This paper was originally presented to staff and students at the AHRB Centre for Editing Lives and Letters research seminar, October 2003.

The Auchinleck Manuscript Project as an exemplar of collaborative research
<http://www.nls.uk/auchinleck/>

July 2003 saw the launch of the online facsimile-edition of the Auchinleck Manuscript, National Library of Scotland Advocates' MS 19.2.1. This paper provides an account of the project's editorial and intellectual goals and the processes and procedures undertaken to achieve these. The aim of documenting the project in this way is to identify some of the issues that researchers encounter when undertaking a humanities research project with online outputs within the current academic context. In particular, this paper considers the possibilities and potentials of collaborative academic endeavour.

How was the project established?

The Auchinleck Manuscript Project was the initiative of the late Professor David Burnley who established the project in September 2000 as a collaboration between two institutions, the National Library of Scotland and the University of Sheffield. There was a team of four at this initial stage of the project: project leader Professor Burnley and Research Associate Dr Alison Wiggins at the University of Sheffield Department of English Language and Linguistics, and at the National Library of Scotland in Edinburgh, archivist Dr Kenny Gibson and web designer Tony Stuart. Ian Hewines subsequently joined the project at the NLS. Funding for the project consisted of a grant from the University of Sheffield to cover the cost of employing the Research Associate part time (.6) for one year and the use of departmental facilities.

Why the Auchinleck Manuscript?

The Auchinleck Manuscript attracted the interest of the research team at Sheffield because of its importance for understanding the development of written English, particularly the varieties of Middle English in use in early fourteenth-century London. Auchinleck was produced in London in the 1330s and its forty-four texts were copied by six scribes under the direction of the main editor-copyist 'Scribe 1'. The manuscript is most famous for the large collection of verse romances it contains but the collection also includes a range of other texts, including

chronicle, hagiography, basic doctrinal instruction, an alphabetical poem in praise of women, a tale of a merchant who betrayed his wife, and a satire on the state of society.[1]

The National Library of Scotland (NLS) regard Auchinleck as one of their most rare and precious items and Auchinleck has a unique place within Scottish literary history. The manuscript came to Scotland in the early eighteenth century and was donated to the Advocates' Library in 1744 by Alexander Boswell, Lord Auchinleck and father of James Boswell the biographer of Samuel Johnson. It first came to prominent notice sixty years later when, in 1804, Sir Walter Scott published an elaborate edition of the Auchinleck romance *Sir Tristrem* with an introduction describing the manuscript.[2] Auchinleck thus came to play an important role within Scott's campaign to promote Scotland through its literature, or in the case of *Sir Tristrem*, through the literature he claimed for Scotland.[3] As a result of Scott's interest, Auchinleck has also come to play a distinctive role in the rise and definition of medieval studies and the conception of the Middle Ages as we know it today.

The need for a full edition of the manuscript reflects its acknowledged wider importance across the fields of textual and cultural studies. Auchinleck is an exceptional manuscript in a number of ways and has become one of the landmarks used by scholars of language and literature to describe the history of the late medieval period. It provides a snapshot of literature in English in the generation before Chaucer. The large size of Auchinleck and number of unique texts it contains make it particularly important in this respect, both in terms of the interest in these texts in their own right and the information they provide for how Chaucer and his contemporaries were influenced by an existing English literary tradition. Auchinleck has also become a key case study for the history of the book and is regarded as the earliest example of a commercially produced English manuscript. The individuals it was produced by and for remain anonymous but the manuscript's physical make-up has been repeatedly scrutinised over the last century as evidence for the workings of London's early book trade. It was the significance of Auchinleck to English language and literature in the

[1] The manuscript originally contained over sixty texts but has suffered substantial losses. For an introduction to the manuscript see <http://www.nls.uk/auchinleck/editorial/importance.html>; for discussions of the scribes see Timothy A. Shonk, 'A Study of the Auchinleck Manuscript: Bookmen and Bookmaking in the Early Fourteenth Century', *Speculum*, 60 (1985): 71-91; Judith C. Mordkoff and Ian C. Cunningham 'New Light on the Signatures in the Auchinleck Manuscript', *Scriptorium*, 36 (1982): 280-92.

