The need for an academic social bookmarking service

A report for the Joint Information Systems Committee
Acknowledgements

UKOLN is funded by the Joint Information Systems Committee (JISC) of the Higher and Further 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.
Table of Contents

1. Executive summary 5
2. Introduction 5
3. Social bookmarking overview 6
   3.1. Potential benefits 6
   3.2. Potential barriers 7
4. Review of existing initiatives 10
   4.1. Social Bookmarking services for the Research Community 10
   4.2. Content analysis of third-party social bookmarking services 12
   4.3. JISC activities related to personalisation and tagging 13
       4.3.1. Studies of personalisation in the JISC Information Environment 13
       4.3.2. Enhanced Tagging (EnTag) project 14
       4.3.3. PERonalised Tagging Information Services (PERTAINS) 15
5. Discussion points 15
   5.1. Sustainability 15
   5.2. Role of third-party services 17
   5.3. Tagger motivation 18
   5.4. Tag quality 18
   5.5. Building communities 19
6. The need for an academic social bookmarking service 20
   6.1. Benefits 20
   6.2. Barriers 21
7. Conclusions 22
8. Recommendations 23
9. References 24
A. Social bookmarking service overview 26
   A.1. Selected service summaries 26
       A.1.1. Scientific papers — Connotea 26
       A.1.2. Scientific papers — CiteULike 27
       A.1.3. Scientific research — Mendeley 28
       A.1.4. Webpages — Delicious 29
       A.1.5. Photographs — Flickr 30
       A.1.6. Books — LibraryThing 31
       A.1.7. Blogs — Technorati 32
   A.2. Language-specific and specialist sites 33
       A.2.1. The Francophone Web 33
       A.2.2. The German-speaking Web 33
       A.2.3. The Dutch-speaking Web 34
       A.2.4. The Spanish-speaking Web 35
1. Executive summary

Currently there is no single academic bookmarking system serving the UK research community. This report aims to investigate the potential barriers and benefits of developing such a system within the UK. At present, the activity around social bookmarking is happening in three broad areas. The first area is the development of social tagging and the enhancement of tagging within existing systems used by researchers in the UK. This work has been undertaken by a variety of projects funded by JISC and builds on popular, well-used services such as COPAC. The second area is the use of general social bookmarking systems such as Delicious and Flickr which are not designed specifically for research use, but have been adopted by researchers and are widely used throughout the world. The third area is the use of third-party social bookmarking systems developed specifically for the needs of the research community, such as CiteULike and Mendeley.

The possible need for a single social bookmarking system is investigated within the context of the use researchers make of all three of these areas. Each area has strengths and weaknesses, and, perhaps just as importantly, each area also has engagement by researchers with the systems they offer. Any new social bookmarking system would need to address the potential risk of attracting users away from their current preferred systems. This may not be down to functionality and how fit for purpose a system is for the researcher: other factors such as familiarity, loyalty to a ‘brand’ and reluctance to change will also, potentially, impact on the uptake of a new system.

In addition, the workflows actually available in mainstream tagging systems allows for private use—that is, saving bookmarks to a personal library within the system without sharing resources and knowledge with the wider community. Instead of perceiving a social bookmarking system as a way of sharing, some users perceive the social bookmarking system as a personal content management system, creating a closed library of their own resources. This approach doesn’t necessarily negate the value of a single system, but it is worth considering how much of the potential benefits of a UK-wide community of researchers would actually be realised.

The possibility of addressing the existing landscape by developing useful tools and applications to integrate with existing systems is also examined. Potentially, this would negate some of the barriers of using third-party systems, such as preservation, archiving, ownership of data and moving data around between systems, all potential problem areas when using third-party systems over which the research community has no control. The barriers here can equally apply to commercial products and non-commercial products, and if it is not possible to prevent users from using third-party systems, perhaps it could be useful to support this usage in a way which addresses the practical issues and, potentially, negates some of the highest risks associated with their use.

2. Introduction

Collaborative tagging tools are increasingly being used to support the sharing and management of Web-based resources. The use of tools like these is generally referred to as social bookmaking or social tagging because users can share the links that reference them together with metadata in the form of user-generated tags. These tags may in some circumstances collectively become (or exhibit) a kind of shared vocabulary known as a ‘folksonomy’, a term first coined by Vander Wal,
2007. The concept of a ‘folk taxonomy’ is much older, and refers to the way in which groups, operating in the vernacular, describe aspects of their environment.

Be this as it may, collections of tags can be used to support enhanced access to resources, e.g., by users browsing through hyperlinked lists of tags or visual representations of them known as ‘tag-clouds’. Social bookmarking tools are one of a class of (so-called) Web 2.0 technologies that potentially support the building of communities around resource types or disciplines of practice.

This report briefly introduces social bookmarking tools, reviews a range of existing activity in the higher education domain, and provides advice to the Joint information Systems Committee (JISC) on the potential need for an academic social bookmarking service in the UK.

3. Social bookmarking overview

Social bookmarking refers to a variety of mechanisms by which a community of practice may share stored hyperlinks, alongside descriptive metadata, frequently including user-contributed tags. Not all uses of shared bookmarking are social, as the technology has benefits for personal information management as well as knowledge management within and between groups. There are many potential uses for the technology, which has been successfully applied to things like overlay journals and the sharing of reading lists. In the higher education context, social bookmarking tools have been used in support of both teaching and research. Examples include the development of frameworks that permit the sharing and annotation of educational resources (e.g., Churchill et al., 2009; Minguillón, 2010) and the use of third-party tools that support the management of bibliographic references and other kinds of research output (Hull et al., 2008, pp. 8–10). Academic libraries have also explored the potential applications of social bookmarking, e.g., to add interactivity to library OPACs or collaborative functionality to reference services (e.g., Redden, 2010).

Potential benefits and barriers need to be viewed in the context of the different and distinctive ways social bookmarking can be executed, some of which do not utilise the sharing (e.g., ‘social’) element of the tools. Thus, the benefits vary depending on the actual workflows used. Essentially, social bookmarking networks offer three workflows:

1. Creating tags but not sharing, which gives a flexible, personal content management system
2. Viewing tags created by others, which offers enhancement of personal knowledge and an expansion of individual resources
3. Sharing tags, which contributes to the wider knowledge pool

3.1. Potential benefits

These fall into the distinct groups, as outlined above. Potentially, by tagging online resources and making those tags available to other users within a social networking environment, users open up their own expertise to others within the network, thus allowing them to learn from the choices made by someone with the same interests. This can be in many kinds of learning relationships, e.g., peer-to-peer, where the users have the same level of knowledge in a subject, tutor/student, where a user with more knowledge is sharing information with those who have less
knowledge, and cross-subject sharing, where experts in specific fields can share overlapping and related knowledge with each other. For the user who creates to share, the social bookmarking environment can become a dissemination channel. For the user who accesses the resources tagged by others, the same environment becomes a place to learn and gather relevant information. The benefits of this pattern of use depend heavily on users sharing the tags they create, which is not mandatory within such networks.

Again, based on the principle of users sharing their tags, social bookmarking tools can be used to generate descriptive metadata in the context of the World Wide Web where available content is so great that traditional indexing approaches are invaluable. The sharing aspects mean that, potentially, a trusted broader and deeper pool of resources and knowledge can be accessed in a more accurate/faster way by individuals than would otherwise be possible online. In short, user-generated tags can enhance information retrieval (e.g., Morrison, 2008).

Trusted sources (i.e., individuals who demonstrate to their peers that they contribute bookmarks to useful and relevant material in a given subject area) can contribute to an online, shared content management system on, potentially, a world scale that can be accessed easily by many. Provenance is generally a significant characteristic of social services, and microblogging and social bookmark sharing services are no exception.

Using the social bookmarking tools without actually sharing tags is another workflow within such environments. Potentially, this offers a user a flexible, personal content management system that is easily accessible, online, without being confined to a particular machine or hard drive. In addition, users have the option of viewing the tags of others and adding trusted resources to their personal collection.

3.2. Potential barriers

Some of the barriers outlined below are more likely to occur in a third-party services, specifically those centred around ownership of content. Other barriers are potentially inherent in all services of this nature, and are focused more on specific workflows or technical issues. As with the potential benefits, the risk of some, though not all, of these barriers is dependent on the specific workflow followed by a user.

Sustainability over time is an issue for both third-party and internal services. Commercial/third-party services may cease to exist, be bought out or sold on, depending on business models, which would result in a loss of access to the information stored. Non-commercial services, run from within the HE community, may cease to be funded, which would result in a loss of access to the information stored.

Transfer of the personal data stored by an individual on such services can be problematic (export of data, migration from one service to another). There is no generic model for social tagging and systems often have differing ideas of what is permissible, so information can be lost when moving between services (i.e., multi-word tags versus space as separating character, capitalisation, special characters, etc.). The process may well incur losses.

