5027MEDCUL Research Methods – Week 9 Doing Communications Research – Building the Digital Archive
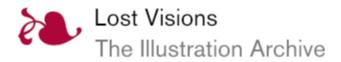
Lost Visions
The Illustration Archive

'Lost Visions: Retrieving the Visual Element of Printed Books from the Nineteenth Century", led by Professor Julia Thomas in the School of English, Communication & Philosophy, Cardiff University, is one of 21 new research projects, funded by the Arts and Humanities Research Council, to address the challenges of working with big data and making the information more accessible and easier to interpret by a lay audience.

In collaboration with experts at Cardiff's School of Computer Sciences & Informatics, Professor Thomas will work on more than a million illustrations from books in the British Library's collection. Although many historical texts have been digitally stored, the illustrations in these texts are frequently without the information needed to help people understand them. The Cardiff team will devise methods that will enable the collection of illustrations to become fully searchable online.

The "Lost Visions" project is creating an internet-based system for enabling users to interact with digitized content in a variety of different ways. A key objective is to allow digitized content to be searchable using crowd sourced, bibliographic and content-based features. These features are captured as keywords, either based on text provided by a user (as part of a crowdsourcing activity), keywords extracted from key positions in the bibliographic data (e.g. name of illustrator, engraver, photographer, book title etc.), or those obtained by the outcome of an image-processing algorithm. A website has been implemented for the project which links into a high performance computing environment at Cardiff University to support image storage and processing.

A key novelty of this project is the combined use of these features to facilitate search.

The system enables users:

(i) to view and interact with digitized images;

(ii) to tag images using keywords or to make use of a pre-defined taxonomy;

(iii) to add images of interest into a "personal" archive. This archive does not allow images to be downloaded, but primarily to be recorded into a "personal space" on the web site;

NAME:

Worksheet

Enter the following address into your browser:

http://lost-visions.cf.ac.uk/

For the purposes of feedback, we need you to set up an account.

Click on "My Account", then "Log In"

Click on "Sign Up"

For your user name, please use the prefix LJMU – so, for example, LJMUHorrocks

Return to the start page and your name should now appear at the top of the screen, signifying that you have logged in

Remember, this is an Archive that is in the process of being built, so some areas of the site will remain blank

Task One

Click on "Tag Images" and work through the 4 questions on the right hand side

Consider your choice of keywords carefully – which will be most relevant to the user? How can you account for variations in meaning/cultural associations?

Are there any "types of image" that you would expect to see listed but which are not?

Consider the "types" listed – why do you think that these are important types to foreground? How will they be useful to a researcher of communications for example?

What are the problems of creating your own captions for each image? What editing would you expect to be undertaken?

Consider the logistics of monitoring crowd-sourced data – potential problems? What recommendations would you make to the designers of the archive?

How do you discern between a "description" and a "caption"? Could these questions be labelled differently?

Do you feel empowered as a user to be able to collaborate in creating an archive?

What recommendations would you make to new users of the site?

Task 2

Use the search facility to see if there are any images recovered that are directly relevant to your chosen research topic for the assessment, bearing in mind that the process of tagging images has only really just begun.

Comment on the ease of use, how easy it is to manipulate the data.

Try using the "My Archive" tab to sort and collate your returns. How useful would this feature be if you were to actually begin writing up your research proposal?

This project enables you to move from the position of consumer to producer, combining the eccentric archive model we examined with the Liddell Hart Collection with the digital model we examined with the Mass Observation Archive. What are the implications for further research and the future of archive studies?

Any other comments?