

THE INTERNET LIBRARY OF EARLY JOURNALS: PRESERVING AND DISSEMINATING THE BRITISH JOURNAL HERITAGE

Clive Field

Paper presented at the Third European Serials Conference, Ireland, September 1996 with the final section revised to reflect progress with the project between September 1996 and January 1997.

The ILEJ project is one of the principal outcomes of the Follett Report, and involves a consortium of university libraries. This paper examines the aims and methodology, and the progress to date.

Introduction

The Internet Library of Early Journals (ILEJ) project is one of some sixty projects being funded by the Joint Information Systems Committee (JISC) under the three-year £15 million Electronic Libraries (eLib) programme, one of the principal outcomes from the United Kingdom Joint Funding Councils' Libraries Review Group Report (the Follett Report) of December 1993.

The ILEJ project is one of only two focused journal digitization projects currently in train under the eLib programme, the other being DIAD (Digitization in Art and Design). In addition, JISC has recently established a general-purpose digitization bureau and consultancy service for the United Kingdom, based at the University of Hertfordshire. ILEJ itself is receiving £338,000 over three years for the project to cover hardware, software and staffing costs.

The ILEJ project is a consortium undertaking of the libraries of the Universities of Birmingham, Leeds, Manchester and Oxford. All are members of the Consortium of University Research Libraries (CURL) and the Research Libraries Group (RLG). Collectively their printed library collections comprise more than 15,000,000 items. The management of the project is being co-ordinated from Oxford, where scanning from microfilm and index-keyboarding will also take place. Servers are located at Oxford and Leeds, and scanning from hard copy is being undertaken at Birmingham and Manchester. All four institutions are represented on the project executive.

Aims and methodology

The original motivation behind the Follett Report's recommendation and eLib's funding of a digitization initiative in the United Kingdom was to save valuable storage space in university libraries. The hope was that by digitizing back-runs of commonly-held and preponderantly in-copyright journals many

Clive D. Field is Librarian and Director of Information Services The University of Birmingham, Edgbaston, Birmingham B15 2TT

libraries would be able to dispose of their hard-copy holdings and re-use the space saved for reader places, under acute pressure as a result of the rapid expansion of higher education student numbers in recent years, or other purposes.

The motivation behind the ILEJ project is very different. It is concerned with less commonly-held, more valuable and out-of-copyright journals, and the intention in digitizing them is not primarily to save space (although there may be an incidental benefit from this in some cases) but: (a) to improve availability of the journals by enabling researchers to access, search and manipulate data from them from the desktop, and (b) to reduce handling of the hard-copy originals, thereby aiding conservation. There is clear potential for digitization to replace preservation microfilming as an archival substitution medium for rare and valuable library materials.

The ILEJ project is being funded to provide a critical mass, upwards of 120,000 digitized images, of early journal literature. In that sense it is developing and supporting a live service rather than being a demonstrator or pilot project. At the same time, however, ILEJ will be controlling for and evaluating a wide range of technical variables in the digitization, retrieval and display processes with a view to recommending, at the end of the project, the most technologically viable and cost-effective strategy for scaling up the digitization of early journals.

The ILEJ project will, therefore, investigate from the perspective of speed, quality and cost the relative merits and demerits of:

- * scanning from hard copy versus microfilm of varying generations and polarities
- * bitonal (monochrome) versus 16/256 greyscale scanning (TIFF versus JPEG images)
- * scanning at variable resolutions
- * bit-mapped images versus optical character recognition (OCR) generated text
- * variable compression rates
- * indexing and retrieval from existing indexes, whether re-keyed from printed form or already in electronic format (and

used under licence), versus full-text searching with fuzzy matching

- * data transfer via ftp or tape
- * presentation via World Wide Web versus X-Windows servers in Oxford and Leeds respectively (PAT versus Excalibur Technologies EFS search engines, with access to EFS possible from Web browsers)

User evaluation will also be an important part of the project, because the acceptability and usability of the digitized outputs from the project will be critical in informing the future direction for larger-scale digitization. The evaluation will establish who is using the digitized images, with what frequency, and for what purpose; probe the acceptability to this user population of image quality, retrieval systems, presentation options, accuracy of OCRd text and content of the digitized journals; and particularly seek to determine the added value which digitized images offer over the hard-copy originals. For the purpose of evaluation, each of the four consortium members will establish focus groups of users drawn from their own and the 'new' university in their city. For more remote users the evaluation will be undertaken electronically.