Ownership of the bookmarks stored on a service may not lie with the individual who uses the service. In general, ownership is not signed entirely over to the service (the Berne convention means that anything created by an individual is, by default, owned by them)—but services may well, specify on registration that they own the right to reuse, exploit and redistribute content.
provided to them by individuals, even if content is subsequently deleted by the individual from their account. It is outside the scope of this study to investigate the individual requirements of individual commercial social bookmarking services, but this risk is mentioned here as a potential hidden and unexpected barrier to the sharing of information and resources.

Archiving/preservation of bookmarks rests with the individual in most cases. This means that the responsibility of saving outside of the service may not occur in a regular and systematic way, or may not occur at all. Thus any problems with the service could result in a loss of valuable resources to the working researcher.

Permission systems on bookmarks often permit private bookmarking, which cannot be distributed to a third party in any case—so a bookmarking preservation service would need to have access to users’ accounts to do a complete job.

The social and community benefits are dependent on individuals sharing their bookmarks—but this is not a requirement of any of the services described in this report, commercial or non-commercial, so there is a risk that no-one will share their resources and the benefits outlined above of sharing and exchanging information will not take place.

The wisdom of crowds is dependent on the existence of a crowd, and people en masse are unpredictable; there are many years’ worth of evidence to demonstrate that social software goes in and out of vogue. As a personal information management system, this makes little difference to the functionality; but systems that depend on evidence from multiple sources are at risk of impairment from a limited user body.

As attention moves from one service to another, there is a tendency for older information to be lost and replicated in another form (i.e., microblogging mentions vs links placed on a social tagging service). Some of these services perform a similar job, but in an ephemeral manner (Twitter, for example, is not expected to make information readily accessible post-hoc, although archiving services do exist for the purpose). Academic/public interest in social tagging may have peaked (figs. 1 and 2).

The perception of individuals using the services may not take on board the sharing aspects, which means that a service can be used as a personal content management system. This wouldn’t negate the usefulness of a service, but would re-focus benefits away from the community/sharing nature of such tools.

The growth in use of these general systems indicates the independent exploration of online tools created outside the academic research community. Researchers, potentially, have the scope to identify such tools for themselves, independently, and apply them in a useful way to their own research work. A service developed within the community would, to compete, need to be sure that it could offer new developments and adapt and incorporate external functional developments quickly and effectively. If collaborative aspects are involved in the functionality of the service, it would also need to attract a sufficiently large contributing audience. Failure to do so could, potentially, result in researchers engaging with external systems through choice, based on perceived benefits and new functionality.

The graph above indicates a slight slowing down of academic interest in 2010 for these services. Potentially, this could indicate that these services are considered well-understood and hence attract less academic interest, or that users are now turning to alternative systems and academic interest is waning as a consequence. There is a low threshold to adoption of more general
Discussion of 'technorati social tagging' over 2000-2010

![Graph showing interest in Technorati's role as a social tagging service from 1999 to 2010.](image)

Figure 1: Interest in Technorati’s role as a social tagging service (see appendix B for methodology)

Discussion of delicious and flickr as social tagging services over 2000-2010

![Graph showing academic interest in Delicious and Flickr as social tagging services from 1999 to 2010.](image)

Figure 2: Academic interest in Delicious and Flickr as social tagging services (see appendix B for methodology)
systems, especially if data migration is made easy. If a new system is released that is perceived as offering more benefits, users may well engage with the new system. However, there is no certainty that this would happen with a system developed within the community—it could equally apply to a system or systems developed for a wider audience outside the academic research sphere, that offers benefits and is fit for purpose from the perspective of researchers.

4. Review of existing initiatives

There is much activity currently in the area of social bookmarking systems. Some of the initiatives have been developed within the UK academic community (e.g., see JISC activities, Section 4.2), while others have been developed by third-party organisations, mainly by the commercial sector. Of these third-party systems, a large number are generic bookmarking systems that are designed for use by any Internet user. However, some of these (e.g., Delicious) have been pressed into action in support of learning or research. Another—much smaller—set of bookmarking tools have been developed specifically with the research community in mind, prominent examples being: CiteULike, Connotea, Mendeley and Zotero. This section aims to give a brief overview of some of these systems to help inform the context in which the development of a social bookmarking system specifically for the UK research community can be assessed. See appendix A for a broader statistical overview of social bookmarking system usage.

The first generation of Web bookmarking sharing services (e.g., Backflip) emerged in the mid-1990s but most struggled at the time to find a viable ongoing business model. At the present time, a large number of generic social bookmarking services exist (see appendix C), including Digg and Reddit (used for sharing news content) and Stumbleupon. Perhaps the most well known generic bookmarking service is the one known as Delicious (formerly del.icio.us), founded in 2003, taken over by Yahoo! in 2005, then sold to Avos Systems (a company owned by the founders of YouTube, Chad Hurley and Steve Chen) in April 2011. Delicious enables registered users to save (and manage) their bookmarks online as well as view what other users are bookmarking. The Delicious interface enables users to track bookmarks that are popular or current, as well as navigate through user-generated tags.

Some of the tagging principles used by social bookmarking sites have also been adopted by Web-based content services like YouTube and Flickr. The latter, for example, was launched as an image sharing service in 2004 and—just like Delicious—was taken over by Yahoo! in 2005. Images are notorious for missing semantic metadata, so Flickr encourages the submitters of content to include tags that would help identify the subject matter and geographical location of images. However, Flickr differs from del.icio.us in that it promotes 'narrow tagging'—allowing owners to tag their uploaded pictures and to give other individuals permission to do so on a case-by-case basis (Simons, 2008).

4.1. Social Bookmarking services for the Research Community

Some third-party systems provide social bookmarking services aimed specifically at the academic community. CiteULike, for example, was created in 2004 by Richard Cameron. On returning to academia, he found that the then available systems did not offer the functionality that he, as a researcher, required

1http://www.citeulike.org/faq/faq.adp
There were a number of social bookmark managers which existed before CiteULike (Delicious, unalog, etc.), but they were all general systems designed to handle arbitrary web links. They didn’t really capture all the metadata (authors, journal name,...) which go with academic articles.

CiteULike focuses on tagging of articles from academic journals, and aims to offer the user a personal library environment from which they can not only build up their collection of journal articles, but also create biographies. Tagging is open and done by readers, authors and information professionals (Kipp, 2010).

Connotea is a similar service, run by Nature Publishing Group (NPG) as a free service that links into various commercial NPG tools and products. It is aimed more specifically at scientists. In addition to journal articles, saving and sharing of, for example, entries from PubMed and Amazon product page information on books, is also supported. Again, the key differences with more general tagging systems lies in the emphasis on bibliographic data, functionality designed to create bibliographies and the concept of a personal library.

Both systems take a flexible approach to the sharing aspect social bookmarking by pointing out the benefits of sharing with colleagues and emphasising the ease of sharing links. CiteULike puts emphasis on the benefits of sharing with a community. Its FAQ states that, ‘the more people who use CiteULike, and the more they use it, the better it becomes as a resource’. Connotea, on the other hand, emphasises choice, ‘you can choose to make them private, or shared with just a select group of other Connotea users if you prefer’.

2Collab was another social bookmarking system aimed at scientific researchers, in this case provided by the publisher Elsevier, who withdrew the service from April 2011 when it decided instead to focus on ‘providing optimal integration with and support for the leading bibliographic management and social-bookmarking tools currently on the market’.

By contrast, Mendeley is a social bookmarking system for the academic community that is currently attracting many users and a lot of interest from the UK HE community. It is not a purely online system, as it combines a desktop application with a web application. A basic version of Mendeley is available free of charge, with additional capacity for storage, groups etc. being available through a premium service available at a cost.

The Mendeley service offers similar functionality to the other academic-specific systems but also has additional features specifically tailored to the scientific research community. Additional features include a stronger emphasis on sharing than other similar systems, with functionality that allows users to synchronise their personal libraries with selected colleagues, potentially creating a more controlled work-style environment for research collaboration than the purely ‘folksonomy’ approach of the more general systems.

Greater emphasis is also placed on integrating the system within the existing research environment, whether this is within the personal desktop environment of users or the greater online research environment. For example, DURA (Direct User Repository Access), a year-long JISC-funded

---

3http://www.connotea.org/about
4http://www.citeulike.org/faq/faq.adp#why
5http://www.connotea.org/about#privacy
6http://www.2collab.com/
7http://www.mendeley.com
8http://www.jisc.ac.uk/whatwedo/programmes/inf11/jiscdepo/dura.aspx
project led by the University of Cambridge, seeks to integrate Mendeley into the institutional repository environment. As part of this, the Mendeley system will also integrate with the Symplectic Elements research management system to reduce re-keying of metadata and offer access to full text articles via the Mendeley personal library for researchers.

