

INTERNET ARCHIVE
WayBackMachine

120 captures
15 May 07 – 20 Oct 14

http://www.ariadne.ac.uk/issue51/emamy-cameron/ Go

MAY JUL AUG
1
2006 2007 2008

ARIADNE

Issue 51 April 2007

Main Articles

Citeulike: A Researcher's Social Bookmarking Service

[Kevin Emary](#) and [Richard Cameron](#) describe a tool which assists researchers gather, collect and share papers.

Main Contents

Section Menu

Email Ariadne

Search Ariadne

Introduction

This article describes Citeulike, a fusion of Web-based social bookmarking services and traditional bibliographic management tools. It discusses how Citeulike turns the linear 'gather, collect, share' process inherent in academic research into a circular 'gather, collect, share and network' process, enabling the sharing and discovery of academic literature and research papers.

What is Citeulike?



Figure 1. Citeulike front page

Citeulike is a Web-based tool to help scientists, researchers and academics store, organise, share and discover links to academic research papers. It has been available as a free Web service since November 2004 and like many successful software tools, it was written to solve a problem the authors were experiencing themselves:

'Collecting material for a bibliography is something which appeared to require an amazing amount of drudgery....So, the obvious idea was that if I use a web browser to read articles, the most convenient way of storing them is by using a web browser too. This becomes even more interesting when you consider the process of jointly authoring a paper.'^[1]

The basic functionality of the tool is simple; when a researcher sees a paper on the Web that interests them, they can click a button and have a link to it added to their personal library.

The screenshot shows the Citeulike web interface. At the top, there are browser tabs and a navigation bar with the Citeulike logo and user information (Logged in as camster). Below the navigation bar, there's a search bar and a list of flags. The main content area is titled 'Where would you like to file it?' and contains a form for posting a new article. The form includes fields for Title, Authors, Tags, Subject, and Priority. The Title field contains the text 'The pathology of embryo death caused by the male-killing Spiroplasma bacterium in Drosophila nebulosa'. The Authors field contains 'Bentley J, Veneti Z, Heraty J, Hurst G'. The Tags field contains 'Spiroplasma selective pathology'. The Subject field is empty. The Priority field has radio buttons for 'Top priority!', 'I really want to read it', 'I will read it', 'I might read it', 'I don't really want to read it', and 'I've already read it'. The 'I might read it' option is selected. There is a 'Post Article' button at the bottom right of the form. On the left side, there is a 'Navigation' menu with links for Home, Discussion list, Your Library (View articles, Recent, To read), Authors and tags, Search, Post article, Manually post article, Export to BibTeX/EndNote, Change password, and Your Watchlist (View watchlist). On the right side, there is a 'Click to add tag' section with a list of tags including 1931, acquisition, algorithms, alive, anaesthesia, api, artificial, band, bands, barabasi, basal-ganglia, bastards, bibtex-import, biogenetics, biography, blast, blog, blogs, boolean, brain, bugs, bulgakov, c-plus-plus, calcium, canalizing, cellular-automata, cfo, and circadian.

Figure 2. Posting page

When a user posts a paper, Citeulike automatically extracts the citation details and stores a link to the paper, along with a set of user-defined tags. The user is then returned to the original Web page, where they can continue reading.

Citeulike has a flexible filing system, based on the tags [2]. Tags provide an open, quick and user-defined classification model that can produce interesting new categorisations.

'the beauty of tagging is that it taps into an existing cognitive process without adding add much cognitive cost. At the cognitive level, people already make local, conceptual observations. Tagging decouples these conceptual observations from concerns about the overall categorical scheme.' - Rahshmi Sinha [3]

By tagging papers they post, users are building a domain-specific 'folksonomy' that describes the paper they are bookmarking in terms that are meaningful to themselves and usually other specialist researchers (in Citeulike's case).

Because everyone's library is stored on the server, it is accessible from any computer, enabling users to share their link library with others and see who else has bookmarked the same papers (their Citeulike 'neighbours'). They can then click through to see the rest of these other users' libraries and in this way discover literature that is relevant to their field but of which they may have been unaware. Tags also provide another simple mechanism whereby users can navigate the libraries and discover new papers.

RSS feeds and Watchlists allow users to track tags and users' libraries that interest them, showing the latest additions to these chosen categories.

