education.au
creative capable connected

FOLKSONOMIES AND TAGGING:

New developments in social bookmarking

Sarah Hayman
education.au

Ark Group Conference: Developing and Improving Classification Schemes
27-29 June, 2007
Rydges World Square, Sydney

Abstract

What is the role of controlled vocabulary in a Web 2.0 world? Can we have the best of both worlds: balancing folksonomies and controlled vocabularies to help communities of users find and share information and resources most relevant to them?

education.au develops and manages Australian online services for education and training. Its goal is to bring people, learning and technology together. ***education.au*** projects are increasingly involved in exploring the use of Web 2.0 developments building on user ideas, knowledge and experience, and how these might be integrated with existing information management systems. This paper presents work being undertaken in this area, particularly in relation to controlled vocabularies, and discusses the challenges faced.

Education Network Australia (**edna**), managed by ***education.au***, is a leading online resource collection and collaborative network for education, with an extensive repository of selected educational resources with metadata created by educators and information specialists. It uses controlled vocabularies for metadata creation and searching, where users receive suggested related terms from an education thesaurus, with their results. We recognise that no formal thesaurus can keep pace with user needs so are interested in exploiting the power of folksonomies.

This paper describes a proof of concept to develop community contributions to managing information and resources, using Taxonomy-Directed Folksonomy. An established taxonomy from the Australian education sector suggests terms for tagging and users can suggest terms. Importantly, the folksonomy will feed back into the taxonomy showing gaps in coverage and helping us to monitor new terms and usage to improve and develop our formal taxonomies.

This model would initially sit alongside the current **edna** repositories, tools and services but will give us valuable user contributed resources as well as information about how users manage resources. Observing terms suggested, chosen and used in folksonomies is a rich source of information for developing our formal systems so that we can indeed get the best of both worlds.

Table of Contents

1.1 Overview of folksonomies and tagging: a brief history, current developments and directions

The term Web 2.0, first used by Tim O'Reilly in 2004, describes a cluster of web-based services with a social collaboration and sharing component, where the community as a whole contributes, takes control, votes and ranks content and contributors. Web 2.0 services include social networking sites, wikis, communication tools, weblogs, social bookmarking, podcasts, RSS feeds (and other forms of many-to-many publishing), social software, and folksonomies. Central to this new Web is the idea of tagging — the adding of keywords to a digital object (e.g. a website, picture, audiofile or videoclip) to categorise it. This activity is effectively subject indexing but generally without a controlled vocabulary.

Tagging of course is not a new concept, especially to librarians, indexers and classification professionals. What is new is that the tagging is being done by everyone, no longer by only a small group of experts, and that the tags are being made public and shared. The development of the internet and the web, and of search engines, led to users doing their own searching. In the Web 2.0 environment users are now also doing their own content creation and information management.

The PEW internet survey of December 2006 (Rainie, 2007) found that 28% of internet users have tagged or categorised content online such as photos, news stories or blog posts. On a typical day online, 7% of internet users say they tag or categorise online content.

Tagging is used in a range of sites for many different types of resources. Tagging is done somewhat differently at different websites, but the following all use some type of user tagging:

Blogs (Technorati: http://technorati.com/)
Bookmarks (Delicious: http://del.icio.us/)
Books (Librarything: http://www.librarything.com/, Amazon: http://www.amazon.com/)
Emails (Gmail: http://mail.google.com/)
Events (http://www.goingtomeet.com/)
People (Tagalag: http://www.tagalag.com/)
Pictures (Flickr: http://www.flickr.com/)
Podcasts (Odeo: http://odeo.com/))

Videos (YouTube: http://www.youtube.com/)

Even perhaps tagging of tags? (http://tagtagger.com/)

In user tagging, after an account has been created a user can apply a tag (or label, or keyword) to a resource; it may be a website, a photograph or video, or a record for a book as in Librarything.