Mendeley offers an interesting alternative approach whereby a third-party system seeks collaborative opportunities to develop inline with requirements within the HE community. This differs from third-party systems, which are developed either in isolation or within the context of a commercial company offering a social bookmarking system for free as a way of integrating with their paid-for products and services.

4.2. Content analysis of third-party social bookmarking services

In evaluating social bookmarking services, it is useful to get a ‘flavour’ of the characteristics involved and of the implications. Social tagging sites vary in much more than in the size of the user population. Figure 3 is a rough-and-ready comparison of three sites—Delicious, Connotea and Citeulike—alongside an indication of the apparent newsworthiness of each keyword entered (as estimated by a Google News archive search).

Publication delay: Note first that Delicious seems to be first to catch up to newsworthy events, such as the Jasmine Revolution. This is predictable, in that peer-reviewed or formally published sources that are more likely to be covered by the other two sites incur a

---

significant delay due to the publication process, so sites focusing on this class of resources will update more slowly.

**Popular topics** such as the subjects of recent news reports, for example, Magnapinna, and cult topics such as the subject of internet memes (e.g., Caramelldansen), attract coverage on Delicious. It is possible that this is due to a lack of academic interest, but lolcats as a keyword attracts almost 200 hits in Google Scholar—so one is left with the conclusion that the academic community with an interest in these topics is not well represented on Connotea/CiteULike.

**Academic subjects** appear to be distributed between the three sites depending on their topic: whilst CiteULike seems from this brief analysis to contain technology, geographical and medical subjects, Connotea seems to contain little technical content.

### 4.3. JISC activities related to personalisation and tagging

JISC activities in this area fall broadly into two strands, the enhancement of existing systems through the introduction of Web 2.0 tools, including tagging, and the enhancement of tagging itself to create more specific and useful tools for the research community. The projects described below broadly illustrate these strands. Overall, the enhancements seem to have been viewed favourably by those researchers who engaged with the systems and gave feedback, but the sample groups were, in some cases, small and largely based within specific disciplines. This does not negate the findings, but the creation of a specific academic social bookmarking system is not necessarily the only way of building on this work. Rather, the projects show the potential usefulness of integrating tagging functionality within existing, well-used systems.

#### 4.3.1. Studies of personalisation in the JISC Information Environment

Between 2007–2008, the JISC services EDINA and MIMAS ran two projects (DPIE1 and DPIE2) under the general heading Development of Personalisation for the Information Environment. These investigated how the JISC Information Environment (IE) could be ‘enhanced with personalisation features and Web 2.0 systems and technologies’\(^9\). The projects reviewed the ways in which (adaptive) user personalisation could be used within the JISC IE to enhance the user experience. The DPIE2 project report (M. Hammond et al., 2008) concluded that would be ‘most appropriate for JISC to leave the development of personalisation to those who provide services currently, and to those who can develop useful functionality based on existing content’.

In addition, a series of features were recommended for a proposed infrastructure that would be able to support the development of personalised services by third parties. They concluded that an environment conducive to the development of personalised services would have the following features\(^10\):

- diverse and ubiquitous sources of data on resources and users, with standards to support interoperability

---


• flexible, extensible platforms for delivering services
• services with flexible/permissive objectives and funding to experiment with new functionality
• a clear approach to acknowledging and addressing privacy and disclosure concerns
• trustworthy services
• effective channels for sharing ideas, processes and good practice
• recognisable/identifiable users
• a relentless focus on defining, realising and measuring benefits, particularly end-user benefits

Between February and September 2009, JISC funded a follow-up project at EDINA on Supporting Personalisation across the JISC Information Environment. The project reported in 2010 that they were exploring two areas:\n
• Location-based searching for JISC IE services: will build a proof of concept demonstrator that takes location information on resources (and services) and uses these to sort and rank the results of searches on the basis of geographic proximity to a user-defined location.

• Shared middleware service of service-content references: will investigate and demonstrate a reference linking knowledge-base, with a focus on multimedia resources, that could act as middleware for personalisation actions by services into the JISC IE.

4.3.2. Enhanced Tagging (EnTag) project

In 2007–2008, the JISC-funded project EnTag (Enhanced Tagging for Discovery) created two demonstrators to investigate ways of using social tagging linked to controlled vocabularies. The project explored ‘[...] ways of enhancing social tagging via controlled vocabularies, with a view to improving the quality of tags for increased information discovery and retrieval’ (Golub et al., 2009). Looking at presenting two groups of users—readers and authors—with tagging functionality enhanced by controlled vocabulary options, the project focussed on investigating the indexing aspects and the retrieval aspects of social tagging versus social tagging in combination with a controlled vocabulary.

The hybrid approach used in the demonstrators illustrated the potential value to both groups of the addition of controlled vocabulary elements to the traditional free-tagging systems. Such specific and specialist terminology, and the enhancement of precision in both tagging and searching/retrieval tasks was seen as beneficial by the test groups of users. However, this was very dependent on the ‘quality and appropriateness of the controlled vocabulary’ and there was great importance placed on an un-cluttered, easy-to-use user interface. The simpler, more
streamlined interface, more similar to conventional tagging, was used more frequently during the study (Golub et al., 2009).

Although users did find benefits to using the enhanced service, this was based very much on the additional specialist vocabulary available. Further testing on wider subject areas and using a wider variety of controlled vocabularies was recommended by this project.

4.3.3. PERonalised Tagging Information Services (PERTAINS)

In 2009, this 5-month project\(^\text{13}\) was funded by JISC to build on the work of the EnTag project. The aim was to pilot tag recommenders in UK national services (Intute and COPAC), as part of JISC Personalisation Programme. The project was led by MIMAS.

The aims of PERTAINS were:

- To build on the En-Tag project to produce a transferable tag recommender tool for the Intute and Copac services
- To provide the Intute and Copac user communities with a tag recommender tool, allowing users to tag with a tool developed in conjunction with their needs as defined by them\(^\text{14}\).

However, the target users were students, rather than the researchers and authors targeted in the EnTag project. Key findings indicated that the addition of suggestions within the tagging tool made students more aware of the benefits of using tagging for search and retrieval purposes, which might suggest that enhancing a system to support research and teaching/learning tasks is valuable to the HE community. Students also stated that they wanted to be able to choose whether or not they shared their tags with the wider community, and that some tags were considered personal/private.

Although the technology and controlled vocabulary aspects of PERTAINS built on the work of the EnTag project, the test user groups were very different and engaged in very different kinds of work and direct comparisons between the two projects as far as user reactions and feedback are concerned should be treated with caution. A useful tool for students is not necessarily a useful tool for researchers and authors within the academic community. For example, participants’ previous engagement with social bookmarking systems, general use of tagging in other contexts, level of knowledge and specific requirements may differ greatly. As such, this could give a very different reaction to and interaction with the concept of enhancing tagging functionality with a controlled vocabulary.

5. Discussion points

5.1. Sustainability

Social bookmarking is an ad hoc and uncontrolled process. In fact, it may not be possible to entirely control this process. There is significant evidence that interface changes can make a

\(^{13}\)PERTAINS project: http://hypermedia.research.glam.ac.uk/kos/pertains

major difference to the sort of information that is provided, as user perception of the functionality of the service changes [citation needed]. Tags may be useful metadata (or annotation) that needs to be maintained (a growing body of research is based on analysis of tagging behaviour).

Tags as ‘annotation layer’ (c.f., research data)

- Commercial products can be—and are—bought, sold, and discontinued. This can raise issues of data ownership, technical problems of transferring data, loss of a system in daily use and the potential disruption of changing systems in the middle of a research project, which could delay schedules and hamper collaboration.

- Non-commercial tools can cease to be funded, and the same issues raised by the transfer of ownership or discontinuation of commercial products would equally apply in this case.

- Tools can break, or change, breaking dependent software products—or suffer breakdowns or data loss. Linked with preservation and archiving issues and the problems of transferring data between systems (detailed below), any research practices relying on a social bookmarking system for sharing of resources, saving resources and general collaboration would be jeopardised by a system being unavailable or unreliable for a period of time.

- Preservation/archiving issues remain a problematic area. It is unclear where responsibility lies, and individual users may fail to take any steps to preserve or archive the content they store in these services. They may be unaware of their vulnerability in this area, not see it as a serious risk or not have the skills and tools to carry out such activities. Failure to preserve and archive can, potentially, result in a serious loss of content, some of which may be critical for a research project.

- Issues around ownership of data have been highlighted by the recent acquisition of Delicious first by Yahoo! and then by AVOS. This has made this a real-life issue rather than a speculative one. Yahoo! issued details of how bookmarks could be transferred along with the announcement of the latest transfer of ownership. However, transfer of bookmarks does require signing a different agreement with the new owner, AVOS. Failure to sign the agreement and transfer bookmarks means that, ‘You will no longer be able to use Delicious or access your bookmarks after the transition is complete’. Such a scenario could happen with any third-party system, but equally raises issues for any system based in the HE community if funding were cut and the system was no longer supported.