[2] W. Scott (ed.), *Sir Tristrem: A Metrical Romance of the Thirteenth Century by Thomas Erceldoune Called The Rhymer* (Edinburgh: Constable, 1804).

[3] See the studies by D. Matthew, '"Quaint Inglis": Walter Scott and the Rise of Middle English Studies', *Studies in Medievalism*, 7 (1995): 33-48; O. D. Macrae-Gibson, 'Walter Scott, the Auchinleck Manuscript, and MS Douce 124', *Neophilologus Mitteilungen*, 50 (1966): 449-54.

widest sense, and to Scottish literary history in particular, that made it the focus of our research interest.

What factors determined the editorial rationale?

It has often been said that one of the great advantages of online publication is the freedom it grants from the constraints of print. That is, freedom from the economics of print publication and from the restrictions of the codex form.[4] In a number of ways these advantages have been felt here. However, online publication does impose its own constraints and the editorial decisions made on this project had to respond to a number of different factors. These were not factors determined by a print publisher but by the objectives and priorities of the two institutions involved in creating, hosting, and maintaining the edition. Following consultations in the initial stages of planning the project, three key criteria were established. These are described below; (1.) and (2.) primarily reflect the priorities of the NLS, and (3.) primarily those of the research team at the University of Sheffield.

[Criteria 1] To improve and ensure both preservation and accessibility

Archives have a duty on the one hand to preserve the documents they hold and on the other to enable access to them. These competing and apparently contradictory priorities present a particular challenge in the case of delicate, rare, and important documents such as the Auchinleck Manuscript, where there is a higher than usual demand from students and scholars to see a particularly fragile document. It was therefore crucial that this edition should offer a readily accessible and high-quality alternative to the manuscript itself. One which would reduce the intensity of use of the original whilst at the same time encouraging and enabling scholarly research.

The decision to make available online high-quality colour images of the manuscript reflects a recognition of the benefits both for manuscript preservation and research that access on demand to such images would offer. A printed facsimile of Auchinleck was published in 1977

[4] For discussions of the theoretical and analytical opportunities that hypermedia offers to editors and readers see Jerome McGann, 'The Rationale of the Hypertext', IATH (1997), accessed 4 February 2004 <http://jefferson.village.virginia.edu/~jjm2f/chum.htlm>; Jerome McGann, 'Imagining What You Don't Know: The Theoretical Goals of the Rossetti Archive', IATH (1997), accessed 4 February 2004 <http://jefferson.village.virginia.edu/public/jjm2f/rationale.html>. See also the article by Peter Robinson which points out a number of the economic, pragmatic, and editorial implications of electronic media, 'The Computer and the Making of Editions' in *A Guide to Editing Middle English* (Ann Arbor: University of Michigan Press, 1998), pp. 249-61.

and, though in many ways an excellent resource, it has certain shortcomings.[5] The print facsimile is something of a 'rare book' itself: it is available only within major university libraries where its use is often confined to the library or its special collections room. Furthermore, as it is a black and white reproduction, important aspects of codicological information have been lost. Colour features, such as rubrication, painted miniatures, other decoration, and certain signatures, cannot be examined and these provide information about the book's production, status, quality, and visual coherence. The use of black and white images and the scarcity of the 1977 facsimile are the result of the high cost of print production. To publish a colour facsimile in electronic form is to bypass both of these problems and present a more flexible, accessible, and faithful alternative to the original manuscript. Of course, an online edition can never offer completely 'fidelity' to the original as the appearance of the images depends upon the graphics card, settings, and calibration of individual users' computers. A colour print facsimile would offer much greater control over what the user actually sees, but the prohibitive expense of colour print makes online reproduction the only viable option for a readily available and accessible colour facsimile. If users ensure their computer monitor is correctly calibrated for optimum viewing, then it can also be a reliable alternative to the original manuscript with a very acceptable level of colour fidelity.[6]