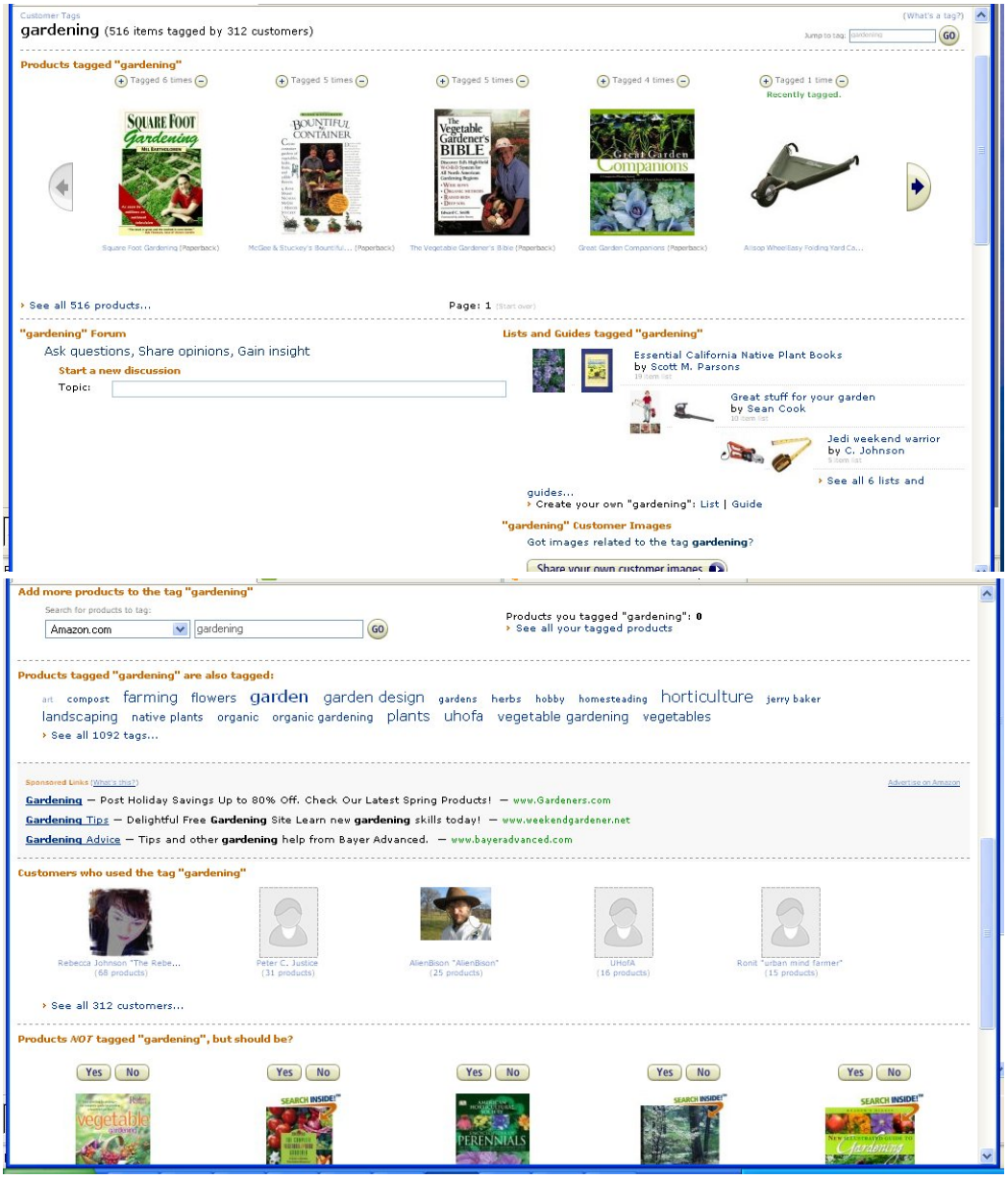
The user chooses a tag that is meaningful to him or her. In most sites it has to be a single word – more about that later. A large number of tags can be applied (e.g. in Flickr the maximum number at the time of writing is 75). Once the tags have been assigned, they act as index terms and they may be public or private. When they are public, the tags together can all be searched by all users, creating a "folksonomy".

It is important to remember that users have complete freedom in the tags they choose and may assign tags for their own organising purposes, without regard to any other users who may wish to make use of them. Even if this is the case, there may still be valuable information in the collection of tags that develops. In many cases however, users are keen to share their tags and will choose tags that others have also used.

Users can add their own tags to already tagged resources. They may use a different word for the same concept or a broader or more specific word for a related concept. The aggregation of all the tags allows a site like Flickr to organise resources better for all users, and also informs the site owners about the popularity of tags and of resources. This can be described as a bottom-up rather than top-down building of categories.

Tags, once assigned, can be grouped, shared, displayed, published and managed in several ways. Typically tags are displayed in a "tag cloud" on many sites, where the graphical display indicates by size, font or colour how many times the tag has been used or how many resources have been assigned that tag. It is possible to see all tags assigned to a resource, all people who have used a particular tag, other tags that have been used for similar items, popular tags, recent tags etc. Del.icio.us allows users to manage their tags by bundling them and renaming them, and provides for a rudimentary hierarchical structure using the "/" between tags. Clearly, tagging is potentially a very powerful information management resource.

It is interesting to see what Amazon offers users in relation to tags. See for example this page on Amazon for the tag *gardening*, at http://www.amazon.com/tag/gardening



The page shows the name of the tag, how often it has been used and by how many people; it provides information about the first 5 of those 516 products tagged *gardening*, with an offer to see all 516 and for each one the opportunity to tag it or remove it; it shows items in this set that were recently given this tag; it shows the users their own tags and tagged products; it shows a tag cloud for products that have been

tagged *gardening*; it shows customers who have used the tag; it suggests other products that could also have been tagged *gardening*.

The last item is interesting as an example of an element of control being added by the service to the collection management that users have done – a suggestion is made of other items that also could be assigned the same tag. Here Amazon's own classification is being used to identify and suggest other items on the same subject – but it is only the users who decide to add a tag.

The selection of options above also illustrates the three distinct entities in the world of tagging: the tag, the item being tagged and the person doing the tagging. They all have a separate existence but vital relationships to each other that can be exploited for information. Amazon has compiled and presented data about all three: the products tagged (and how often and how recently), the tag itself (noting which other tags have also been used with this tag) and the person doing the tagging (others who have also used the same tag). This richness of data has huge potential to provide multiple layers of information, and we discuss later in this paper how we aim to employ a similar architecture in our proof of concept model for tagging educational online resources.

1.1.1 Language about tagging

In an emerging field, names of new phenomena are not always clear and many different terms are currently used for the activity of tagging and the results of the tagging activity.

Tagging has come to be the most frequently used name for the action of applying a label to an item; however some sites do talk about labels and others enable tagging seamlessly so that it may not be apparent that it is being used.

A definition of tagging from Wikipedia:

> "A tag is a (relevant) keyword or term associated with or assigned to a piece of information (like picture, article, or video clip), thus describing the item and enabling keyword-based classification of information it is applied to." (Wikipedia, 2007)

The word tag has also been used for some time in the context of HTML (hypertext markup language) where it refers to formatting codes used in HTML documents. HTML tags indicate how parts of a document will appear when displayed by browsing software.