Journal sample

The ILEJ project is concerned exclusively with early British journals. It is not intended to compete with other journal digitization projects in other parts of the world, for example JSTOR: The Mellon Foundation Journal Storage Project of American economic and history journals and the Australian Cooperative Digitisation Project (Ferguson), 1840-45. However, close contact will naturally be maintained with these other projects and experiences shared. As RLG members, the consortium is also obviously tracking RLG's own efforts in digitization (currently in the fields of sexuality and migration), which have British content relevance; one of the ILEJ partners (Leeds) is directly involved with the sexuality project ("Studies in scarlet").

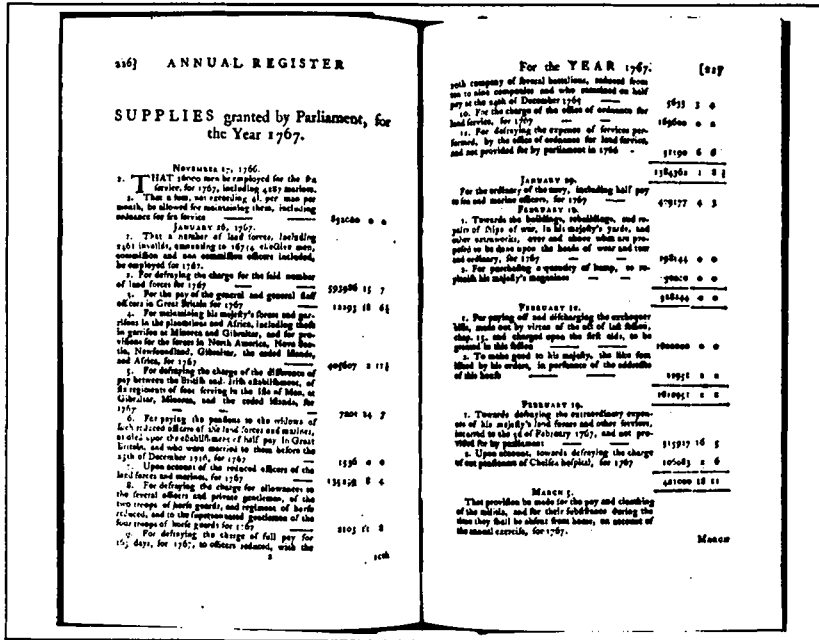
The ILEJ project seeks to create digitized versions of substantial runs (at least twenty consecutive years for each title) of six key early

British journals. Depending upon throughput, it is possible that additional titles may be added subsequently. The six core journals are Annual Register, Gentleman's Magazine and Philosophical Transactions of the Royal Society for the eighteenth century, and Blackwood's Edinburgh Magazine, The Builder and Notes and Queries for the nineteenth century. Brief profiles of these journals follow:

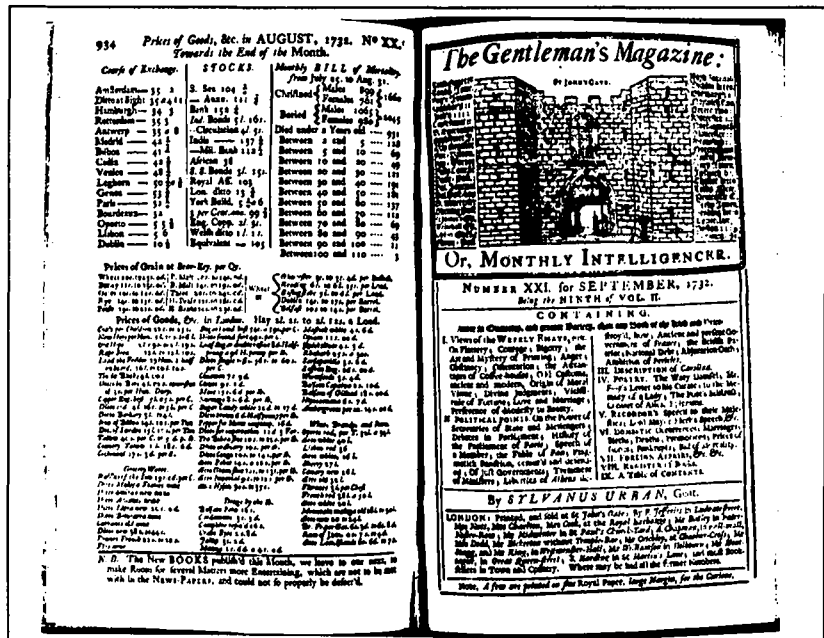
Annual Register: started in 1758, an annual survey of European and world events from a

British perspective, but including biographical notices, parliamentary and legal reports, and some book reviews, divided into topical sections with chronological sub-divisions

Gentleman's Magazine: started in 1731, a Britain-focused miscellany of information about people, places and events, including news summaries, parliamentary reports, biographies and obituary notices, poems, essays, and a register of current publications



Annual Register



Gentleman's Magazine

190 THE BUILDER. [APRIL 26, 1849.]

HALL AND STAIRCASE OF THE NEW BRITISH MUSEUM.
115 & 116, AVENUE.

THIS HALL AND STAIRCASE OF THE NEW BRITISH MUSEUM.
 Here is a representation of part of the Hall, in the British Museum, and the staircase, as the same affective portion of the museum.
 The artist is Charles Duncanson. The ceiling is a hollow brick ceiling over St. George's Hall, Liverpool. The hall is a grand hall, with a high ceiling, and is supported by large columns. The staircase is a grand staircase, with a high ceiling, and is supported by large columns. The drawing is a detailed architectural illustration, showing the texture of the walls and the arrangement of the architectural elements.

**BLACKWOOD'S
EDINBURGH MAGAZINE.**

No. COCCV. JULY, 1849. Vol. LXVI.

Miss Mortales.
No. II.
CHRISTOPHER UNDER CANVASS.

ENCAMPMENT AT CLANCH. TIME—EVENING, A.M.
SCENERY—The Portal of the Partition.
NORTH—BULLER—SEWARD.

I know there is nothing you dislike so much as personal observations—
 NORTH.
 On myself to myself—not at all on others.
 BULLER.
 Yet I cannot help telling you to your face, sir, that you are one of the
 finest looking old men—
 NORTH.
 Elderly gentlemen, if you please, sir.
 BULLER.
 In Britain, in Europe, in the World. I am perfectly serious, sir. You are,
 NORTH.
 You needed not to say you were perfectly serious; for I suffer no man to
 be ironical on Me, Mr Buller. I am.
 BULLER.
 Such a change since we came to Clanch! Seward was equally shocked,
 with myself, at your looks on board the Steamer. So lean—so bent—so
 hollow—so haggard—in a word—so aged!
 NORTH.
 Were you shocked, Seward?
 SEWARD.
 Buller has such a blunt way with him that he often makes me blush. I
 was not shocked, my dear sir, but I was affected.
 BULLER.
 Turning to me, he said in a whisper, "What a wreck!"
 VOL. LXVI.—NO. COCCV.

