Electronic annotation for researchers: a feasibility study
Item 5: Project Summary

A. Context: the value of annotation

Researchers in any discipline will read a large number of documents, such as technical papers related to their topic. If the document is on paper, researchers will almost always annotate each document they read, typically by underlining passages in the document and writing comments in the margin. In some cases the need for the annotation will be ephemeral, e.g. the researcher just has to review a document. Often, however, the need is *long-term* and the annotated document will be kept and re-read at a later date, perhaps years in the future. A central purpose of the annotations is to aid this re-reading by capturing the researcher's original opinions, e.g. on what parts are especially important and relevant.

Nowadays researchers increasingly obtain documents from electronic sources, such as the web. Often documents will be read, or at least browsed to identify important points, on a computer screen. If a document is reasonably long/important most of us prefer to print it and then read the paper form. Nevertheless more and more reading and browsing is done from the screen.

Facilities for electronic annotation are currently poor. They are often tied to one type of document, e.g. a Word document, and overall do not come close to offering the convenience of paper annotation.

Potentially, however, electronic annotations have huge advantages over paper ones, because the power of the computer can be used to organise the annotations and make them easy to retrieve. Thus the researcher may issue a command to retrieve all the annotations made over the last 5 years that relate to the topic X. X may, for example, be a specific gene or a piece of legislation. The computer can then retrieve all such annotations, together with the documents they relate to.

Some possible approaches to electronic annotation lie in hardware that unites the electronic and paper worlds, such as the Digital Desk, Anoto pens or Intelligent Paper. It will, however, be at least ten years before such hardware can cover all annotation needs. Thus in the short to medium term, at least, potential approaches must lie in software. Even in the long term software will always be part of the solution.

B. The project and its objectives

Our project aims to study researchers' needs for annotation, and to outline potential software solutions. We will concentrate on annotations for long-term use and on informal personal annotations rather than group ones (which Cathy Marshall has found to be fundamentally different from personal ones). Our research hypothesis is that the purposes of annotation cross all disciplines. Thus, for example, (a) scientific researchers face issues with massive amounts of literature, and fast technical change; (b) language researchers face issues of translation, annotating parallel texts and accessing documents in a source language that is not their own; (c) researchers in English might want to study the sequence of annotations made by past authors (Wordsworth added annotations to the draft of his *Prelude* over 60 years); (d) social scientists and historians will want to

track changes in government policies over time and across different countries — they are especially concerned with chronology, both of original documents and annotations.

The *objective* is therefore to identify and support the needs of electronic annotation. If this leads to the identification of basic common principles covering all disciplines, our research hypothesis will be supported.

C. Significance and originality

Our topic is of wide significance, since it can bring benefits to all researchers. This project is a first step along a path which could lead to work of great impact.

Our work is original in making a cross-disciplinary approach to electronic annotation, and in seeking basic principles common to all.

D: Method

The project will be in three stages.

The *first stage* (Months 1–6) will be to look at the working practices of researchers, and to evaluate their needs for electronic annotation. For convenience, and to reduce cost, this study will be largely centred on Exeter. As well as the various departments of Exeter University, it will involve the Meteorological Office, the Devon Record Office, and an overseas institution, the University of Würzberg. The study will not only cover textual documents, but also images and multimedia objects, and, innovatively, annotation of procedures (e.g. annotation of why a certain web link was followed or a certain option was set). We would expect a centre of focus for the study to be the web, but we will also cover other software that researchers regularly use.

It is important, before starting the study, to specify a coherent methodological framework. We plan to spend the first month doing this; Professor Brown will play a major part here – the main RA on the project will not start until the second month, i.e. after the framework has been well-defined and tried out. Obviously there is no absolute division between annotation and other document-manipulation technologies such as hypertext linking, bookmarks, history trails, etc. Our approach will be catholic in the sense that we will be looking at meeting researchers' needs, rather than being technology led. The first stage will, we trust, result in a novel and wide-ranging 'annotation theory', which captures unifying principles underlying apparently different annotation practices.

The *second stage* (Months 4–10, i.e. overlapping with the first stage) will be to propose new software solutions to meet researchers' needs, and to implement 'proof-of-concept' demonstrators. A cornerstone here is to use computer power to add value to electronic annotations, thus compensating for their disadvantages over paper. This value lies in retrieval, comparison, proactive suggestion of relevant material, etc. We envisage some proposed solutions might be ambitious and futuristic, whereas others might be simpler and more restricted (e.g. text only). Our goal is to find more general and more powerful methods of electronic annotation, that could help *all* researchers in their work. If we just end up with a huge and diverse set of individual features, and thus leviathan software to implement these features, we will have failed.

The *third stage* (Months 9–12) of the project will be to show these software solutions to researchers, investigate how well they meet researchers' needs and revise our theories and software as necessary. Given that the solutions will be demonstrators rather than polished software products we do not expect production use of these solutions; the aim will be to identify the solutions offering the best potential for future research and implementation, and to point the way to more powerful annotation practices. We also plan to take advantage of our links with Southampton University, a leader in the project field, to refine ideas.

We intend to produce published papers for all three stages. In addition we will maintain a web site to present results and to promote wider discussion of ideas. The final deliverable for the

project will be a report with a provisional title of 'Electronic annotation: making research reading more productive for everyone'.

E: Reasons for approaching the Leverhulme Trust

The key reason for approaching Leverhulme is that our work is cross-disciplinary. As outlined in Section B, researchers in different disciplines make use of annotation in different ways. However we believe that a lot of these differences may be superficial in that they relate to the different practices in using documents, rather than to underlying principles. If we are successful in identifying such principles we will help in a small way to bring disciplines together.