As well as browsing their neighbours' tags and libraries, users can discover papers on the Citeulike front page where the latest papers that have been posted are displayed (see Figure 1. above).

Another point of discovery is the set of subject specific pages, where the latest links to papers posted are displayed according to the subject under which users have classified them. This is currently a simple closed classification consisting of Computer Science, Biological Science, Social Science, Medicine, Engineering, Economics/Business, Arts/Humanities, Mathematics, Physics, Chemistry, Philosophy and Earth/Environmental Science.

Within these categories, as well as display of papers by latest addition, there is a voting system whereby users can vote on a particular paper that they find interesting, resulting in that paper's promotion up the list of latest papers.

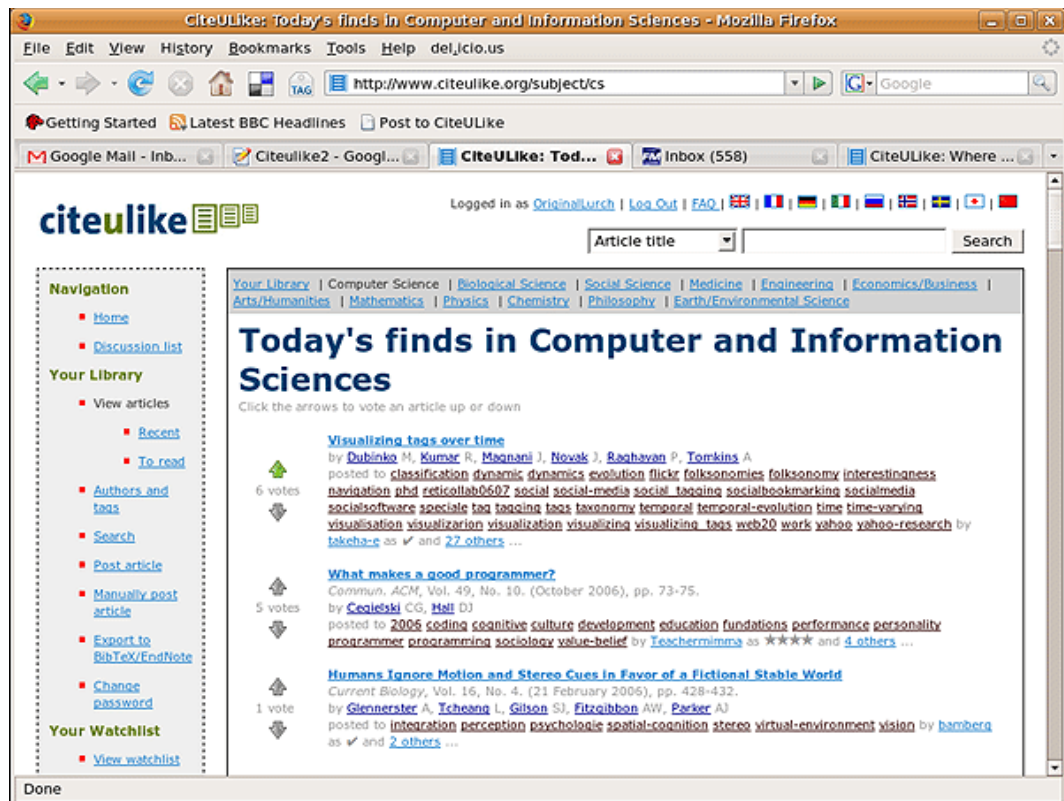


Figure 3. Latest papers in Computer Science

The emerging dataset of papers, tags and the relations between the two offers many intriguing avenues for investigation and data mining. Clusters of tags can be investigated for patterns, and papers with similar tags can be grouped and relations between them exposed. There are already several independent projects under way to produce analyses from the Citeulike dataset.

It is important to note that the tagging is initially done for the individual user's personal benefit and the community benefits arise as a consequence of this behaviour. Having said that, it is also clear that contributing to the community, or at least to a group, is also an important part of the motivation of many users (which can create a possible dichotomy in the choice of tag words (personal vs generic)).

Gather, Collect, Share

Citeulike fuses together two separate categories of software: the new 'Web 2.0' breed of social bookmarking services (del.icio.us [4] etc) and traditional bibliographic management software (EndNote etc). While Web bookmarks are simple URLs, citations are a bit more complex and include metadata like journal names, authors, page numbers etc.

