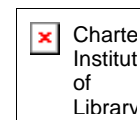


Information Scotland

The Journal of the Chartered Institute of Library and Information Professionals in Scotland



ISSN 1743-5471



[Issue contents](#) | [Journal contents](#) | [About the online edition of the journal](#)

April 2007 Volume 5(2)

Digitisation

Peak of achievement

Accounts of the first exciting explorations of the Scottish Highlands are being made accessible by digitising the Scottish Mountaineering Club Journal. Alan Dawson explains how ebook technology is making the results as easy to use as possible.

“Let thy words be few”

These are the first five words in the first issue of the [Scottish Mountaineering Club Journal](#), published in January 1890. Well over a century later, the club and its journal are still thriving, and their published words are far from few. The club has changed substantially over the years, and membership is now restricted to accomplished rock and ice climbers, but in its early years most members were simply men who climbed hills, who would now be regarded as hillwalkers rather than mountaineers. The early issues of the SMCJ therefore document an exciting period in exploration of the Scottish highlands, containing the first recorded descriptions of numerous Scottish hills and crags, as well as articles on geology, photography, deer forests, snow cover, gaelic names, aesthetics, physiology, equipment, and expeditions abroad.

The first six volumes (36 issues) of the SMCJ have recently been digitised at the Centre for Digital Library Research (CDLR), thanks to a grant of GBP 3,000 from the Scottish Mountaineering Trust. All 36 issues are now freely available to anyone, for personal or educational use, via the Glasgow Digital Library (run by CDLR).

Munros

“Surely it is better to follow a standard, even if occasionally wrong.”

(H. T. Munro, SMCJ volume 2 number 6, p330)

Issue 6 of volume 1 of the SMCJ contains the first publication of the tables that later became known as the Munros (Scottish mountains over 3,000 feet high). The list has been revised many times since, most recently in 1997, leading many to a call for a return to the original list compiled by Hugh T. Munro. However, inspection of the original tables shows why this would not be a good idea:

- The original tables contain no grid references

- The heights of numerous hills are now known more accurately than in 1891
- Numerous names and spellings now differ
- Munro was busy making revisions to his list, and published some corrections in the SM CJ in 1893

The first list of Munro is much larger. Although Munro did designate a subset of his 538 hills to be 'separate mountains', he clearly regarded the full set as the standard list, so it is not clear how the set of hills regarded as Munros has shrunk from 538 (all mountains) to the current 284 (separate mountains). Perhaps the answer is hidden away in one of those volumes that have yet to be digitised.

Methodology

The methods used at CDLR to create the online version of the SM CJ are similar to those used to create accessible and easily usable ebooks in HTML format rather than the more cumbersome and problematic PDF, or other proprietary format. This methodology is described in detail elsewhere,[\[2\]](#) but can be summarised as follows:

1. Capture text and images using scanner or digital camera
2. Convert text to machine-readable form, via OCR
3. Convert images from TIF (kept for preservation) to JPG
4. Assemble text for each issue into single Word document
5. Proofread text and apply structure using Word styles: headings, quotes, tables, notes, indexes etc
6. Insert references to image file names
7. Convert from Word to HTML (using a Word macro), retaining only text and structure, not formatting
8. Import all HTML files into an Access database
9. Generate web pages from database
10. Generate cumulative indexes to all issues from database
11. Publish generated pages on web server, with manually created stylesheet to control formatting
12. Add link to web pages and await visit from Google robots to add searchability

This methodology takes longer than producing facsimile pages in PDF or image format (especially step 5), but has many advantages. Web pages are relatively small, quick to load, Google-friendly, fully compliant with accessibility legislation, and viewable on any browser on any machine, with no plug-in software needed. Furthermore, each page has a different but precise HTML title, generated from the article title, author and date.[\[3\]](#)

The methodology also adds value to the paper publications, rather than merely digitising them. Although the collection will be searchable and discoverable via Google, the use of accented characters and the spelling variations of proper names mean that browsing is at least as effective as searching. It is therefore important for the indexes to function effectively across the whole collection, not just within an issue or volume. This has been achieved in two ways: by converting the original indexes to each bound volume into a single cumulative index, with links to the specific issue and relevant page; and by adding further indexes that do not appear in the paper version, such as indexes of authors, events, illustrations, places and reviews. All index entries are stored in the Word documents along with the text, so that creation of index pages can be fully automated.

Many of the issues that arise from using this methodology have been addressed and resolved in earlier work on ebook creation.[\[4\]](#) For example, policies are needed on error correction, punctuation, capitalisation, image placement, footnotes, character sets, etc. The aim is to strike a balance between access and preservation by faithfully capturing the content and structure of the original work without having to preserve typesetting or artefacts of the printing process, so that the end result is highly accurate but can take advantage of current styles and standards.

Development of the ebook methodology to make it applicable to a journal, including creation of the cumulative indexes, has made the SMCJ a useful focus for digital library research and development, as well as being valuable historical content.

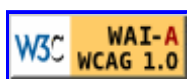
Setbacks

The process of producing the online SMCJ has been far from smooth. Copies of the paper journals had been borrowed from the SMC library in Glasgow, but this closed due to building renovation and sale, so issues had to be located elsewhere or borrowed privately. The project then ran out of funding halfway through proofreading of volume 4. Yet these were minor issues compared to the death of Rob Milne on Everest in May 2005. Rob had been SMC publications manager and steered the digitisation proposal through the SMT committee. Shortly after Rob's death the SMC librarian, Ian Angell, was desperately unlucky to fall into a rock crevasse descending from Ben Donich near Arrochar, and he too was killed. The project therefore took longer than originally envisaged, but it was also important not to abandon it. The names of Rob Milne and Ian Angell deserve to be credited and remembered alongside those heroes from a much earlier generation of mountaineering in Scotland, whose recorded exploits are now readily available to all.

Alan Dawson is Senior Researcher/Programmer, Centre for Digital Library Research, University of Strathclyde.

References

- 1 For a concise summary of revisions see Statistical Topics in Hillwalking, by Chris Crocker and Graham Jackson: www.biber.fsnet.co.uk/
- 2 The ebook methodology project report and toolkit is available from the Arts and Humanities Data Service: <http://ahds.ac.uk/collections/ebook-methodology/>
- 3 For more details of this technique see Optimising Metadata to Make High-value Content more Accessible to Google Users, by Alan Dawson & Val Hamilton, 2006: <http://cdlr.strath.ac.uk/pubs/dawsona/ad200503.htm>
- 4 See Twenty Issues in ebook Creation, by Alan Dawson & Jake Wallis, 2005: <http://cdlr.strath.ac.uk/pubs/dawsona/ad200501.htm>



[Information Scotland](#) Vol. 5(2) April 2007

© Chartered Institute of Library and Information Professionals in Scotland

[Disclaimer](#)

Information Scotland is delivered online by the [SAPIENS](#) electronic publishing service based at the [Centre for Digital Library Research](#). [SLAINTE](#) (Scottish libraries across the Internet) offers further information about librarianship and information management in Scotland.

Last updated: 19-Jun-2007