The terminology about user tagging is still fairly fluid and many terms for the same phenomena are being used, often in slightly different ways, with debate starting about the exact usage and meaning of the terms. These terms currently include:

> Collaborative tagging, shared tagging, user tagging, social bookmarking, collaborative bookmarking; folksonomies, tagsonomies, tagonomies, collabularies, tagosphere, folksonomic zeitgeist.

The term folksonomy was coined by Thomas Vander Wal in 2004 to signify what he called a “user-generated classification, emerging through bottom-up consensus”. It is a fusion of the words folk and taxonomy.

There is debate about the nature of these concepts and terms. Some writers have distinguished between a *folksonomy* (a collection of tags created by an individual for personal use) and a *collabulary* (a collective vocabulary). Other writers however use *folksonomy* to mean a collective vocabulary.

In this paper we use *tagging* to refer to labelling of web items, *user tagging* when that tagging is done by the user, and *folksonomy* to refer to the collection of user tags.

1.1.2 Different approaches to tagging

Elements to consider are: who does the tagging?; is it collaborative? is it intended for categorisation? what use is made of the tags?

Different sites use tagging in different ways. Tagging is not always done by users, as in the sites listed above. In some cases tagging is done by “experts” with the results still displayed as a tag cloud. An example of this approach is Surf the News.com (http://news.com.com/2243-12_3-0.html) where News.com editors tag each published story with one or more topics. Here the tagging is done not by users but by news editors. In the tag cloud, the most assigned topics are large and red and those used least are small and gray. The data is sorted by most-used topic to least-used topic. The data is shown for the last 30 days of stories. If a topic is not assigned within that time, the topic will not appear on the page. This is tagging but not folksonomy.

Another approach to tagging is to develop an application that will take RSS feeds and analyse the keywords in the feeds to produce tag clouds. An example of this was NewsCloud (http://www.revsys.com/aboutnewscloud/), an application that took all of the RSS feeds from the Washington Post website and built a tag cloud from the keywords. Each story's full text was pulled from the website and indexed by these

keywords. There were typically around 23,000 news stories and 71,000 keywords being indexed at any given time. This is an example of machine tagging.

A similar application designed for people to put on their own sites to create tag clouds is ZoomClouds (http://zoomclouds.egrupos.net/cloud/ZC_CNN/). A web site owner can use ZoomClouds to put small (or large) tag clouds on a website and control the look and feel of the tag cloud to provide visitors with a way to see what terms are more often mentioned in the website. Again, these are examples of tagging but not folksonomies.

Even when user tagging is the activity on the site, different users will be behaving differently. Some users choose tags they wish to make public and to share. Other users keep their tags private and simply use this as a way of storing and organising information for their own purposes. You can have a completely private area in Del.icio.us as a way of managing your own bookmarks simply so that you can access it from anywhere but without making it a social networking activity. You can even search other people's tags if you wish and still not share your own.

Other users may make their tags public but deliberately choose private words so that they know few others will use them and they can keep control over a manageable number of items tagged. For example, a recent blog posting stated:

> "Tags are not for taxonomists. They are not for the creation of some enormous world-wide content-finding system. They are PERSONAL. It's nice that they can provide some benefit in a social context, but I know what *my* 'frob' tag specifies, and I frankly don't care what anyone else's does. When I want to find something, I use *search*." (Feinberg, 2007).

In many cases tags are designed to be shared but only within a small group or already defined community which may have its own idiosyncratic use of language and wish to limit the number of users of the tags in order to control their own information access and management. This is a type of folksonomy but for a restricted group.

The ***education.au*** team working on the **myedna** project share a common invented tag, *myednapoc,* to inform each other about websites and resources of interest for discussion. (See http://del.icio.us/tag/myednapoc/).

Our view is that search facilities and collection management can be enhanced and enriched by noting and analysing what users' tags do. Purely personal ones tags, like the example *frob* above, and narrow interest group folksonomies would be excluded through low frequency but it can be of value to know what those users who do like to share their tags choose as the most meaningful words. We can thus obtain information about current linguistic usage as well as about topics of current interest.

Many tagging sites cater for multiple levels of interaction and tagging activity. Although private uses of this type of facility are also valid, it is, we argue, only when tags are publicly shared that a folksonomy develops.

1.1.3 Folksonomies

A folksonomy is essentially the name given to a collection of tags built up by the action of user tagging, effectively a user generated taxonomy as opposed to an authoritative hierarchical taxonomy like Library of Congress Subject Headings or a subject thesaurus. A key feature of a folksonomy is that tags may be reused many times, providing information about the popularity of the tags themselves (which synonyms come to be more popular over time) as well as information about emerging areas of interest.

The essence of folksonomies is that the tags allocated are chosen by the user. This is a fascinating new development for those of us who have been working in the field of controlled vocabularies and we are keen to see what terms users will choose for tags, how they will use them and how they will organise them. What new vocabularies and taxonomies will emerge from these clouds of tags developed and used by anyone who has an interest and inclination to do so?

There are many studies already emerging in this area. These include an examination of numbers of people using tags (Rainie, 2007), a project investigating the potential of folksonomy in academia (TNN, 2007) and a comparison of the use of tags in Librarything and Amazon that has generated considerable discussion (Spalding, 2007). A study analysing the structure of collaborative tagging systems found “regularities in user activity, tag frequencies, kinds of tags used, bursts of popularity in bookmarking and a remarkable stability in the relative proportions of tags within a given url.” (Golder and Huberman, 2006).

1.2 Contrasting folksonomies and formal taxonomies (some benefits and disadvantages)

It is worth reviewing some features of folksonomies and comparing them to formal classification systems.