- There is no generalised privacy model for tags, which raises issues regarding migration and user expectation of data reuse.

\[\text{http://www.Delicious.com/help/transition} \text{ 'What if I don't opt-in to migrate my bookmarks to AVOS?'} \text{ (information removed from page)}\]
Transferring of bookmarks from one system to another is another significant risk area. For example, in the recent acquisition of Delicious by AVOS, users were given an option to sign up to a new agreement with AVOS and transfer their bookmarks to the new system. However, users were given no information or guidance on transferring their bookmarks to any other system. Failure to transfer the bookmarks to the new system by a certain date, as noted above, would result in a loss of all current bookmarks. This real-life scenario highlights the potential risks in the technical difficulties of transferring data in a complete and usable form between two potentially un-related systems.

5.2. Role of third-party services

Due to the distributed nature of tagging services, tagging functionality can be added to academic resource services such as repositories in a variety of ways; systems such as Delicious can easily be referenced within web pages as an actionable link. Alternatively, academic-specific services may be adapted to provide a new service offering tagging functionality on the basis of existing infrastructure.

- Users may well go with whatever the popular systems are at the time (e.g., Delicious) despite what else may be available from academic specific services. Further investigation into user behaviour and takeup of academic-specific systems versus general, popular systems would be valuable, as it is difficult to control the systems individual users select for their own work. Options of supporting choices through developing additional tools to aid preservation, archiving and transfer of data between systems may be a viable alternative approach to trying to enforce use of one system over another.

- Established use of and familiarity with generic services could be a significant barrier to the used of tagging functionality within academic specific services. Again, a non-system-specific support package could be a possible solution to managing the risks of third-party systems and also supporting user choice in systems adopted for research purposes.

- Dictating use of academic specific services won’t work in the research community, so there is a need to make a strong case to consider developing them, i.e., there must be evidence that they will be used. More appropriate and better functionality often doesn’t win users over and will not necessarily stop them continuing to use the services they are already familiar with. It may well be more effective to guide people through best practice, independent of the peculiarities of a given tool.

- Sustainability could be an issue—once people rely on the functionality, it would need to be maintained. This raises particular difficulty in the current UK funding environment, in which sustainability has become a key problem. It is worth noting again that sustainability has been proven to be problematic for many commercial services. A consciously distributed approach, encouraging the widespread deposit of data over a number of bookmarking services, is likely to be the most stable over time. However, this raises issues of intra-site incompatibility and mapping issues as well as the potential duplication of effort.

- Consideration could be given to putting effort into offering a supporting preservation/data transfer service that could be used across many services. This would leave academic researchers free to use their preferred service, but ensures backup and continuity of bookmarks, thus mitigating risks.
5.3. Tagger motivation

Bookmarks are usually created for personal use, but are sometimes created for the wider community (Arakji et al., 2009) or some intermediate grouping, i.e., a community of interest or practice.

- The community model only works if a critical mass of individuals share their bookmarks—but the critical mass is not a constant (e.g., Benbunan-Fich et al., 2010, p. 145). It depends on the effort individual users are willing to put in, such as focused groups asserting their own schemas. For example, there are many examples of online communities, such as LiveJournal groups, making use of tags to index older content. Where the cost of indexing correctly is a prerequisite to posting, the motivation for posting seems to be enough to overcome the obstacle of accurate (if limited) indexing.

- This does not negate the benefits of using these services to create a personal content management system that is flexible and easy to use, but this is separate from the social bookmarking/sharing functionality.

- The motivation for sharing and/or not sharing need to be understood on a number of levels. Perception and subsequent actions here could be influenced by the way in which researchers in different disciplines work. One motivation may not fit all. Some sciences are based on collaboration, other areas, like history, tend to produce individual works in isolation. This is the culture of the discipline, and the way researchers are familiar with working, so will affect their use of a social bookmarking tool.

- Finding on the basis of survey work that motivational bookmarking, intentionally bookmarking resources for public consumption or sharing, locally or generally, is common, Arakji et al. suggest that circumstantial contributions are statistically insignificant—but that there remains a possibility that personal tags are simply marked as private. That said, Arakji et al. note that the study is self-limited by its online survey methodology.

5.4. Tag quality

Tag quality is a recurring issue with any social bookmarking system. When users attempt to create their own metadata, even in a professional capacity such as a researcher tagging research material, the system is open to confusion, lack of context and lack of precision which could, potentially, make discovery difficult.

- Missing hierarchy and context is a problem area, although various solutions have been proposed and implemented on given tagging systems to solve this issue (for example, Delicious used tag bundles for this purpose).

- Using terms for different purposes—subject, genre, value judgements (e.g., ‘funny’), provenance (e.g., ‘import-04-03-10’)

- No differentiation between description of the content and description of the site

- Overly personalised (e.g., ‘read’, ‘not read’; in LibraryThing, ‘my dad’ in Flickr)—again, this is not entirely without value, but the usefulness of these terms can be limited for purposes such as object retrieval based on general search terms. That said, if the use case chosen involved profiling a user’s interests and/or social network, this information would...
be of value; our judgements of the worth of user tags are contextual on our expectations of dataset usage.

- Little quality control (misspellings, compounding of words, grammatical standardisation). Standardisation of vocabulary can be difficult when technical and discipline-specific tags are used and agreement on terms that are accepted by the community can be an issue. On an international level, spelling and grammar vary greatly. The problem is therefore difficult to solve centrally.

These points overall raise the question of filtering, performing triage on the tag corpus to retrieve the information that is of use for a given purpose. This reflects the nature of social tagging as a raw resource requiring significant pre-processing for many practical uses. Many of these problems may also be resolved during the tag creating process, for example, by providing a language-specific spellchecker for input sanitisation. Work carried out by the EnTag project and the PERTAINS project on the integration of a controlled vocabulary with conventional tagging demonstrates another approach to resolving this difficulty.

5.5. Building communities

The marginal cost of creating tags is low, but many use cases for tagging systems depends on maintaining a critical mass of users (network effects). The benefits of building a successful, dynamic community where researchers engage with each other, share resources, forge collaborations and add to their knowledge and expertise are significant. However, there are many barriers, both technical and cultural, to building a successful and sustainable community (Benbunan-Fich et al., 2010; Corrocher, 2011). These barriers need to be assessed in the light of what it is feasible to achieve and the cost benefit of effort involved in such an undertaking.

- Free riders are a type of user associated with social networking sites. Basically, a free rider will access and make use of the content added and shared openly by other users of a site without sharing any content of their own. They do not necessarily need to be a problem; watching their progression through the site can provide useful evidence of the perceived relevance of applied tags. However, they do not contribute to the general development of the online community in which they operate.

- Some users are not necessarily free riders, but have a different perception of what the social bookmarking systems are for. Rather than focusing on the community/sharing aspects, they use a social bookmarking system as a personal content management system. This means, in practice, that they save bookmarks and create a personal library but do not share their resources with others.

- Potentially, to build a research-focussed community, common agreed terms need to be adopted that are useful for highly specialised users.

- Different disciplines may do things differently. A potentially useful focus may be to explore many small communities rather than a ‘one service fits all’ approach across the research community. Procter et al. (2010, p. 4053) suggest that the need to focus on innovation arising from community-based activities may be more generally characteristic of Web 2.0 use in scholarly contexts.
• Different kinds of content may be favoured by different disciplines, e.g., images/photos may be more important to some than others; books, general web pages, usage and importance may not be equal across disciplines. This means that different tools may be needed if JISC create their own service to serve a given user community; for example, a data-focused tagging service may have very different requirements to an image-focused tagging system. This may lead to the perception of ‘reinventing the wheel’. The question remains, will people find/create the tools they need quicker in the commercial environment, and adapt them more easily to their needs?

• It is not clear from current available information if there are genuine communities in the research environment, or, instead, a collection of individuals using bookmarking services as personal content management systems with some also using the option of sharing resources. The term ‘community’ may cover many types of grouping: research groups; lectures sharing a reading list with students in CS101; loose-knit groupings of art students with a shared interest in abstract sculpture. The organisation of these groups depends very much on their purpose and the underlying human relationships, and may very well reflect asymmetry through the rate, nature and character of individual contributions, the existence and nature of moderation and repair mechanisms, and so forth.

• Some disciplines may not have a culture of sharing resources in the pre-web environment. For example, if researchers in certain fields don’t traditionally, share books, articles etc. in paper format, they will probably not share in electronic and online resources either. The provision of alternative mechanisms may not suffice to ensure cultural change.