The gather-collect-share model found in traditional bibliographic software is a linear process. **Gathering** literature is conducted by querying an OPAC database or a scientific publisher's site using a Web browser.

Desktop software such as EndNote will then allow the user to **collect** the articles which he or she wishes to keep for future reference, and the collection process stores sufficient metadata for the article (the title, authors, journal name, page number) in a format which allows for its ultimate **sharing** with others by citing it in the author's own publication.

When the publication appears in print, the whole gather-collect-share process starts again with a different researcher in a different institution.

Citeulike fulfils two roles:

Firstly, it makes the existing model of collecting information easier for the end-user. A Web browser is the natural tool for exploring lists of publications, and our premise was that it ought equally to be the natural tool with which to collect bibliographic records. Social bookmarking services such as del.icio.us allow the user to store links to Web pages in an online account - all at the click of a button.

On the other hand, academic users have traditionally had to switch back-and-forth between Web browser and external application as they alternate between gathering and collecting modes. This process is time-consuming and error-prone.

Citeulike solves this problem by operating like a standard social bookmarking service (the user clicks a bookmarklet in order to post an article to his or her account), but it also extracts all the relevant metadata required to create a proper bibliographic record automatically from the publisher's site. Citeulike supports most of the major publishers [5], and the 'gather' and 'collect' steps of the process work seamlessly for the user without having to leave the Web browser, with none of the drudgery traditionally associated with keeping one's personal bibliographic database up to date.

Gather, Collect, Share, Repeat

The second role fulfilled by Citeulike is that it has actually changed the traditional method for discovering and sharing information. The

linear gather-collect-share process has turned into a virtuous circle. Because users' collections are now stored on a Web server rather than in a proprietary bibliographic database locked away on a desktop computer, it is now possible for users of Citeulike to browse each other's collections.

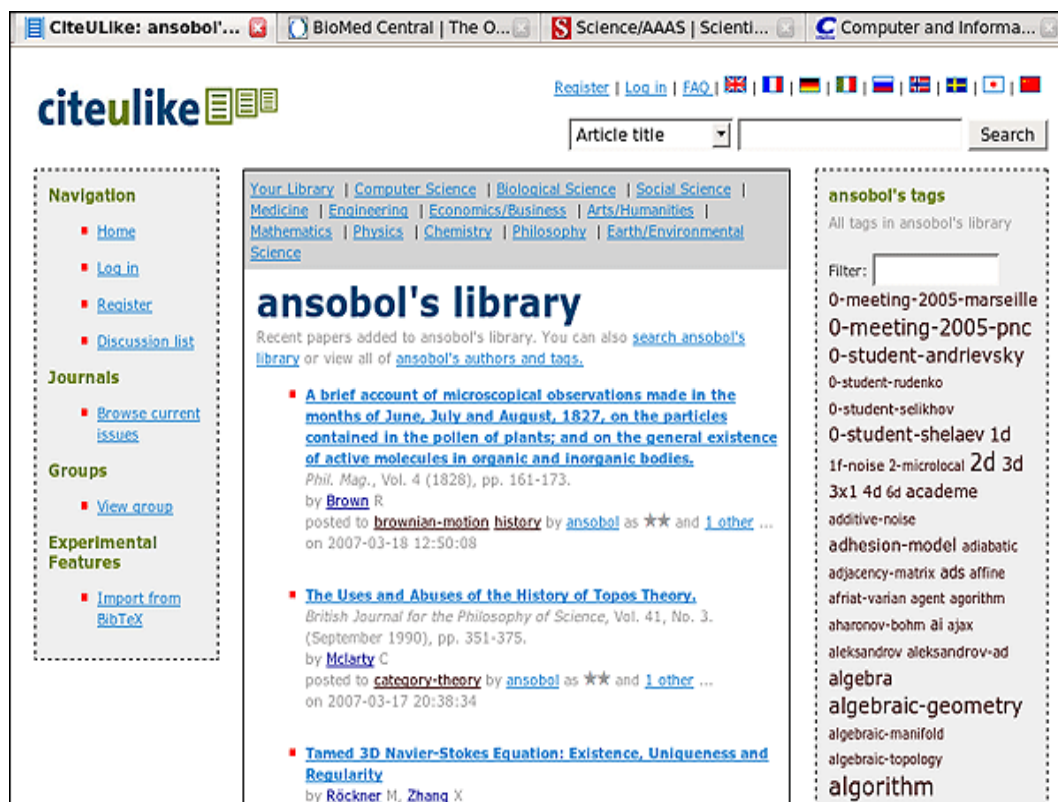


Figure 4. Browsing a user's library

Users of Citeulike can browse or search through collections of articles bookmarked by other people with similar interests.