1.2.1 Benefits of folksonomies

The following are characteristics of tagging and folksonomies that can be seen as beneficial features.

They are multidimensional: users can assign a large number of tags to express a concept and can combine them.

Users can use their own language: words that have meaning for them. These words are likely to be current and reflect local usage.

Users select concepts that have meaning for them as individuals and analyse items to highlight what is important to them.

Tags can be shared, creating knowledge through aggregation.

> "We now have millions and millions of people who are saying, in public, what they think pages and images are about. That's crucial information that we can use to pull together new ideas and information across the endless sea we've created for ourselves." (Rainie, 2007)

Instead of having to store an item in a single folder, it can be tagged with many different terms and each of these could be used to generate an instant collection (e.g. if a collection of photographs contains photographs with tags such as birthday, family, holiday, Europe, sub-collections can be readily assembled by searching for single tags or pairs.)

Public tagging has been described as having an altruistic appeal, allowing people to contribute to a shared knowledge base. Social tagging fosters the development of communities around similar interests and viewpoints.

Social tagging provides information to professional providers and managers of information about areas of interest and how they are being described. It is a new window on the way our users are thinking and can provide insight into their information needs and habits.

Tagging is very quick, simple and straightforward. Users can apply tags without formal training in classification or indexing.

Clay Shirky has identified a number of advantages of tagging systems, including:

> "Market Logic: […] where you deal with individual motivation, but group value.
>
> User and Time are Core Attributes: […] because you can derive 'this is who this link is was tagged by' and 'this is when it was tagged, you can start to do inclusion and exclusion around people and time, not just tags. You can start to do grouping. You can start to do decay.
>
> Signal Loss from Expression: […] in a world where enough points of view are likely to provide some commonality, the aggregate signal loss falls with scale in tagging systems, while it grows with scale in systems with single points of view." (Shirky, 2005)

As folksonomies grow, the larger scale can bring some organisation into the tagging process; judicious users will evaluate tags and tend to use existing tags to assist with forming useful connections. Thus the folksonomy can develop its own tagging conventions through group consensus rather than an externally imposed and possibly dated formal system.

1.2.2 Disadvantages of folksonomies

It will be readily apparent that many of the features of folksonomies listed above as advantages can also lead to problems for effective classification and information management.

The simplicity and ease of use of tagging can result in poorly chosen and applied tags. While it could be argued that this is a necessary feature of user tagging and insignificant, nevertheless, the following issues need to be considered.

Tags can be applied at different levels of specificity by different users (or even by the same user at different times) e.g. the tag *cats* may be used in one case and *animals* or *pets* in another. Or the tag *Kitty* may simply be used.

Different terms may be used for the same concept (again by different users or by the same user – users will not necessarily be consistent). So *felines* may be used for some items and *cats* for others.

A person searching for pictures of cats will have to use many different terms to be sure of finding all items.

Tags with personal meaning only are frequently used (example on Flickr: *viewfrommywindow*). This tag on its own is of virtually no use to anyone else.

Conversely, the same term can be used for different concepts. Typically, no information about the meaning of a tag is provided (although some systems, del.icio.us being one, do allow tag descriptions; see http://del.icio.us/help/tagdescriptions). The word *play* could occur in an educational resource collection in the drama context or the games context. The word *tag* itself has more than one meaning. Without even considering the issue of other languages, English itself has a huge number of words with multiple meanings. Vocabularies have been built for specific communities where the meanings chosen are appropriate for that context. To some extent this will also apply in user tagging on the internet, but even within communities there can be ambiguities of meaning.

Uncontrolled tagging can result in a mixture of types of things, names of things, genres and formats. Many of these problems can arise even with specialist indexers, for example using *video* as a subject heading when the item is a video, when it should only have that subject heading if it is *about* videos. If it is already difficult for people to comply with requirements such as these, it will be far more difficult to have precision when there are no indexing guidelines other than those developed by individual users for their own practice and unlikely to be made explicit.

Regular indexing and cataloguing rules such as singular vs plural forms, use of hyphens and spelling conventions are not established in a folksonomy.

People's choice of tags may change as new trends evolve — e.g., it is likely that *blog*, *weblog*, *blogs* and *blogging* will all be used for the same concept.

Many systems only allow single word tags. It may be difficult to difficult to assign terms to complex concepts using only a single word and running two or more words together is difficult in many ways – the resulting words will be highly idiosyncratic and difficult to read and to search with precision.

Social tagging systems are vulnerable to spam and malicious practice.

A more subtle issue is that people may behave differently (consciously or unconsciously) when tagging other people's items as opposed to their own. The objectivity of a professional indexer is not necessarily a feature of social tagging.

Another high level concern is that over time tags may come to represent a dominant view, discouraging usage of less popular concepts (and terminology) which become disproportionately overwhelmed by the majority. Users will tend to use popular tags and may not realise that there is a more precise term available for their concept, or may be subtly discouraged from creating their own tags. Zeldman writes: “Network effects being exponential, what is immediately mildly popular quickly becomes artificially very popular, while what has yet to become popular never will be.” (Zeldman, 2005)

Overall, social tagging and tags are uncontrolled and tags are not connected to each other by a reference structure, which in formal systems is used to link related terms and narrower or broader terms. The creation and application of tags by users who are not experts in information management lead to the problems described above.

However there are also clearly great benefits in user tagging and folksonomies, especially in the richness, currency, relevance and diversity of the terms used, and the collections of resources created. It is important to try to retain those qualities in any attempt to control folksonomies.