6. The need for an academic social bookmarking service

The need for an academic social bookmarking service can most usefully be viewed within the context of the existing services currently in use and the related developments already funded by JISC in this area. This section will look at the benefits and barriers of developing an academic bookmarking service within this context and indicate in both sections where existing services could provide an alternative.

6.1. Benefits

• Community control of content ownership on agreed terms.

• Single service would potentially offer easier integration with existing systems used within the research environment.

• Sustainability and reliability could, potentially, be built into such a system.

• Bespoke functionality for researchers could be included, to potentially make the system a more useful tool than others currently available.

• A single community could, perhaps, develop around such a system, bringing the benefits of collaboration and knowledge to the UK research community.

• Transfer of data between systems could, potentially, be negated by the use of a single system.
• Future developments would be driven from within the community and so could, perhaps, reflect the ongoing requirements of researchers in a more specific way than third-party systems.

• Third-party systems aimed at academic researchers tend to focus on science research. A system developed within the community could have the potential to investigate and incorporate the requirements of other subject areas and therefore engage much more widely with the existing research community.

6.2. Barriers

• Many of the benefits would rely on a critical mass of user engagement with the system. This, however, cannot be enforced within the research community, so any new system would need to be able to compete with existing systems and win over users. This is a very difficult task, and the risk of failure would be high.

• Existing systems may already be providing the functionality researchers require, so reinventing this wheel could, potentially be costly and provide no real benefits to users.

• Ownership of data, preservation and archiving issues could, perhaps, be addressed by creating tools and services which could integrate with other systems yet be owned by the community. This may be a more cost-effective and workable approach, which does not carry the high risk of asking users to swap from a currently used and favoured system and also minimises the risks inherent in using third-party systems.

• The difficulties of transferring bookmarks between systems, and the high risk of loss or partial loss of data in doing this, could be addressed by developing tools to perform this task across multiple services. Again, this would both minimise the risk of losing data and also allow researchers to continue to use the existing systems, without a need to transfer to a new system.

• The difficulty in building communities is a high risk when developing a single research system. Users may, potentially, not engage in community activities of sharing for many reasons, and it could, perhaps, be more productive to nurture existing collaborative and community activities within any currently active environments.

• The perception among researchers of any social bookmarking tool as a personal content management system rather than a social network may mean lack of engagement with a system that offers UK-wide sharing as a benefit. Participation in community activities may not be a motivating factor for some of the research community and could negate the benefits, as they would perceive them, of transferring systems from their existing choice.

A single system may incorporate work already done within the UK HE community (e.g., CO-PAC personalisation, EnTag and PERTAINS work on the integration of controlled vocabulary, DURA work on integrating tagging with institutional repositories in a personal research library environment), but may also lose some of the value these projects brought into the community by engaging with and enhancing existing services and systems. An alternative approach would be to assess the impact of such work and perhaps build on the areas where users had expressed interest and shown engagement already.
7. Conclusions

It may not be possible to control the selection of systems. Equally, it may not be possible to control the use of systems by individual researchers. If popular, more general third-party systems are being used on the ground within the research community, this raises issues of preservation, ownership of content, the moving of data from one system to another in the event of a system ceasing to exist and the possibility of researchers losing valuable information. However, enforcing a system may meet with resistance and is unlikely to be adopted by the community.

The abundance and obvious popularity of third-party systems needs to be considered when investigating building new systems. Re-inventing the wheel would not be a good option. It may be that working with systems that are already embedded within the research community to further integrate them into the wider research environment (e.g., the work being done to integrate Mendeley with institutional repositories) and working to add enhanced tagging functionality to existing, popular and well-used services and systems (e.g., the personalisation developments for COPAC) is a useful starting point for further development plans.

There is some considerable difference in the functionality offered by the academic-focused social bookmarking systems (e.g., Connotea, CiteULike, Mendeley) and the more general systems (Delicious, Flickr etc). Indeed, there is considerable variance in the extent to which many systems may be described as enabling social tagging at all, as opposed to free-text indexing for personal or groupware information management. However, there are no firm indications as yet in how this different functionality actually affects uptake or use by the academic community—Flickr, for example, although often categorised as a more personal tagging environment, has been adopted widely by many communities with some tags being used that are obviously designed to enable discovery by a wider audience than family and friends. It could be that if a system is seen as offering tools that are fit for purpose, then that defines the selection made by a community of users.

The development of tagging functionality with controlled vocabulary options potentially offers a more precise tagging option and has had positive feedback from users, but testing has been carried out within small, subject-specific groups (EnTag) and with groups that have very different potential user stories, e.g., researchers and students (EnTag, PERTAINS respectively). Thus it is difficult to conclude the relevance to and the potential impact of such developments within the wider HE community.

The very different workflows within all social bookmarking tools facilitated by the option to keep tags private or to share tags is a significant factor in creating either a personal content management system or in building a community. There is value to the research community in having a flexible, online personal content management system, but this is an entirely different way of using these systems than sharing with the wider community or even sharing with a defined group of collaborators. All the systems encourage the sharing of tags and community building, but all also offer a personal library option. It is at present unclear what usage patterns are being created in practice, and this could be a determining factor in selecting the most useful functionality to develop in the future.
8. Recommendations

- Further research into the use of social bookmarking tools within different disciplines and subject areas is needed. Currently, most third-party academic-specific systems are aimed at the scientific research community, for example. Any new systems or enhancements to existing systems would need, potentially, to be of use to the whole research community. Subject-specific needs could well be a determining factor in uptake, as could the existing culture of different research areas.

- Consideration of how systems are being used in practice, i.e., whether the distinct workflows of sharing/community building activities and the privacy options that can be used to create a personal content management system are used equally or if one workflow is favoured over another would be beneficial. If one workflow is favoured over another, future developments need to take this into account.

- The issue of supporting existing systems favoured by the academic community is an area which, potentially, be of great practical value to working researchers. The main barriers to using third-party systems lie in the areas of backup, preservation, potential loss of data, data migration from one system to another and ownership/rights management of content. If third-party systems are being used successfully, it may be beneficial to the research community to acknowledge this and focus on offering support in these areas by creating functionality that can be applied to any system of choice and offers backup facilities and ‘safe’ storage of user-created content outside of the system. This would also facilitate early adoption of new systems and allow researchers to exploit developments within existing systems, giving a faster rate of access to new functionality than a complete system based purely within the HE community could, potentially, offer.

- The area of cost savings by allowing user tagging of resources as mentioned in the JORUMOpen report could offer potential benefits in sustaining services and resources that currently rely on professional indexing for discovery of content. Whether this could be applied in all areas without a loss of quality great enough to hinder discovery is unclear, but in the current economic climate further investigation could explore this option as potentially beneficial to the wider community.

- Collaboration with third-party suppliers of systems to facilitate embedding of tagging within the research environment (such as the recent work with Mendeley through the DURA project) is an area worthy of further investigation. Such collaborations offer the potential to bring new developments by experts into the research environment as an integrated service to enhance the user experience of both the third-party social bookmarking system and related systems (such as institutional repositories) developed within the HE community. Tapping into a popular system and the expertise behind it could, potentially, offer a more efficient way of integrating the latest specialist functionality of social bookmarking systems into the research environment than in-house development starting from scratch.

- Integrating social bookmarking tools into existing popular systems and services already developed within the HE community (such as JORUM and COPAC) is, potentially, a useful way of gauging interest and enhancing the user experience. However, care should be taken to follow an integrated approach and not create silos where systems have isolated social bookmarking functionality that does not place the user at the centre of the experience. Ideally, a user would want the opportunity to create, save and share bookmarks across
such systems, not have to be in a system-specific environment to access each set of data and resources.

- Finally, we note that there is considerable variation over time in the usage patterns of these systems both individually and in aggregate. Older systems are either abandoned or move over towards usage as a commonplace production service. As the tide turns, services increase and decrease in perceived value. There is a significant risk of developing services that, if they capture user interest at all, do not retain it in the longer term, and for this reason we suggest that attention should in the longer term focus more closely on dataset-centric activities than on an individual data collection system/UI.

9. References


A. Social bookmarking service overview

A.1. Selected service summaries

A.1.1. Scientific papers — Connotea

Name: Connotea
URL: http://www.connotea.org

Description: Third-party social bookmarking system designed specifically for scientific researchers. Owned by Nature Publishing Group (NPG).

Connotea popularity over time. (See appendix B for source/analysis)
A.1.2. Scientific papers — CiteULike

Name: CiteULike

URL: http://www.citeulike.org

Description Third-party social bookmarking system founded in 2004 by Richard Cameron, an academic researcher who found that existing social bookmarking systems were too general for his research needs.

Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond

By: Jim Black


CiteULike popularity over time.
(See appendix B for source/analysis)
A.1.3. Scientific research — Mendeley

Name: Mendeley
URL: http://www.mendeley.com

Description: Third-party social bookmarking system that offers desktop and online interfaces, aimed at academic researchers as a reference management tool and a social network. Originally the service was more science focused, but is now expanding into other subject areas. Mendeley is currently participating in the JISC-funded DURA project which aims to integrate institutional repositories with the Mendeley user environment.