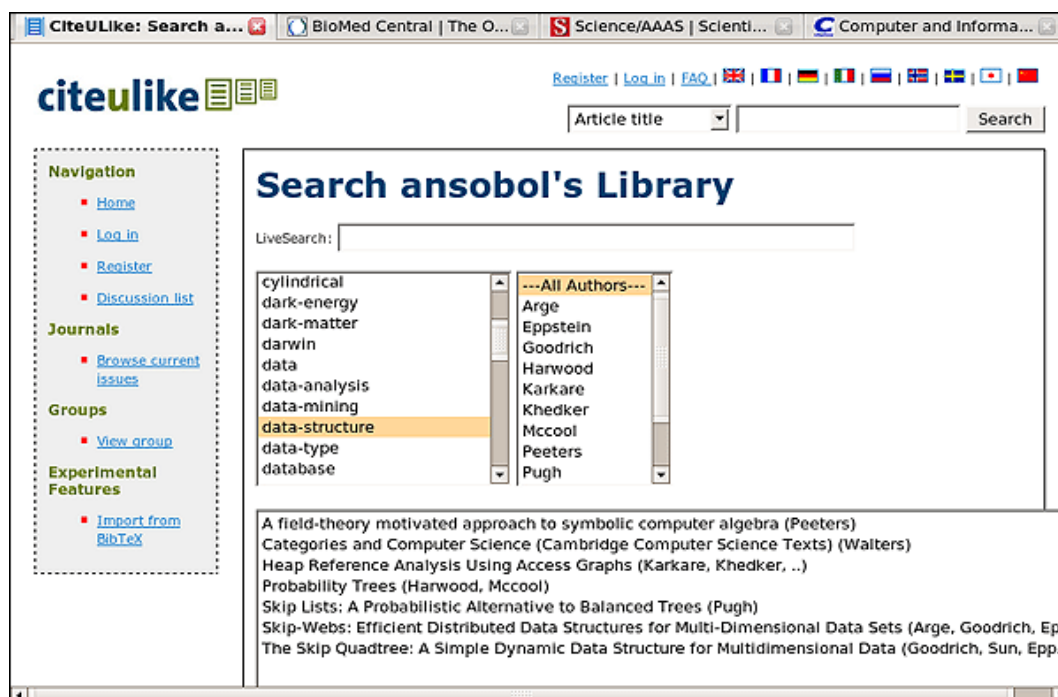


Figure 5. Searching a user's library

Due to its specialisation in a particular niche (only catering for academic articles), Citeulike has value to researchers beyond a service aimed at a generalised audience. Within this niche, new papers are more easily discovered, relevant clusters of interest are naturally formed and the tags are likely to be meaningful to users.

The tagging within a niche is also more specific. For example, a tag consisting of the term 'evolution' applied to the World Wide Web as a whole could correspond to many possible interpretations of the word. Within the context of peer-reviewed articles, the scope of such a tag is likely to be much narrower, and users searching on that term will retrieve many more targeted results.

An interesting question arises: how far should this specialisation be taken? Is there a requirement for a separate bookmarking service for each individual academic discipline or sub-discipline?

It could be argued that a single unified bookmarking system filtered by tags is the answer to this problem, but the utility of a specialist service versus a generalised one as demonstrated by Citeulike weighs against this. Separate services would not be of benefit to cross-disciplinary fields or users, or indeed the discovery of relations between aspects of separate disciplines. Perhaps in choosing the overall category of academic research as a specialisation Citeulike has fallen naturally into a rational balance in this regard.

As noted above, Citeulike partly addresses this issue with the user classification of papers into subject areas (primarily to enhance discovery). At the time of writing, we believe that Citeulike is the only social bookmarking service that has attempted to combine tagging with a closed classification like this (albeit a simple one).