Recent research by Kipp and Campbell shows some interesting clustering patterns emerging from user tagging. These writers suggest it may be complementary to conventional indexing, with the two approaches enriching each other. Further, they consider whether “user tagging extends beyond the traditional objectives of subject access, and expresses a dynamic relationship between document and user, and between subject and task, which may lead to new ways of modeling subject access” (Kipp and Campbell, 2006) write:

1.2.3 Combining folksonomies and formal classification

Is it possible to combine the two approaches and gain benefits from both? Some attempts have been made already and a few are mentioned here in a consideration of some future developments for social tagging. We then discuss our own model: the taxonomy-directed folksonomy for the **myedna** proof of concept.

In some ways it is too early to tell what will be the results of the Web 2.0 developments and in what directions social tagging and folksonomies will go. One prediction by David Weinberger about social tagging:

> “Because it's useful when there's lots of information and the information is truly meaningful to individuals, it'll be adopted more and more widely. But we're also going to invent new ways to harvest tagging. Flickr, for example, is already able

> to cluster photographs by subject with impressive accuracy just by analyzing their tags, so that photos of Gerald Ford are separated from photos of Ford Motor cars. We'll also undoubtedly figure out how to intersect tags with social networks, so that the tags created by people we know and respect have more 'weight' when we search for tagged items. In fact, by analyzing how various social groups use tags, we can do better at understanding how seemingly different worldviews map to one another." (Weinberger, 2007)

Another development, described in Wikipedia:

> "Although 'tagging' is often promoted as an alternative to organization by a hierarchy of categories, more and more online resources seem to use a hybrid system, where items are organized into broad categories, with finer classification distinctions being made by the use of tags." (Wikipedia, 2007)

Some libraries are allowing users to tag catalogue items. In the academic sphere, the University of Pennsylvania's Penntags project has been developed for readers to tag catalogued items. It enables them to track resources for a research project and simultaneously make the results available to future researchers. (University of Pennsylvania, 2007)

In the public library sector, Ann Arbor District Library (AADL) has developed a set of social networking tools called the SOPAC, integrated into the library catalogue. It gives users the ability to rate, review, comment on, and tag items. (Blyberg, 2007)

1.3 A proposed model for combining a folksonomy approach with a taxonomy for information management in the education sector: a taxonomy directed folksonomy proof of concept model

1.3.1 Nature of proof of concept

As mentioned above, ***education.au*** aims to develop a **myedna** service where users can contribute, customise, manage and share their own resources. As a new development in the Web 2.0 era this will aim to make the best use of collaborative

technologies and philosophies. We have begun this process by developing a proof of concept model to test the concept, develop the skills of the team and explore possibilities without being limited by our capabilities at this early stage.

A proof of concept model is a model that explores what could be achieved, before a decision is taken about whether it is produced and implemented. It explores possibilities and develops skills internally without the commitment of large scale resources initially. It is ideally suited to a time of rapid technological development as it can be fast and flexible.

> "[It is] a high risk, high trust and low governance project that creates a conceptual solution for the client. Focus solely on the concept. Forget security. Forget firewalls. Forget QA etc. What is delivered is a bare-knuckled prototype which demonstrates how it could solve the issues your client deals with. The project has no limits other than Time (approx 3 months) and Resource a small team working part time within a small financial budget. Concluding each POC month is a 'Show and Tell' to an audience of sponsors, users and internal teams - each one being more public than the last.
>
> Project Outcomes:
>
> - Innovation: new ideas and new approaches from more directed conceptual thinking.
>
> - Better starting position - much closer to the solution suited to the user needs.
>
> - Better position to ask really pointed questions and address issues such as accessibility, security, privacy and quality.
>
> - Agile in response to ideas and feedback particularly with shorter feedback times" (Cotton, 2007)

1.3.2 Key features of the myedna proof of concept

The aim is to develop a service based on the notion of sharing learning or "Watch me learn". It will be a personal learning space that will:

- record a learning journey
- accommodate all types of learning: formal, informal, "a-ha moments", work-related
- provide an online space organised by a person to meet his or her needs
- remind users what they have learned
- allow people to share information, resources, stories and narrative (in many formats) with multiple audiences

- record assets only once, and reuse them, catalogue, edit, comment and share for different contexts.

The interface will be customisable by the user to reflect individual needs, interests, preferences and whims. As a function provided by a national education service, it should be a public space so that each person's personal learning journey will add to the community's knowledge and learning. It will be a portal containing a range of features, one of which will be a facility to "stitch" RSS feeds together. It will provide for storage of documents, images and other files that users will be able to access from anywhere any time.