Mendeley popularity over time. (See appendix B for source/analysis)

---

17 JISC DURA (Direct User Repository Project) project blog: http://jisc-dura.blogspot.com/
A.1.4. Webpages — Delicious

Name: Delicious

URL: http://www.delicious.com

Description: Third-party social bookmarking site that is for general use. Very popular and widely used. Bought by Yahoo! in 2005, recently taken over by AVOS, thus raising very practical concerns about sustainability and reliability of commercial systems, migration of data and ownership of content stored by users.

Delicious popularity over time.
(See appendix B for source/analysis)
A.1.5. Photographs — Flickr

**Name:** Flickr

**URL:** http://www.flickr.com

**Description** Third-party photo sharing system that uses tagging. Examples of tags can be highly personal. Flickr is used by many kinds of specialist groups, including the research community, to share images.

![Flickr popularity over time. (See appendix B for source/analysis)](image)
A.1.6. Books — LibraryThing

**Name:** LibraryThing

**URL:** http://www.librarything.com

**Description**  Third-party social bookmarking system focusing on books. Book titles and bibliographic data can be tagged and shared. Designed for the general user, but has been adapted by specific groups to support sharing of books of all kinds.

![LibraryThing popularity over time.](See appendix B for source/analysis)
A.1.7. Blogs — Technorati

Name: Technorati

URL: http://technorati.com

Description  Third-party social tagging system that supports the tagging of blogs and blog posts.  

Aimed at the general user, it has been adopted and used for more specific groups, including the research community.

Technorati popularity over time. (See appendix B for source/analysis)
A.2. Language-specific and specialist sites

In this section, we take a brief and fragmentary look at locally popular sites—for example, social sites on the francophone Web. This allows us to gain an idea of the relative popularity of different kinds of social site. Analysis of relatively low-volume sites is of interest to us in that it helps us to understand what the parameters of such sites are: just how many people are required to build a ‘networking effect’? What happens when this is not achieved? Is it possible for small, niche sites to remain healthy and useful, rather than targets for spam? Academic-specific services are in themselves a small niche market, and a UK-specific service much more so.

A.2.1. The Francophone Web

Just as academic communities make use of generalist sites to index or store their links, so too is it nonsensical to try to draw boundaries around French-language use of the Web. As a consequence, language-specific sites tend to be relatively small and little-used. ComScore (2009) suggest that the most popular French-language social networking site is Facebook.

Here we see many small sites that perform functionality similar to Delicious or Digg, and a number of others that focus exclusively on certain types of content (e.g., collecting tutorials). The global ranking of each site is very low, although on a content level some appear relatively healthy (maintained, with little spam).

A.2.2. The German-speaking Web

Things look rather different on the German-speaking web, with a fairly popular Facebook-alternative site (studivz.net) attracting many views. (Note that this and many subsequent graphs use a logarithmic Y axis.)

---

Comparing the social bookmarking sites mentioned here: folkd.com, mister-wong.de, linkarena.com— with bibsonomy.org, we find that these sites appear to be similar in numeric, although not necessarily demographic, popularity (the below graph demonstrates this comparison).

Note, however, that the more popular sites mentioned here appear to be reducing in popularity. A content-level analysis of the larger alternatives would allow us to check whether the users have simply ceased using social bookmarking applications, or whether they have moved to using a different service.

A.2.3. The Dutch-speaking Web

The story in the Netherlands is similar. A Dutch-language tagging site with low popularity survives, despite problems with spam (tagmos.nl); another, watvindenwijover (‘what we think about’) has closed down. EKudos, a Dutch-language Digg equivalent, survives with relevant content. The most popular of the sites shown in this graph, hyves.nl—a Facebook equivalent—remains extremely popular in the Netherlands.
A.2.4. The Spanish-speaking Web

Again, the Spanish-speaking world has embraced marcadoras sociales (otherwise known as the etiquetado social) alongside other social web sites—although again, the low popularity of many sites suggests that the majority of Spanish-language link sharing is happening elsewhere.

The most popular sites are, as one might predict, not those dedicated to tagging. Meneame (named following a Mexican children’s song, in which an incrementing number of elephants stand on a spiderweb!) is, again, a competitor to Digg. Migente.com is a social/dating network. Tuenti.com is an invitation-only social network. Bitacoras.com is a digg-style system that focuses on blog posts. By contrast, keepyourlinks.com, dir.eccion.es, favoriting.com and so forth focus on social tagging.
A.2.5. The Chinese-speaking Web

This illustrates the dangers of depending on third party analytical websites. For whatever reason, the most popular Chinese social websites tend to differentiate themselves from the large parent company through the use of a subdomain (i.e., bookmarks сервис.something.com versus bookmarks сервис.com). Alexa statistics are blind to subdomain usage, and therefore cannot differentiate this information at all. Compete.com recognises the distinction, but reports suspiciously low access numbers, suggesting that there is still a problem with the count.

The divergence between these results makes it difficult to draw a meaningful conclusion, other than to repeat anecdotally that there are several social tagging-centric websites in common use in China, most of which are services offered by large companies. QQ.com is Tencent holdings, publisher of the most popular IM software in China (based originally around ICQ). Baidu is an extremely popular search engine (Alexa rank: 7). Sina is a Chinese-language news and entertainment portal, who also run a Twitter equivalent, weibo.com, visible in the Compete.com graph, above.

The least popular of these sites, mister-wong.cn, is a localised version of a German site. Many other Chinese social tagging sites, such as gootou.com, igooi.com and myurlworld.com, simply no longer appear to exist, perhaps in itself an indication of the level of competition in this sphere.
Figure 4: Number of applications on the Android and iPhone Markets designed to integrate with various social web sites.

A.2.6. Conclusions

Across nations and languages, a pattern emerges: social tagging sites, even at best, are seldom as popular as social news sites and other types of social site such as Facebook and equivalents. ‘Niche’ tagging sites can exist—eventually even flourish—but to do so requires at the very least good luck, unique selling points, or significant investment in maintenance and content management. Alternatively, a significant amount of ‘seed’ data may be made available through the site—that is, preexisting content that can draw attention even before any social content has been created. Perhaps the most important lesson from this analysis is the sheer number of sites that are created, only to fail.

A second conclusion: it is not enough simply to explore the number of accesses of a site. Rather, an effort should be made to focus on the ‘health’ and genre of the site—the appropriateness of its content, the amount of spam, and the type of content that is represented there.

A.3. Mobile access to social tagging sites

The number of mobile applications designed to integrate to various social sites may provide some suggestion of the popularity of each site with mobile users.

Whilst Facebook seems to be the most popular, it is noticeable that Delicious and Meneami have also attracted some attention from developers. Mendeley, Connotea, Citeulike and Bibsonomy, by comparison, appear to have attracted no interest from Android developers; for the latter three, there exists no iPhone application either.

Surprisingly, given the fact that the iPhone app store contains many more applications than the Android store, the Android Market seems to have a wider diversity of applications in some areas. For example, the Android Market offers three Hyves applications, while the iPhone store
contains only one. On the other hand, a Mendeley application exists for the iPhone, whilst none exists for Android.

Broadly speaking, then, application popularity seems to echo site access statistics; social tagging sites remain less popular than social news and social networking in any given category, and Delicious is by far the most popular of the general-purpose URL tagging sites.

A.4. Overall distribution of social sites

As with many types of rank data there is a good chance that, if enough sites are considered, a more or less approximately Zipfian distribution will be identified. The sites discussed here are no exception. As can be seen in fig. 5, the distribution of sites discussed here drops off quickly from a peak of highest popularity, leaving the proverbial ‘long tail’ of less popular resources. There is nothing unexpected about this, but it is worth bearing in mind that a generalist site of a popular genre will be in direct competition with a large number of other sites within that genre—and that, as we have seen here, the primary site within that genre will receive exponentially more page views than the second-ranked.

This is of limited importance for many classes of development, such as infrastructural services, but for a social site in particular, there is an argument to be made that lack of contributions materially reduces the value of the site for the remaining audience. There is also an argument to be made that an absolute tipping point exists, beyond which the ‘network effect’ takes effect and the site becomes self-sustaining; predicting exactly where that tipping point may occur for any given system appears to be a complex and contextually bound problem.

A.5. Benchmarking the social web

Generic analysis, such as the ‘popularity’ graphs given above (see appendix B for methodology) provides only a very approximate idea of reach. The following graphs compare site popularity between similar sites.

This figure shows us that Delicious.com is by far the most popular external-resource-tagging service of the set. For context, two of these services (ma.gnolia and rawsugar) no longer exist.

---

Recent views of these services (about 0.5 per million) therefore do not relate to continuing use of the service.