It is worth noting that, from users' point of view, it is far more practical to have a single social bookmarking service to store all their research papers, from whatever source (rather than several services). This is also true from the sharing and collaborative point of view. For this reason, we would argue that journal publishers and database providers should link to services like Citeulike in order to provide bookmarking functionality for their subscribers. Additionally, the natural sharing and discovery of papers amongst users on Citeulike has obvious promotional benefits for the output of publishers. Because Citeulike is an independent service, papers posted to Citeulike are more likely to reach an audience beyond a particular publisher's natural constituency.

Gather, Collect, Share, Network

As already noted, a further consequence of the "everything in your browser" model of Citeulike is that users will inevitably discover others with similar interests. The fact that two users read similar literature probably indicates that they will potentially have a professional interest in each other. The bibliographic data forms a *fabric* binding people together.

Professional networks build up between researchers in the same field. Rather than requiring the facility to chat to friends, researchers welcome tools which let them carry out tasks associated with their work collaboratively. To further serve that end, Citeulike provides group allowing people who already have working relationships and are, say, collaborating on a publication to share their bibliographic database

Additionally, it allows for globally distributed researchers with a common theme to build up a shared collection of literature they find relevant.

Publisher-specific Data

As well as browsing Citeulike themselves, publishers can obtain feeds from the database of tags being used by Citeulike users of their particular journals.

The screenshot shows the Citeulike web interface. At the top, there's a navigation bar with links like 'Register', 'Log in', 'FAQ', and a search bar. Below this, a sidebar on the left contains 'Navigation' (Home, Log in, Register, Discussion list), 'Journals' (Browse current issues), 'Groups' (View group), and 'Experimental Features' (Import from BibTeX). The main content area is titled 'Tag bioinformatics' and lists recent papers classified by this tag. Three articles are visible, all from the journal 'Science'. The first article is 'Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families' by Lockless SW and Ranganathan R. The second is 'The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease' by Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, and Golub TR. The third is 'COMPUTATIONAL BIOLOGY: Bioinformatics—Trying to Swim in a Sea of Data' from Science, Vol. 291, No. 5507. On the right, there's a section titled 'Everyone's Tags' showing a list of tags like 'adaptation', 'algorithm', 'analysis', etc., with a filter input.

Figure 6. A list of articles from the journal 'Science' which have been tagged 'bioinformatics'

Publishers can encourage the sharing of papers on Citeulike by adding a post to Citeulike link at the article level on their publications. Alongside the link, publishers could also choose to display the tags used and number of users tagging at the article level on their sites.

Architecture and Future Directions

In terms of technical architecture, the software is built on PostgreSQL, Tcl and Memcached. The database and Web servers reside on scalable, redundant, professionally hosted Linux servers and the database is backed up every 15 minutes. The design of the Citeulike Web site is simple, clear and functional. This should not be lost through careless addition of excessive extra functionality.

Unlike other services, Citeulike is remarkably free of spam links and the technical design decisions that have prevented spammers' invasions will continue to be a focus of activity.

The current development schedule includes building on the existing group's functionality, to make it easier for users to separate personal and group associated bookmarks and extending and improving on the tagging tools by introducing things like tag bundling which would give a further level of user-defined classification and enable mass tag-editing operations. Private bookmarks are a feature request that has been resisted so far in order to keep with Citeulike's community-orientated philosophy, however there are probably good reasons why certain researchers wish to keep their bookmarks private and this is under review.

Using open source components, it costs surprisingly little to build and run Web services today which in the past would have required millions of pounds of investment, and this is surely a trend that will continue. Citeulike benefits from this trend; however the authors do intend to make it a self-sustaining resource. There are a number of ways in which this could be achieved however they all depend on a continued expansion of the user base, which is where efforts are concentrated at the moment. Citeulike has grown virally through word of mouth to its current size: 33,000 users generating 200,000 distinct visits per month, (see [Current Statistics](#) below). The network effects that created this scale continue to accelerate. The obvious and least intrusive way to promote its use is through library and information management professionals, as well links from publishers alluded to above.

The most intriguing area for future experimentation is mining the tag and article data that is being created. Is it possible that large-scale datasets from bookmarking and tagging can be used to supplement traditional peer review and citation analysis? This is a hard problem to solve, but there must be some implicit crowd knowledge in the patterns formed. Dario Taraborelli has written a thought provoking post on this:

'Collaborative metadata cannot offer the same guarantees as standard selection processes (insofar as they do not rely on experts' reviews and are less immune to biases and manipulations). However, they are an interesting solution for producing evaluative representations of scientific content on a large scale.' [6]

Conclusion

Citeulike is a tool that has gained a significant audience in the academic community. Through helping users keep track of their own bibliographies, it naturally creates an environment that facilitates sharing and consumption of academic literature. Publishers can encourage Citeulike's use amongst their readers, thereby benefiting from enhanced exposure for their content and greater user engagement with content. Many are doing this by placing posting links at the article level on their content, and will soon be displaying statistics as well as popular tags from Citeulike on their sites.