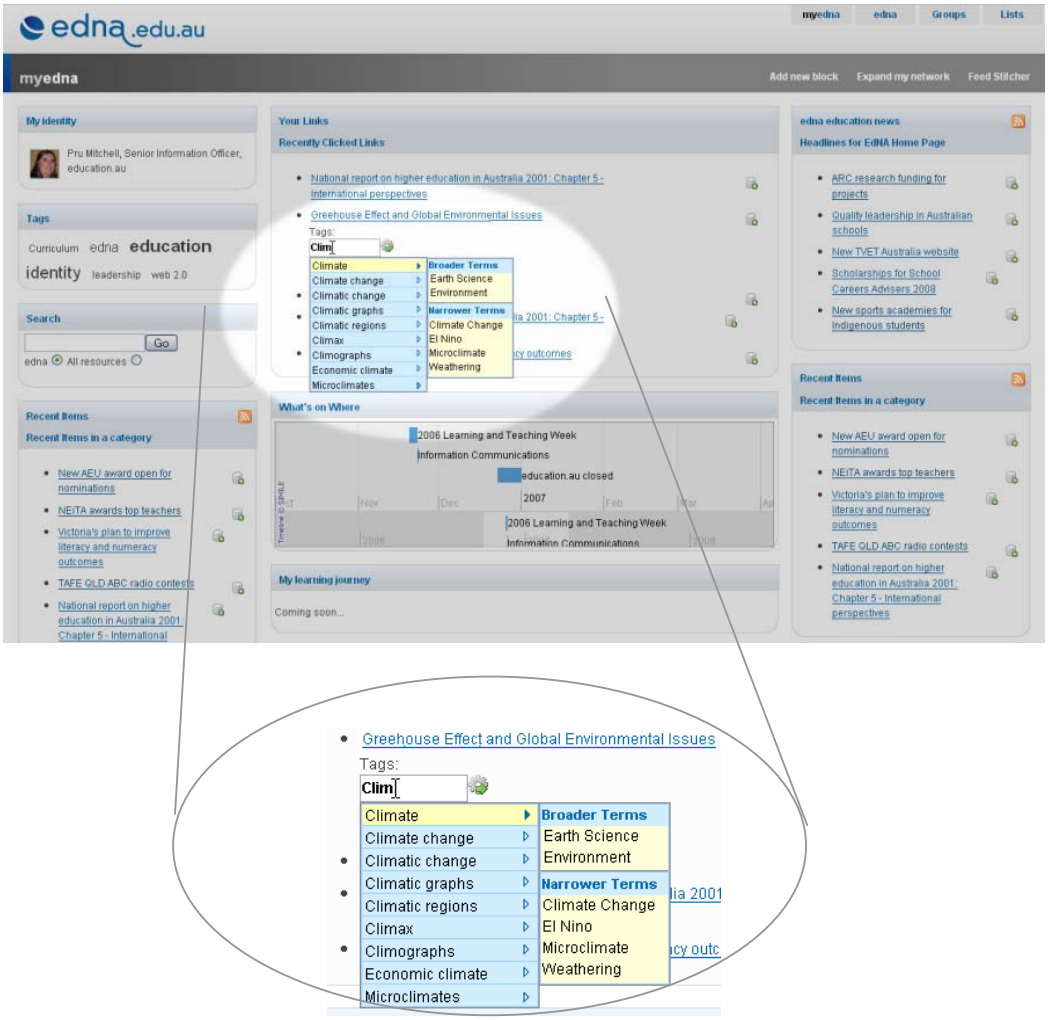
It will also include a section called **edna Links** where users can collect, store and label information about favourite sites as bookmarks using tags. Our proof of concept model for this area includes the feature we have named the Taxonomy-directed Folksonomy.

1.3.3 The taxonomy-directed folksonomy

Tagging was described above as a type of subject indexing, performed by users, without a controlled vocabulary. In our **myedna** proof of concept model we are aiming to combine user tagging with a controlled vocabulary and harness the best of both worlds. A recent article about the concept of Library 2.0 stated:

> "tags and standardized subjects are not mutually exclusive. The catalog of Library 2.0 would enable users to follow both standardized and user-tagged subjects; whichever makes most sense to them. In turn, they can add tags to resources. The user responds to the system, the system to the user. This tagged catalog is an open catalog, a customized, user-centered catalog. It is library science at its best." (Maness, 2006)

This is similar to the philosophy underlying the proposed **myedna** development.



In the **edna Links** area of the **myedna** portal, users will be provided with a box in which to enter their own tags for resources. As they type a tag, they will be prompted by a thesaurus, which will suggest terms that match the term they have entered.

The algorithm for matching terms is still under development. We are exploring ways to mesh taxonomy based functionality with the ease of use that users are familiar with from a social tagging context. Taxonomy based functionality which we are looking at includes:

- exact match, suggestions for broader or narrower terms, or prompt with a term which contains the user's term in its scope note
- choice of thesaurus or one may be supplied to fit a user's signon if he or she is from a particular educational sector.

We are using the Schools Online Thesaurus (ScOT) - used for schools for this project. **edna** makes use of a number of specialist education thesauri in addition to ScOT, including VOCED - a vocational education research thesaurus and ATED (Australian thesaurus of education descriptors) – particularly suitable for higher education. As each of these thesauri adheres to the controlled vocabularies standard, ANSI/NISO Z39.19-2005 (National Information Standards Organization, 2005), it would be technically feasible for users to be given a choice of thesauri for tagging.

Users may still choose to use their own terms. Tags will be collated over time and a tag cloud produced and displayed. The tags in the clouds will come both from user tags and from tags selected from thesauri. This collection of tags will be a folksonomy that has been directed by a taxonomy.

A possible future development would be to give users information about the tags as they are about to choose one – so they not only see related tags but also scope notes or guidelines for usage. This would give further taxonomic direction to the folksonomy.

The folksonomy thus created will generate valuable information.

Information about the tags

it will indicate which thesaurus terms are useful to our users

it will indicate new terms for existing concepts that should be considered for our thesauri (either as preferred or non-preferred terms)

it will indicate new concepts and suggest terms for them

Information about the items tagged (resources)

it will indicate which items are considered of value by our users simply because of the number of times they have been tagged, and, if ratings are included, which are valued by valued taggers

Information about the people doing the tagging

it will indicate what tags and items a person has used, and each person will have a profile about their learning journey

In our current model for **edna Links** we propose to display results for each of the three facets or dimensions we have identified: Tags, Resources and People. If you select a

tag you will see all items assigned that tag and all people using that tag. If you select a resource you will see all tags assigned to that resource (as a tag cloud), all people who have tagged that resource, and ratings given to that resource. If you select a person you will see all tags they have used and all resources they have tagged.