Blackwoods

The Builder

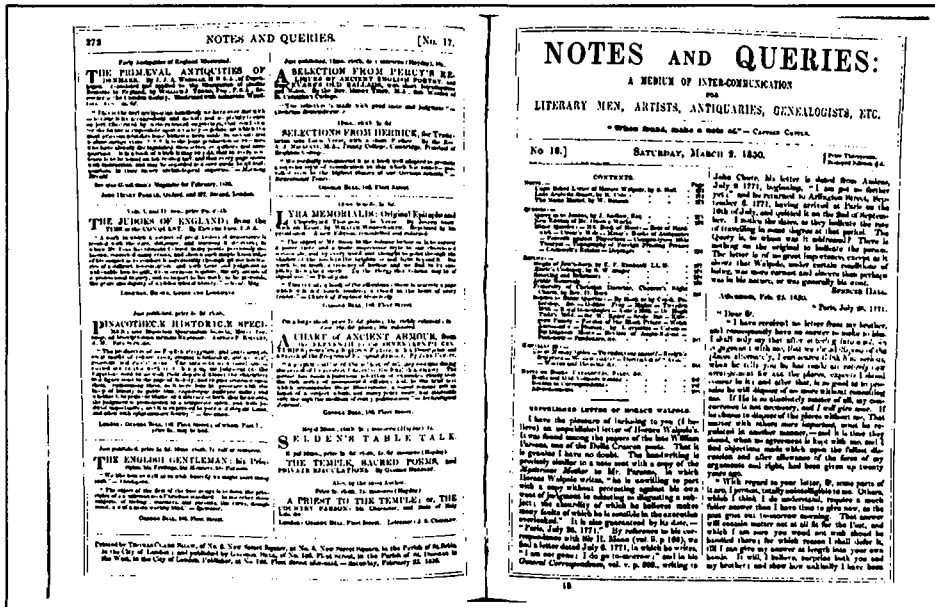
Philosophical Transactions

[329]

**PHILOSOPHICAL
TRANSACTIONS.**

**XXXIV. Astronomical Observations made at
Chislehurst, in Kent, in the course of the
Year 1772. By the Rev. Francis Wollaston, LL.B. F. R. S.**

Read, Mar. 16, 1774.
 MY having these two last winters communicated to this society, what astronomical observations I had occasionally made in the course of each year, seems to be a duty upon me to continue the same now. And I am the rather inclined to do so; because I could wish we were favoured with the correspondences of observations of all our worthy brethren, and therefore ought not to be backward to throw in my mite towards a general stock.
 My instruments and situation are the same as before described; and the following tables are in the same form as the last year. My clock has been kept
 Vol. LXIV. U a going



Notes and Queries

Philosophical Transactions of the Royal Society: started in 1660, initially as a forum for the publication of scientific papers of both a general and a specialized nature, although increasingly a learned journal carrying refereed papers from established scientists

Blackwood's Edinburgh Magazine: started in 1817 (as a Tory rival to the Whig Edinburgh Review), a medium for imaginative literature, publishing English poetry, essays and especially prose fiction, and pioneering the presentation of European literature (particularly German) to a British audience

The Builder: started in 1843, a mine of information on domestic and foreign building developments from the perspective of the architect, engineer, constructor and art historian, including accounts of new buildings, materials, processes and books, and articles on ancient monuments and other historic buildings

Notes and Queries: started in 1849, "a medium of intercommunication for literary men, artists, antiquaries, genealogists, etc.", carrying brief reports of completed research on humanities and related subjects and questions inviting answers in subsequent issues

These six titles were carefully chosen according to a set of inter-related criteria so as to create a critical mass of material which could be considered to be broadly representative of pre-1900 journals as a whole and which would

test the technological variables already identified. These criteria included:

- * wide subject range, covering science and technology as well as the arts
- * perceived user demand in the United Kingdom higher education sector
- * conservation priorities
- * diversity of typefaces and print and paper quality (variable show-through, foxing, page alignment, type-face weight)
- * short and long article formats
- * one-, two- or three-column page formats and variable page size
- * balance between text and illustrations (both line-drawings and half-tones)
- * availability of copies in the consortium libraries

Progress to date

The project effectively began during the first half of the 1995/96 academic session as soon as two half-time project staff had been recruited at Oxford and Leeds to augment the efforts of the original project executive (comprising two members from each of the four university libraries). During the first phase of the project a number of key tasks were successfully completed, including: drafting and signature of a formal memorandum of understanding to

regulate the operation of the consortium and the subsequent exploitation of the digitized images; a comprehensive page-by-page audit of the holdings of the target journals at Birmingham and Manchester, to highlight deficiencies in the copies and potential scanning problems; simulations with such hardware and software as were readily to hand in the libraries to gain practical experience of issues of image capture, transfer, compression, conversion to OCR, retrieval and display; and installation and configuration of the Oxford and Leeds servers. The actual choice of scanning hardware and software was to prove much more problematic, however, not least because, for reasons of conservation, flatbed scanning from hard-copy originals, whether dismembered volumes or not, had been rejected at the outset.