[Criteria 2] To offer a resource for education and outreach

The Auchinleck Manuscript Project is part of the NLS's evolving digital library <http://www.nls.uk/digitallibrary/index.html> which provides the widest possible public access to a range of educational resources. It was therefore important that the project address the issue of accessibility in the widest sense. This included ensuring that the site would function correctly across different browsers and operating systems. But it also determined how the project's expectations of its users were to be defined. We did not expect users coming to this site to be equipped with a high level of computer literacy or any specialist skills beyond the ability to use a web browser. Be they University Professors, A-level students, or the general enthusiast, it was our belief that specialist computer skills should not be expected of users. It was therefore a priority to produce a site that incorporated web design best practice and offered a user interface with clearly presented materials, consistent built-in navigation, and minimum barriers to access.

[5] Derek Pearsall and Ian Cunningham (eds), *The Auchinleck Manuscript* (London: Scolar Press, 1977).
[6] The original images have been archived by the NLS. These are high-resolution tif images, which include colour swatches, and therefore provide the possibility in the future for generating high-quality faithful colour images on screen or in print. The first set of updates to the Auchinleck site will refer users to calibration instructions.

Accessibility would define the site's structure and interface, and influence decisions over content. However, we were concerned to strike a balance between the desire to produce a scholarly edition and the need to make the site attractive to as wide a range of users as possible. The fulfilment of these dual aims is reflected in the content of the final edition. The material expected of a scholarly facsimile edition has been included in full, but this has been supplemented by additional apparatus to aid the non-specialist. This apparatus includes the following. (i.) Graphic icons (with an explanation in the <alt> tag) mark links to folio images and textual notes; so, for example, users would click on an image of a folio rather than on the conventional abbreviations 'fol. 7ra' and 'fol. 7.rb' to view the image of folio 7 recto <http://www.nls.uk/auchinleck/mss/tars.html>. (ii.) Glossaries and visual images have been incorporated to clarify technical descriptions; so, for example, users can refer to dynamically embedded word glosses and visual illustrations in the necessarily technical codicological description of the manuscript <http://www.nls.uk/auchinleck/editorial/physical.html>. (iii.) Introductory material has been divided into sections which can be accessed under a browse-able list of titles. (iv.) Summaries head the various sections to orientate the user. (v.) Information of a technical nature is always presented first as a non-technical summary, with the option to click to access a more detailed description if desired.

[Criteria 3] To provide a searchable corpus suitable for linguistic research

The criteria of primary interest and importance to the Sheffield research team was to create an edition which would offer a searchable corpus as a resource for linguistic analysis. Auchinleck is an important representative of a stage in the development of the language of late medieval London. Traditionally, the development of London English has been described in terms of a series of incipient standards (Types I, II, III, and IV, Auchinleck being a representative of Type II).[7] It is a model which has been highly influential in thinking about the development of standard English and the changing status of the language prior to the age of print. However, it is a model which was developed by manual sampling from texts in the era before computer assistance was available. By making this important set of texts available together, in searchable form, our aim has been to offer scholars the opportunity to perform 'whole-data' linguistic analysis of texts in their original context. It was our conviction that this approach would enable scholars to describe the development of standard English and the complexities of linguistic usage with greater depth and appreciation than has previously been possible. Our assertion was that the greatest value of electronic media for scholarly editing is its ability to offer the facilities for user interaction and analysis.

[7] M. L. Samuels, 'Some Applications of Middle English Dialectology', *English Studies*, 44 (1963): 81-94.

How did these criteria influence the form of the final edition?