Insights can be gained through analysis of the emerging dataset of tags, and given a sufficiently large dataset, supplemental forms of discovery and rating of scientific literature could emerge.

Ultimately Citeulike works because it is useful to its users. It automates a repetitive bibliographic management task and it offers a complementary alternative to search engines and databases of academic literature through socially mediated retrieval and discovery of papers.

Current Statistics

As of 13 March 2007, Citeulike currently has 33,000 registered users and is gaining new registrations at the rate of 100 per day (up from per day 6 months ago). Of that 33,000, 45% go on to post articles to the site, many simply 'lurk' (i.e. browse other users' libraries but do not post themselves), and some disappear.

Citeulike receives in excess of 200,000 distinct visits (defined by Google Analytics as a set of page views by a unique user with a timeout after 30 minutes of inactivity) per month, with each visit generating an average of 2.77 page views. Of that 200,000 around 40,000 are visits from unique users who have previously visited the site on multiple occasions.

There are currently 505,402 items posted in the database (counting n if n people post the same article); 1,676,130 tags (counting n if there are 'n' tags applied to an article); and 130,548 distinct words used as tags. These numbers are growing exponentially.

There are over 800 user-created special interest groups.

Citeulike has an international audience and has been translated (by enthusiastic users) into 8 different languages including Japanese and Chinese (the largest single group of users by country is the USA).

References

1. Citeulike FAQ <http://www.citeulike.org/faq/all.adp>
2. Wikipedia entry on Tags <http://en.wikipedia.org/wiki/Tags>
3. Rashmi Sinha blog post 'A cognitive analysis of tagging' http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html
4. del.icio.us, the grandfather of social bookmarking sites: <http://www.del.icio.us>
5. Supported sites:
AIP Scitation; Amazon; American Chem. Soc. Publications; American Geophysical Union; American Meteorological Society; Anthrosource; Association for Computing Machinery (ACM) portal; BMJ; BioMed Central; Blackwell Synergy; CSIRO Publishing; CiteSeer; Cryptology ePrint Archive; HighWire; IEEE Xplore; Ingenta; IngentaConnect; IoP Electronic Journals; JSTOR; MathSciNet; MetaPress; NASA Astrophysics Data System; Nature; PLoS; PLoS Biology; Physical Review Online Archive; Project MUSE; PubMed; PubMed Central; Science; ScienceDirect; Scopus; SpringerLink; Wiley InterScience; arXiv.org e-Print archive.
6. Dario Taraborelli (2007). 'Soft Peer review, social software and distributed evaluation'
<http://www.academicproductivity.com/blog/2007/soft-peer-review-social-software-and-distributed-scientific-evaluation/>

Author Details

Kevin Emary

Citeulike

Email: kevin@citeulike.orgWeb site: <http://www.citeulike.org/>**Richard Cameron**

Citeulike

Email: Richard@citeulike.orgWeb site: <http://www.citeulike.org/>[Return to](#)

Article Title: "Citeulike: A Researcher's Social Bookmarking Service"

Author: Kevin Emary and Richard Cameron

Publication Date: 30-April-2007 Publication: Ariadne Issue 51

Originating URL: <http://www.ariadne.ac.uk/issue51/emamy-cameron/>[Copyright and citation information](#) File last modified: Wednesday, 09-May-2007 20:05:56 UTC[Main Contents](#)[Section Menu](#)[Email Ariadne](#)[Search Ariadne](#)

Ariadne is published every three months by [UKOLN](#). UKOLN is funded by [MLA](#) the Museums, Libraries and Archives Council, the [Joint Information Systems Committee \(JISC\)](#) of the Higher Education Funding Councils, as well as by project funding from the JISC and the [European Union](#). UKOLN also receives support from the [University of Bath](#) where it is based. Material referred to on this page is [copyright Ariadne \(University of Bath\) and original authors](#).