For scanning from hard copy five attributes were sought: overhead/cradle operation, to protect bound and fragile originals; speed; resolutions up to 600 dpi (for 'future-proofing', retaining maximum flexibility for the re-use of the images); software correction for defects in originals (for example, page curvature), to maximize image quality; and greyscale scanning for output to a PC. As yet, unfortunately, there is no scanner on the market known to the project which meets these criteria in full, and the project has had to compromise on the Minolta PS3000P scanner, two of which have been purchased, one each for Manchester and Birmingham. The PS3000P meets three of the criteria immediately but, although it already scans in greyscale, it will not be able to output greyscale images to a PC until at least Spring 1997 when the IMAX compression card is expected to be upgraded to PCI technology; nor can it currently handle resolutions beyond 400 dpi, which limits the future-proofing potential.

Similar difficulties with greyscales have arisen in respect of scanning from microfilm, which will be employed initially for a Mellon preservation microfilm run of the *Gentleman's Magazine* from Cambridge University Library. Although the Mekel M500XL-G greyscale microfilm digitization camera was identified very early in the project as potentially the most suitable scanner, and its suitability subsequently confirmed by test scans of microfilm of the *Gentleman's Magazine* undertaken by Zuma

Corporation for the consortium, the camera only went into full-scale production at the very end of 1996 (although the bitonal version has been around for some time), and the project does not expect to take delivery of its scanner until Spring 1997.

The non-availability of the necessary scanning hardware and software meant that the production phase of the project had to be delayed until October 1996 when it started, on Manchester's PS3000P, with the bitonal scanning at 300 dpi of *Notes and Queries*, whose typography was considered to be sufficiently non-problematical as not to require greyscale scanning. Through trial and error, the operator at Manchester has now devised a scanning methodology to optimize throughput and quality of image capture, and the technical staff at Leeds have identified a suite of software (currently Image Alchemy, ScanFix and Omnipage respectively) to handle image processing, cleaning and OCR. Although there have been quite a few problems at the OCR stage, where the system has a tendency to crash (seemingly related to an inability to OCR portions of text which, as a result of page undulation, contain broken characters), by January 1997 the first ten volumes of *Notes and Queries* (for 1849-54) had been successfully scanned and mounted on an experimental Web site, with up to 128 character and Boolean fuzzy-searching capability through EFS WebFile. At that stage the project felt sufficiently confident to publicize its imminent availability as a live service through ten mailbase lists in history, literature and related disciplines, and to appeal for remote users who would provide feedback.

As a condition of eLib funding access to the digitized images for the duration of the project will be available free of charge to the United Kingdom higher education community across the national academic network. Access by other users may be chargeable. The consortium intends to develop the project on a larger scale once eLib support has ended, possibly on a partnership funding model which will imply charging. Almost certainly, that partnership will involve one or more commercial publishers. A number (including Chadwyck-Healey) have already expressed interest in

principle in working with the consortium to expand the project and to link into their own electronic products.

At the same time, the ILEJ project will be keen to forge links with other emerging British initiatives in digitization, in order to explore common technological ground and any potential for the sharing of resources and expertise. Specifically, these links could be with the national digitization centre at the University

of Hertfordshire (of whose advisory group one of the ILEJ project executive is a member), the Knowledge Gallery hosted by De Montfort University, the British Library's Digital Library Development Project, and any developments arising from the National Preservation Office, of which CURL is a co-funder.

Up-to-date information about the project may be found on the ILEJ Home Page at:

<http://www.bodley.ox.ac.uk/ilej/>