The hypermedia environment made it possible to combine two traditionally distinct types of edition: a non-critical edition of the texts and a photographic facsimile of the folios. Whilst the team at the NLS arranged for all the folios of the manuscript to be digitised, the Sheffield research team proceeded with the preparation of the texts. With over 300 folios, 44 texts, and six scribes, the transcription and editing of Auchinleck was no mean task. The first issue of primary importance was to devise a technical and editorial policy for transcription and the decisions we made were influenced by the need to accommodate a number different factors. It was important that (i.) the texts were appropriate for linguistic analysis; (ii.) the texts were readily searchable; (iii.) that the texts complied with the NLS's objective that the site was accessible to as wide a usership as possible; (iv.) that production was technically feasible; (v.) that the texts would be able to be produced within the project's time frame and within the limits of the resources available.

The resources available on this project were insufficient to permit all the features of the texts to be encoded in SGML to the level of *The Canterbury Tales* Project or the *Piers Plowman Electronic Archive*.[8] Such encoding would have made it possible to offer users the ability to choose the way they view the texts (with the option of, say, one style of display for palaeographic information, another for those interested in orthography, as well as a regularised version for the non-specialist). It would also have offered options for accurate searching for different linguistic forms, and for processing and storing the texts. However, this project simply did not have available staff resources to make possible this kind of deep tagging for the entire manuscript. The challenge for us was to find a less labour-intensive solution which would still enable the objectives of the project to be fulfilled.

The decision was made to produce text-only semi-regularised transcriptions, held in HTML, such as would be suitable to be searched with a commercial software package.[9] This decision, to simplify the transcriber's task and thereby reduce the labour involved, meant that it was possible to produce the texts for the project with the equivalent of one full-time RA in post for 6 months. The ability to match available resources to an appropriate technical specification is

[8] For further details see the web sites of the *Canterbury Tales Project* <http://www.cta.dmu.ac.uk/projects/ctp/> and the *Piers Plowman Electronic Archive* <http://jefferson.village.virginia.edu/seenet/piers/>.

[9] The search software purchased for the project was dtSearch. For further information see <http://www.nls.uk/auchinleck/search/index.html> and <http://www.dtsearch.com>. The transcriptions were done by the researchers at the University of Sheffield in HTML; a second layer of XHTML was added by the web team at the NLS.

crucial when designing a research project with electronic outputs. There are three points to this triangle: research objectives, available resources, and a project's technical specification. Each of these points must be assessed, weighted, and adjusted to ensure that a project is led by its research objectives and desired research outcomes. Our goal was to produce an searchable edition of the entire manuscript; limited staff resources necessitated developing a technical specification which was not overly labour intensive.

As we had decided against the kind of deep tagging which would have made different views of the same text possible, it was therefore important that the transcription we did produce be suitable both to be displayed and searched. In order to facilitate searching with the dtSearch software, we reduced the number of non-orthographic variants in the texts. The transcription policy specified that abbreviations should be silently expanded and that word spacing and variant letter forms be regularised (so, for example, long and short versions of *s* and different forms of *r* are all represented by the graphs 's' and 'r' in the transcription). Abbreviations and letter forms are of potential interest to palaeographical analysis; word spacing is of potential interest to considerations of pronunciation and meter. However, each of these remain areas of specialist interest. Furthermore, sample analyses confirmed that these features are highly conventional in Auchinleck and the information would, anyway, be available in the images of the folios. For these reasons the decision was made to prioritise the orthographic information in order to facilitate searching for different spelling forms.

These aspects of regularisation were also desirable because they improved accessibility and resulted in texts which were clearer and more readable. Accessibility was further enhanced by the decision to add modern punctuation and capitalisation – an aspect of regularisation which did not present difficulties for the search engine or detract from the ability to search for different spelling forms. The decision was also made to permit editorial intervention into the texts; obvious scribal errors have been corrected and, in a couple of instances, where large sections of text are missing because of a lost folio, this has been reconstructed with text from another manuscript. Had we decided to regard the manuscript as purely a spelling data-set then we may not have included these editorial additions. However we had to balance this against producing editions of the texts which would have wider application, which would stimulate interest in the manuscript more widely, and which could be utilised by tutors and students. Those undertaking intricate spelling analysis are able very quickly to identify editorial insertions.