We have also included the facility for users to make comments on resources, tags and people, rate resources and contribute to threaded discussions.

Guy and Tonkin have written:

> "One missed area of opportunity is that of more discussion tools through which users can share reasons for tagging things in a certain way. At the moment there is little discussion on folksonomy sites about the appropriateness of tags. Most of the sites do not offer the opportunity to provide actual text feedback, though some allow you to change other users' metadata." (Guy and Tonkin, 2006)

In this proposed model we have built in the opportunity for users to: identify, bookmark and evaluate resources of interest to their community; choose tags from an appropriate existing formal taxonomy; suggest new tags; comment on the tags; comment on the resources; find other users with similar interests; discuss any of the above.

1.3.4 Future directions

myedna is a project that aims to develop a culture of active user participation enabling our organisation to adapt itself, interact with our community and respond to the needs of our users through a participatory cycle of feedback, service development and reevaluation.

The taxonomy-directed folksonomy described above is one of the key features that will allow us to make that aim a reality. It is our hope that it will move from a proof of concept to a fully developed facility so that we can test it with our users and build a taxonomy-directed folksonomy over time populated with real data. The information we collect about the terms our users choose or create, and suggestions they make about the terms, will inform future formal taxonomy developments, especially in areas of new terminology and concepts, such as elearning. The information we receive from them about resources will enable us to enhance our collection development and management to ensure up-to-date and extensive coverage.

Our anticipated outcomes are:

- Educators will be able to manage their educational resources with tags meaningful to them and others
- A greater number of resources will be identified and shared by the wider **edna** community
- There will be more consistent categorisation of resources
- Resources identified by the community will come to light for full cataloguing by ***education.au*** specialist indexers
- The **edna** community will form fluid social networks and discussions around tags and resources.

In this project we have aimed to make the best possible use of emerging technology and current practice in the Web 2.0 environment by exploring how we can unite it with best practice in formal classification and information management to improve outcomes for the Australian education community.

1.4 Social tagging in the enterprise

Much of the discussion thus far is applicable to the special case of social tagging used for classification within an enterprise, but there are some particular issues and challenges associated with organisational information which are worth considering. There is interest in this area and a considerable amount of recent writing about it, but only a few instances of implementation. The references listed at the end of the paper contain some recent discussions. As we have already done for social tagging in general, we can consider the benefits and disadvantages of folksonomies in the context of business information and look at some examples, and consider some of the issues and challenges that have been noted.

1.4.1 Benefits of social tagging in the enterprise

Folksonomies are user-generated, using users' own words which are meaningful to them and potentially alleviating frustrations with rigid taxonomies, email overload and inadequate search mechanisms. This is a "grass sorts" approach to managing knowledge in organisations.

Tagging can be relatively inexpensive and quick and simple to implement.

Because tags are created by employees, they can be useful in facilitating workplace democracy. This form of expertise location can encourage collaboration and sharing of resources within the company.

Using folksonomies, where employees do their own tagging of resources, documents and websites, can enable distribution of management tasks for greater efficiency. Tagging distributes the time consuming process of creating and classifying content beyond the content creator.

The activity of collaborative tagging in the enterprise can enable the formation of social networks around tags or topics. These networks will reflect the interests and expertise of users contributing to the tag. The tagging activity of a user contributes to the expertise of the user and influences the expertise of other users.

Tagging can be a way of identifying expertise and interests beyond the confines of narrow job descriptions and foster multiskilling and greater interest and work satisfaction for employees.

It is a way of capturing employees' knowledge and evaluation of various internet and intranet information resources. Collective knowledge in an organisation can be created and exploited, with cross-fertilisation occurring. This can be knowledge and experience that were previously unrecorded or recognised.

As an information resource, there are many potential integration points with other corporate applications.

While precision may be lacking in user tagging, in many practical usage scenarios the trade-off between simplicity and precision can be worthwhile.

Many niche areas can get better coverage as well. As the costs to create a new tag are low, it becomes feasible to add many tags freely to many content objects.

Enterprise social software can provide persistence, structure and transparency to otherwise transient informal interactions between workers in an organisation.

Allowing the contribution of personal and collaborative information management in the corporate intranet may be an opportunity to utilise fully the metadata features of corporate document and records management systems.

An excellent Gartner report (Andrews, 2007) identified some important and subtle "pros" and "cons" in relation to using social software in an organisation. I have cited them all in this paper, as they are points worth considering:

> "Serendipitous unexpected perspectives: When users demonstrate interest in information objects, they create a complex, interconnected trail of behaviors that reveal unpredictable associations. A user interested in maps of the Amazon river may also be interested in a particular chemical compound or family of compounds that the user believes may be valuable; another user's interest in the Amazon maps could deliver a valuable association between this user's and

the other user's research that they could never have discovered through conventional examination.

Organic changes in definitions: When documents are created, the author might be encouraged to tag them according to his understanding of their contents. Such an understanding will grow stale as time passes, and the documents may have ultimate value that the author could not have imagined at creation time. The contents of the document might not reveal that value, either, with the same creativity that a user might through explicit or implicit behaviors. A user could classify a medical case, for example, as potentially related to a newly named syndrome that had not been conceived at the time the document was written.

Expertise exploitation: People's behavior can affect relevancy results in a constellation of information about a person's interests that is potentially invisible in their authorship. Documents that a person reviews and searches the person conducts may reveal an incipient interest that the person is unable to demonstrate in formal documents yet, or that he or she need not demonstrate because the necessary documents already exist. Social search recording can reveal that person as an expert to future searchers.

