be standardised.

Resolution of these issues would have required a substantial investment of time and was outside the scope of the project.

Content

Quality and accuracy of the student work

- Quality: Although a striking variation in academic quality is clear to the experienced eye, the project has not considered the grades given to individual student work, nor to markers' general comments on student work. The project has not taken any editorial decisions based on assessment of the quality of the assignments. It is thus up to the user of the data to assess quality of the student work. Perhaps a mechanism is needed to provide follow up tutorial support to users of the database, so that student errors are not reproduced into the future.

- Accuracy: There has been no general editorial policy to establish the accuracy of the material contained in student assignments. Where obvious errors and inconsistencies have been observed, such material is included in the database in parentheses: '*'. Where editorial comment was felt to be essential, this is included in square brackets: [*]. These mechanisms do not represent a systematic editorial campaign to ensure quality and accuracy.

Marker input

The marker input written on many of the assignments is very searching and extensive. Further extensive marker material has been produced on separate pages which did not inform the project. Marker comment has not been included in the database due to lack of time. However, it is a major resource, both for its own content, and the context and criticism that it adds to the student work.

Editorial neutrality

As discussed above (Development), the project is based on editorial judgement for the abstraction of data from student assignments, and for the organisation of that data in the database. It is acknowledged that editorial judgement can never be wholly objective. The personal attitudes, experience and interests of the database designer have undoubtedly influenced the content of the project. A particular area of sensitivity concerns buildings that have been restored rather than repaired.

Quality and accuracy of the keyword selection process

The most accurate, useful and editorially neutral method of capturing the data contained in the student assignments would have been to 'scan them complete'. Given that resource constraints made this impossible, every effort has been made to consistently abstract the most significant key-words. The choice of key-words has been tested on a panel of non-specialist readers and found by them to be both complete and accurate. However, no claim to perfection is made. There will be errors and

omissions. Student users of the database will need tutorial support in this area. It is likely that user experience will yield useful data on how the key-word selection can be improved and extended.

Editing criteria

No universal list of editing criteria for the selection of keywords could be produced. The format and content of the student assignments was so varied that no single set of key-words fitted all. Instead, an editorial approach was applied with key-words generated by:

- Discussion of general data sources
- Specific references to historical context and significance
- Discussion of unusual or illustrative details
- Discussion of the process of data acquisition

Discussion of the database format

Abstraction of data from the student work formed the bulk of the editorial effort, but it was not the only area where editorial judgement played a formative role in the project. The organisation of data within the database is an integral part of the editorial process. Once a decision has been made to include a particular field for data entry, this automatically biases the editor`s eye towards relevant data in the original text. In other words, the structure of the database shapes the data that is chosen to fill it. The choice of data entry fields has evolved as the project progressed and it is hoped that a workable structure has resulted. The real test will be future student use. It is to be hoped that resource will be found if experience suggests that the format of the database can be usefully modified.

Testing and Refinement

The database will be tested as a pilot on the MA/PgDip. Arch. Conservation students attending ARCH523X during autumn 2009. This will allow any minor amendments to be incorporated.

Future development

Introduction

Substantial opportunities exist for developing the project further. The list below suggests ways that the data could be linked to external data sources. There is also the possibility of extending the database to include student work from other modules run within the School of Architecture at Plymouth or co-operating with other academic institutions.

Links to external data sources

The database could be linked to published sources such as :

- Pevsner : The Buildings of England
- List of historic buildings
- RCHM databases
- At risk register

Linking might involve further substantial editing and manually inputting or might be achieved by relatively simple electronic merging of data. Obviously considerable discussion between interested parties would be required.

Conclusions

The successful completion of the project has highlighted a number of key issues and possibilities :

- That student assignments are not perfect, factually or academically, but often contain large amounts of unique, valid and useful data obtained by significant research effort. The output of the ARCH 523X students used to build this database is far more than a regurgitation of lecture notes. It is an original research based resource.
- There is a substantial and valuable body of student work that is not accessible to future students. It is very unusual for university libraries to hold copies of students work and no examples have been found of a systematic attempt to organise a comprehensive database of the contents of such work.
- That the process of archiving and editing is seriously hampered by a paper based submissions policy. Throughout the construction of this database considerable time and resource has been wasted by re-scanning material that was produced digitally in the first place. At the least students should be required to produce digital versions of their assignments in addition to paper copies.

Bruce Induni / Linda Watson 30.07.08