More problematic than what we chose to regularise or reconstruct was what we chose to retain. We decided to represent scribal usage of *u* and *v*, long and short *i*, and yogh and thorn.

It was felt that these characters provided important information about spelling and orthography and should not therefore be regularised. To give one example, Auchinleck Scribe 3 uses yogh and thorn in a very idiosyncratic way and seems to consistently confuse them, using yogh where a or ð sound would be expected. It has been suggested that he was perhaps a non-English scribe unaccustomed to these characters, clearly a factor of interest to understanding the manuscript and its language as a whole.[10] It is certainly the kind of issue that should be accommodated by a project that proclaims to be concerned with spelling analysis. To replace yogh with 'g' / 'y' and thorn with 'th' would have completely gone against our research objectives.

On the other hand, the stated priorities of the NLS were such that they would have favoured much more extensive orthographic regularisation for two reasons. First, the inclusion of yogh and thorn instantly makes the texts appear less approachable for the non-specialist. The decision, ultimately, to include these characters reflects the emphasis that all parties put upon the project's underlying research objectives. It also reflects a realistic recognition that, although use of certain characters may repel some users, an orthographically faithful edition is likely to be regarded as a positive bonus to teachers and tutors wishing to set students language-based exercises.[11] The second reason is that the ability to reliably reproduce the Anglo-Saxon character yogh online across different web browsers and operating systems is more technically demanding than reproducing an alternative graph (such as 'g', 'y', or 'gh') from the standard roman alphabet. The initial concern was that inclusion of these characters may necessitate a compulsory download for users (who would have to download and install a customised font before they could view the site correctly). As this would be a major barrier to accessibility it was an option we did not want to have to take. However, a solution to this issue was found and by dynamically embedding the customised font within the site's style sheet it has been possible to avoid the need for any downloads for the majority of users.[12]

What would you do differently?

The technical specification that has been described has ensured that this project fulfilled the stated objectives of each institution and did so within the specified time frame. The

[10] The suggestion was made by A. J. Bliss, 'Notes on the Auchinleck MS', *Speculum*, 26 (1957): 652-58.
[11] Feedback from undergraduate tutors has since confirmed this to be the case.
[12] The site is optimised for use on Internet Explorer (version 5 and above) and Netscape (version 6 and above) browsers running on Microsoft Windows operating systems. Other users may have to download and install the Auchinleck font to correctly view all the characters on the site.

Auchinleck Project was a project led by its research objectives and an IT solution was customised according to the resources available. This is not to say that this is the only or, indeed, the best technical solution to this particular brief. Given greater resources, a more sophisticated technical solution can be envisaged. This is a example of a project that would have been ideally suited to having two or more web portals, with each portal accessing the same data but providing different interfaces that required different levels of knowledge and expertise from the user. It would have been possible to design one portal for an identifiable group of scholars and a second for non-specialist users, each with access to a different level of transcription and editorial apparatus. The 'scholarly' edition could have omitted editorial additions and emendations from other manuscripts and offered a diplomatic transcript. Whereas the second portal could have offered regularised transcripts and editorially corrected texts. Further portals fulfilling the requirements of other users would be possible, such as, say, a semi-regularised transcription for students, or a fully graphic transcription for those interested in palaeography.

What were the advantages of cross-institutional collaboration?