Note that the y axis on this graph is presented using a logarithmic scale. That is, according to the data presented here, del.icio.us receives on the order of 100× more views than the nearest neighbour.

There is some evidence (notably in the first quarter of 2009) of migration between sites. During January of 2009, ma.gnolia suffered a major data loss incident, losing users’ data entirely. There appears to be a corresponding ‘bump’ in Delicious usage. That said, a rule of thumb commonly applied in the usage of Alexa rankings is: they are reasonably trustworthy in the top 10,000 web sites, and of relatively little value thereafter, because there is not enough data to generate reliable statistics. The fact is that most of these sites are not within the top 10,000 web sites. As of the first quarter of 2011, Bibsonomy was ranked at 4,923 by Alexa, Connotea at 10,971, CiteULike at 12,273, Mendeley at 19,003, and ma.gnolia.com, despite the fact that the service no longer exists, at 56,601. Delicious, by contrast, is ranked at 232.

An interesting question, in view of potential similarities between social tagging sites and other social web sites in some use cases, is the relative popularity of these sites. The following graph compares Delicious with a number of popular sites.

Note that this graph again presents the y axis using a logarithmic scale. Digg and Reddit are social news websites; StumbleUpon is a site that uses collaborative filtering of various kinds, including tagging of web pages from the StumbleUpon custom toolbar, to characterise and recommend sites. Technorati demonstrates a significant downward trend over three years, whilst Twitter rose to prominence, alongside linkedin.

Taken as a pair, these figures suggest a number of things:

1. social bookmarking sites (as accessed directly from the Web) are not as popular as social news sites, which receive about four times as many views per million globally.
Figure 7: Comparing popular social sites

2. In general, there is little obvious inverse relationship between the use of social news sites, micro-blogging sites, etc. and decrease in use of social bookmarking sites. That is, these numbers give us no reason to believe that users choose to engage with either one or the other—there is no obvious evidence of migration between sites.

3. There is, however, some evidence that there are ‘winners’ and ‘losers’ in individual categories of site.

4. Social tagging-specific sites are unlikely to be the most popular mechanism for link sharing, although they may be the second most popular dedicated mechanism (after social news sites). It is possible that they contain a larger number of unique links (more and more effectively indexed content) than other types of social site, but content-level analysis would be required to demonstrate this. A balanced approach to use cases such as exploring or preserving link sharing activity is likely to benefit from exploring information sharing via non-dedicated channels, such as Twitter, as well as social bookmarking per se.
B. Quantitative analysis of site popularity and content

B.1. Site popularity and reach

This information has been estimated using Alexa statistics as a primary source throughout this document. Alexa site histories are available for a small charge using Amazon's Web Service interface\(^\text{20}\).

There are many reasons to be cautious about these statistics. Firstly, the data collection method used by Alexa makes use of information retrieved from users who have installed the 'Alexa toolbar', a browser add-on. It is likely that a large demographic of users will avoid installing any such toolbars, so the statistics retrieved from this are unlikely to be entirely representative. Secondly, social tagging sites are services; sites that merely use social tagging as part of their indexing process are usually services. Flickr can be accessed from all sorts of applications, from Android apps to Adobe Lightroom. Social tagging applications exist for many platforms, for most of which the Alexa toolbar is unlikely to be available. This means that the statistics retrieved from these sites are unlikely to accurately reflect mobile usage of content- and information-management sites.

As Alexa statistics are limited in reliability, the following graphs have been included for completeness, and to allow the reader a chance to compare Alexa findings with an alternative source. These are taken from compete.com, a competitor of Alexa.

![Figure 8: The relative popularity of Facebook, Twitter, Linkedin, Digg and Stumbleupon (2010-04 to 2011-04) (according to compete.com statistics)](image)

As can be seen, figs. 8 to 10, broadly speaking, demonstrate similar trends to those predicted by Alexa, suggesting that these statistics are sufficiently reliable to support a general conversation about the size of the audience attracted by a given site.

B.2. Site content

The brief survey demonstrating the effect of 'newsworthiness' on coverage in social tagging sites (see main body of report) uses a simple methodology—terms in each category were chosen from memory. The count was made, where possible, through an API search. Google News and Scholar were handled similarly; unfortunately, neither offers a dedicated API.

\[^\text{20}\text{http://aws.amazon.com/awis/}\]
Figure 9: The relative popularity of Digg, Reddit, Delicious, and Bibsonomy [2010-04 to 2011-04] (according to compete.com statistics)

Figure 10: The relative popularity of tagging sites according to compete.com statistics [2010-04 to 2011-04]. Compare with fig. 7

It is suggested that this approach to classification of tag system is potentially useful, for several reasons:

- Firstly, we may improve our own understanding of which systems should be recommended for a given purpose.

- Secondly, being better able to characterise the usage of a given database means that we are better able to use that data in support of our own aims. If we find that a database contains factual descriptions, popular culture references, or indications as to user opinion or rating, then we are better able to describe both the resources indexed and the opinions of the indexer.

- Thirdly, some of the information held in tag databases may usefully be preserved beyond the life of the tag database (which, as we have seen here, may be relatively short-lived). Given the cost of preserving large quantities of data, and the high prevalence of spam in many sites, a triage process may be beneficial in informing this decision.
### C. Ranking of Web 2.0 sites