Significant advantages have been gained from this cross-institutional collaboration between the NLS and University of Sheffield. First was the range of expertise the project was able to call on from both sides. Collaboration was essential to bring together a core team that combined editorial, codicological, and linguistic skills with professional-level IT competence. In addition to this, the project benefited from regular contact with each institution's distinctive research culture. The NLS has a staff with considerable experience and specialist knowledge in the field of manuscript digitisation and it was a task for which the library was fully equipped. At Sheffield we were keen to stimulate interest and share project materials with other scholars whose research interests overlapped with ours. We found that this resulted in direct and significant benefits for the project. Some of these benefits were tangible: a number of scholars contributed an extra layer of proof reading and transcription checking, and contributed to the compilation of the bibliography.[13] Other valuable contributions came in the form of feedback and comments from members of the academic community in response to papers and online demonstrations presented at various academic conferences.[14]

[13] Dr Guzmán Mancho, University of Alcalá; Professor Kisei Sakemi, University of Hiroshima; Dr Míceál Vaughan, University of Washington.
[14] Particular mention should be made here of the comments from those attending the conference 'Manuscript Facsimiles: Fidelity or Betrayal?', University of Edinburgh, 2001.

The second major advantage of collaboration has been with regard to the long-term maintenance of the site and preservation of the edition. The site is hosted by the NLS, the same place where the manuscript is held. The project has therefore remained unaffected by the nomadic tendencies of individual researchers. Since 2000 the research team at Sheffield has dispersed but the NLS has provided a single stable point of contact and location for the edition.

At the time of writing we are in discussion with the Oxford Text Archive (OTA) regarding deposit of the textual output of the project.[15] This was not a requirement of the original funding award but is something that the project members subsequently decided was important for the scholarly community at large. The OTA are principally interested in providing solutions for the storage and dissemination of texts, which they display in TEI (SGML or XML) format or in ASCII. The Auchinleck texts accessed via the OTA site, then, will be in a different form to those on the NLS site. The OTA will not provide a full reproduction of the NLS web site, and only on the NLS web site will the texts be presented with the interface, folio images, graphics, style sheets, specialist font, and navigation options described above. As the OTA does not archive software or have its own search facility, they will not offer the ability to search the Auchinleck texts from their site. In the case of this project, then, the NLS and OTA will have different, complementary roles as long-term hosts. The NLS will host the site as a whole, complete with images and in a searchable, dynamic, interactive, and highly accessible form. The OTA will ensure the texts are visible to the academic community (and they will include a link to the NLS site) and will provide the texts in TEI Lite or ASCII for scholars or students wishing to process the texts in either of these formats.[16]

The process of having arrived at this particular solution to preservation, dissemination, and access, has raised two issues which are likely to be of relevance to future projects. First is the issue of the appropriate use of text encoding. The Auchinleck Manuscript Project did not originally produce texts in TEI Lite as advocated by the OTA.[17] It was not the format chosen by the NLS and it was not required to produce the Lexicon, to use the search software purchased by the project, or to fulfil other project goals. This does not present a problem for depositing the texts with the OTA. But it does require conversion of the texts into TEI and ASCII form, which will be done using a customised script. It is important that projects in the

[15] The Oxford Text Archive <http://ota.ahds.ac.uk/> is part of the Arts and Humanities Data Service (AHDS) <http://ahds.ac.uk/>.
[16] Dissemination of the NLS site has also involved registration with general search engines and directories (Google, Yahoo!) and subject specific gateways and portals (HUMBUL, the British Academy Portal, History Online, BUBL, The Labyrinth, Aedean).
[17] For a full description of TEI Lite see <http://www.tei-c.org/>.

early stages of planning are aware that this is something that is possible, but that it relies upon consistency in the original tagging, and is something which should be included in the project timetable and allocated necessary resources (either from within the project itself or by agreement with the OTA). The second issue concerns copyright and licensing. Cases can be envisaged of collaborative projects where academics are required to deposit the results of their funded research, but the library or other institution wants to retain ownership and control of the sources. The implications of this issue for individuals and institutions should be explored and established in the early stages of planning any collaborative endeavour with electronic outputs.