<table>
<thead>
<tr>
<th>Site</th>
<th>Views per millions</th>
<th>Rank</th>
<th>Country</th>
<th>Site classification</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>facebook</td>
<td>46554</td>
<td>2</td>
<td>US</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>baidu</td>
<td>16447</td>
<td>8</td>
<td>China</td>
<td>Search engine/User-contributed content</td>
<td></td>
</tr>
<tr>
<td>shuqian.qq.com</td>
<td>7296</td>
<td>10</td>
<td>China</td>
<td>Social site</td>
<td></td>
</tr>
<tr>
<td>myspace.com</td>
<td>5514</td>
<td>16</td>
<td>US</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>twitter.com</td>
<td>3813</td>
<td>12</td>
<td>US</td>
<td>Social network/microblogging</td>
<td></td>
</tr>
<tr>
<td>sina.com.cn</td>
<td>3804</td>
<td>15</td>
<td>China</td>
<td>Search engine/User-contributed content</td>
<td></td>
</tr>
<tr>
<td>linkedin.com</td>
<td>2308</td>
<td>38</td>
<td>US</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>flickr.com</td>
<td>2116</td>
<td>33</td>
<td>US</td>
<td>Photo sharing</td>
<td></td>
</tr>
<tr>
<td>wordpress.com</td>
<td>1401</td>
<td>18</td>
<td>US</td>
<td>Blogging</td>
<td></td>
</tr>
<tr>
<td>livejournal.com</td>
<td>992</td>
<td>81</td>
<td>US</td>
<td>Blogging</td>
<td></td>
</tr>
<tr>
<td>studievz</td>
<td>644</td>
<td>249</td>
<td>Germany</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>naver.com</td>
<td>632</td>
<td>237</td>
<td>South Korea</td>
<td>Search engine/User-contributed content</td>
<td></td>
</tr>
<tr>
<td>digg.com</td>
<td>616</td>
<td>95</td>
<td>US</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>hyves.nl</td>
<td>593</td>
<td>264</td>
<td>Netherlands</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>livedoor.com</td>
<td>576</td>
<td>86</td>
<td>Japan</td>
<td>Search engine/User-contributed content</td>
<td></td>
</tr>
<tr>
<td>ning.com</td>
<td>459</td>
<td>130</td>
<td>US</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>orkut.com</td>
<td>381</td>
<td>58</td>
<td>US</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>stumbleupon.com</td>
<td>284</td>
<td>217</td>
<td>US</td>
<td>Search engine/User-contributed content</td>
<td></td>
</tr>
<tr>
<td>reddit.com</td>
<td>224</td>
<td>323</td>
<td>US</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>hatena.ne.jp</td>
<td>194</td>
<td>222</td>
<td>Japan</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>delicious.com</td>
<td>118</td>
<td>347</td>
<td>US</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>tuenti.com</td>
<td>84</td>
<td>358</td>
<td>Spain</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>xanga.com</td>
<td>55</td>
<td>1039</td>
<td>US</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>technorati.com</td>
<td>47</td>
<td>933</td>
<td>US</td>
<td>Blogging</td>
<td></td>
</tr>
<tr>
<td>folkd.com</td>
<td>35</td>
<td>1333</td>
<td>Germany</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>diigo.com</td>
<td>28</td>
<td>2414</td>
<td>US</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>meneame</td>
<td>22</td>
<td>3162</td>
<td>Spain</td>
<td>Social bookmarking/Social news</td>
<td></td>
</tr>
<tr>
<td>mister-wong.com</td>
<td>22</td>
<td>2723</td>
<td>Germany/US</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>wykop.pl</td>
<td>20</td>
<td>4635</td>
<td>Poland</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>moikrug.ru</td>
<td>19</td>
<td>3371</td>
<td>Russia</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>secondlife.com</td>
<td>16</td>
<td>3945</td>
<td>US</td>
<td>Virtual world</td>
<td></td>
</tr>
<tr>
<td>faves.com</td>
<td>15</td>
<td>4260</td>
<td>US</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>blinklist.com</td>
<td>13</td>
<td>3879</td>
<td>US</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>migente.com</td>
<td>12</td>
<td>14356</td>
<td>Latin America</td>
<td>Social network</td>
<td></td>
</tr>
<tr>
<td>Site</td>
<td>Views per millions</td>
<td>Rank</td>
<td>Country</td>
<td>Site classification</td>
<td>Note</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------</td>
<td>------</td>
<td>---------</td>
<td>----------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>connotea.org</td>
<td>12</td>
<td>5322</td>
<td>US</td>
<td>Social bookmarking/Academic</td>
<td></td>
</tr>
<tr>
<td>evernote.com</td>
<td>11</td>
<td>4380</td>
<td>US</td>
<td>Data management</td>
<td></td>
</tr>
<tr>
<td>newsxine.com</td>
<td>11</td>
<td>4248</td>
<td>US</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>librarything.com</td>
<td>11</td>
<td>4992</td>
<td>US</td>
<td>Social bookmarking/Books</td>
<td></td>
</tr>
<tr>
<td>bibsonomy.org</td>
<td>10</td>
<td>6175</td>
<td>Germany</td>
<td>Social bookmarking/Academic</td>
<td></td>
</tr>
<tr>
<td>linkarena.com</td>
<td>8</td>
<td>7662</td>
<td>Germany</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>citelike.org</td>
<td>7</td>
<td>11237</td>
<td>UK</td>
<td>Social bookmarking/Academic</td>
<td></td>
</tr>
<tr>
<td>yigg.de</td>
<td>7</td>
<td>8872</td>
<td>Germany</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>bitacoras.com</td>
<td>7</td>
<td>10124</td>
<td>Spain</td>
<td>Blogging</td>
<td></td>
</tr>
<tr>
<td>simpy.com</td>
<td>6</td>
<td>11080</td>
<td>US</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>choix.jp</td>
<td>4</td>
<td>8885</td>
<td>Japan</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>rambhai.com</td>
<td>4</td>
<td>16239</td>
<td>India</td>
<td>Social networking</td>
<td></td>
</tr>
<tr>
<td>news2.ru</td>
<td>4</td>
<td>15190</td>
<td>Russia</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>instapaper.com</td>
<td>4</td>
<td>15377</td>
<td>US</td>
<td>Data management</td>
<td></td>
</tr>
<tr>
<td>mister-wong.ru</td>
<td>3</td>
<td>36801</td>
<td>Germany/Russia</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>moemesto.ru</td>
<td>3</td>
<td>19742</td>
<td>Russia</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>mar.gar.in</td>
<td>2</td>
<td>1350242</td>
<td>South Korea</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>netvouz.com</td>
<td>2</td>
<td>26260</td>
<td>Sweden</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>blogmarks.net</td>
<td>2</td>
<td>21253</td>
<td>Germany</td>
<td>Social bookmarking</td>
<td>Registration closed/2010</td>
</tr>
<tr>
<td>ekudos.nl</td>
<td>1</td>
<td>47936</td>
<td>Netherlands</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>mendeley.com</td>
<td>1</td>
<td>52876</td>
<td>US</td>
<td>Data management</td>
<td></td>
</tr>
<tr>
<td>gnolia.com</td>
<td>1</td>
<td>36137</td>
<td>US</td>
<td>Social bookmarking</td>
<td>Closed/2010</td>
</tr>
<tr>
<td>scoopeo.com</td>
<td>1</td>
<td>82877</td>
<td>France</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>fuzz.fr</td>
<td>1</td>
<td>89014</td>
<td>France</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>pinboard.in</td>
<td>1</td>
<td>239353</td>
<td>Israel</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>zootool.com</td>
<td>0</td>
<td>310870</td>
<td>Germany</td>
<td>Data management</td>
<td></td>
</tr>
<tr>
<td>mister-wong.fr</td>
<td>0</td>
<td>169626</td>
<td>Germany/France</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>mister-wong.cn</td>
<td>0</td>
<td>433966</td>
<td>Germany/China</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>bookmarks.fr</td>
<td>0</td>
<td>357447</td>
<td>France</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>rawsugar.com</td>
<td>0</td>
<td>156514</td>
<td>Israel</td>
<td>Social bookmarking</td>
<td>Closed</td>
</tr>
<tr>
<td>dir.eccion.es</td>
<td>0</td>
<td>487839</td>
<td>Spain</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>tutmarks.com</td>
<td>0</td>
<td>322708</td>
<td>France</td>
<td>Social news/tutorials</td>
<td></td>
</tr>
<tr>
<td>tagmos.nl</td>
<td>0</td>
<td>486454</td>
<td>Netherlands</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>linkologia.pl</td>
<td>0</td>
<td>303711</td>
<td>Poland</td>
<td>Social news</td>
<td></td>
</tr>
<tr>
<td>favoriting.com</td>
<td>0</td>
<td>1123512</td>
<td>Spain</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>bookeet.com</td>
<td>0</td>
<td>986519</td>
<td>France</td>
<td>Social bookmarking</td>
<td></td>
</tr>
</tbody>
</table>

*Continues next page...*
<table>
<thead>
<tr>
<th>Site</th>
<th>Views per millions</th>
<th>Rank&lt;sup&gt;21&lt;/sup&gt;</th>
<th>Country</th>
<th>Site classification</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>miselegidos.com</td>
<td>0</td>
<td>1428526</td>
<td>Spain</td>
<td>Social bookmarking/Broadcaster</td>
<td></td>
</tr>
<tr>
<td>misfavoritos.es</td>
<td>0</td>
<td>1767957</td>
<td>Spain</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>Utx.ambience.ru</td>
<td>0</td>
<td>1817158</td>
<td>Russia</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>keepyourlinks.com</td>
<td>0</td>
<td>N/A</td>
<td>Spain</td>
<td>Social bookmarking</td>
<td></td>
</tr>
<tr>
<td>weibo.com</td>
<td>0</td>
<td>1747058</td>
<td>South Korea</td>
<td>Microblogging</td>
<td></td>
</tr>
<tr>
<td>jaka.fr</td>
<td>0</td>
<td>N/A</td>
<td>France</td>
<td>Social news</td>
<td>Closed</td>
</tr>
<tr>
<td>watvindenwijover.nl</td>
<td>0</td>
<td>N/A</td>
<td>Netherlands</td>
<td>Social bookmarking</td>
<td>Closed</td>
</tr>
<tr>
<td>contarte.info</td>
<td>0</td>
<td>1800304</td>
<td>Spain</td>
<td>Social bookmarking</td>
<td>Closed</td>
</tr>
<tr>
<td>redcreo.net</td>
<td>0</td>
<td>1842408</td>
<td>Spain</td>
<td>Social bookmarking</td>
<td></td>
</tr>
</tbody>
</table>

<sup>21</sup> Alexa daily ranks averaged over January 2010
D. Extract from Project Brief

Specification for scoping the need for an academic social bookmarking service

Introduction and Background

Social bookmarking is a method for Internet users to store, organize, search, and manage bookmarks of web pages on the Internet with the help of metadata. Such metadata is typically in the form of tags that collectively and/or collaboratively become a folksonomy. A folksonomy is also described as social tagging, ‘the process by which many users add metadata in the form of keywords to shared content’ (Golder et al., 2006).

Within the JISC IE programme a number of projects at EDINA and MIMAS have started to explore the potential of social bookmarking services. This is as a result of recommendations arising from the JISC Developing Personalisation for the Information Environment (DPIE1) and Developing Personalisation for the Information Environment DPIE2 projects (Details of the projects at EDINA and MIMAS will be supplied with this specification).

This specification is for UKOLN to undertake work to scope whether it is desirable to build specific academic social bookmarking services as part of the JISC IE, or whether it is appropriate to use existing social bookmarking services.

Aims and Objectives:

1. Evaluate the social bookmarking demonstrator work that is currently being done at EDINA and MIMAS as part of the Personalisation work
2. Evaluate social bookmarking services which already exist on the web—and have gained a ‘network’ effect such as del.icio.us (now called Delicious), Citeulike, Connotea, etc.
3. Clearly identify the advantages and disadvantages of establishing an academic bookmarking service or for JISC services to simply use/rely on those that have already gained the network effect on the web.
4. Identify technical and architectural issues in both approaches
5. Identify barriers and enablers in both approaches
6. Identify issues surrounding sustainability in both approaches