Establish institutional perspective for the future: Social search can establish a pattern of knowledge and understanding about certain topics that may be revealed later, even if the users who populate its categories have moved on to new jobs or different companies. The pattern, recorded in explicit objects such as document rankings, can provide valuable future perspective on what was perceived as interesting or valuable within an enterprise." (Andrews, 2007)

Finally, Niall Cook has provided a good summary, listing 15 uses of corporate bookmarking:

1) Monitoring news/blog coverage of your company
2) Consumer and competitor research
3) Identifying subject matter experts within your company
4) Connecting with people who share your interests
5) Seeing what colleagues are finding interesting right now
6) Subscribing to links that your team members are bookmarking
7) Finding the most popular pages on your intranet
8) Mine tags to classify content and supplement intranet search
9) Seeing what a piece of intranet content actually means to staff
10) Searching the collective corporate brain (rather than documents that get "published")

11) Use the tagging folksonomy to refine the corporate taxonomy
12) Capture and share information about clients and prospects
13) Track industry trends
14) Analyse the connections between employees across teams and geographies
15) Identifying potential hires and tracking alumni (Cook, 2007)

1.4.2 Disadvantages of social tagging in the enterprise

There can be a lack of precision in choosing terms, as described above (no synonym/antonym control, related terms, context, etc.).

There is no guarantee people will tag in a consistent, unbiased way, and there is even the possibility of intentional malice directed at an enterprise.

It is possible that the hierarchical nature of most enterprise organisation may make it difficult (will equal weight be given to all views in the social tagging mechanisms?). The informal and egalitarian nature of social software approaches may not fit well with a top-down decision-making structure.

Not all people enjoy or participate in social software – will only certain views be heard or captured?

People may have privacy concerns – may not wish their colleagues, managers or subordinates to see what they are interested in, viewing and tagging. The organisation may also have concerns about the visibility of a tag in highly confidential environments.

Some resources are only available through an Internet connection. Often, these resources are unavailable due to lack of connectivity to the Internet, corporate firewalls or corporate policy. In the case of internal resources, on an intranet or in a document management system, not all may be visible to all users making, a coherent folksonomy more difficult to develop.

Andrews, in the Gartner mentioned above, suggests the following potential problems with using “social search” in an organisation, quoted here in full:

> “Vulnerability to concerted damage efforts: If users have a selfish interest in results sets, then social search can enable them to aggressively pursue their agendas. Ranking documents for their relevancy can become a popularity contest or "log rolling" exercise. Tagging is particularly risky, because users can

very simply grasp the impact that their tags have on the prominence of their documents or the likelihood they will be found.

Ambiguity of implicit methods: When results of a search engine's installation are "wrong" or less relevant than expected, enterprises desire to fix what has gone wrong by "tuning" the engine, or resetting its capabilities to better the results. Social search methods, particularly those that rely on implicit ranking and behaviors, are more difficult to tune because of their comparative opacity. For example, an administrator with no prior baseline to measure against can be confused by how to value a user's bookmarking of a document versus the user's decision to print it out.

Desirability of a large universe or known universe: The more people who use a search installation, the better their interests may be measured, exploited and incorporated. A small number of users, especially a group without predictable commonality that will allow them to aid each other, cannot effectively exploit the capabilities of social search as well, because of the lack of engaged users. In fact, participant self-selection bias can be a problem, even in a larger installation. Which participants choose to involve themselves is a challenging element to measure, but, particularly as users invite compatriots to participate, the shared characteristics of the users must be considered in exploiting their behavior for relevancy." (Andrews, 2007)

Andrews states that despite these points, negative effects can be mitigated, and he stresses the value of any fresh perspective on relevancy (such as offered by social software) which will help an organisation ascertain appropriate documents for employees to find.

1.4.3 Examples of social tagging in the enterprise

There are some current examples of social tagging in organisations. Several writers have predicted that with the rise of social tagging outside the firewall, there will soon be a significant rise of social tagging inside the firewall, effectively enterprise tagging.

1.4.3.1 IBM Dogear

IBM has developed Dogear, an enterprise-scale social bookmarking system. Millen, Feinberg and Kerr have written an article that "describes the design challenges and early lessons learned from a friendly trial of the technology." (Millen et al, 2005)

Their article outlines the growth of personal bookmarks into social bookmarks and discusses whether organisations can benefit from social bookmarking tools. They

discuss desirable features that can help ensure the successful development of social bookmarking communities:

> Identity and Transparency (IBM has chosen to implement dogear using real names to facilitate communication among users of the application “since the various corporate collaboration tools (e.g., corporate directories, e-mail, chat,) all use real-name identities”)
>
> Alerting and Discovery (it is a very useful feature for users to be alerted when a resource has been tagged by a user of interest, or with a keyword of interest – Dogear does this)
>
> Designing for Extensibility: Enterprise Remixing (the design of the IBM system encourages enterprise remixing "by giving potential exploiters a number of easily parsed data formats from which to choose").
>
> Exploiting the Links. (An enterprise bookmarking system, over time, represents an inherent assessment of valued information resources. Collections of links can be exploited by the enterprise to augment enterprise search applications, supplement directory information about a particular user; and to augment workgroup online content. (Millen et al, 2005)

Bob Zurek, from IBM writes about his experience of Dogear as a user:

> “What I find most useful about the Dogear project is how it delivers and surfaces relevant information on demand that my colleagues have tagged while I’m using a search engine. This tagged information, essentially the social bookmark is surfaced right along side my search results. The value of the information from inside and shared by my colleagues has been very valuable”. (Zurek, 2006)

1.4.3.2 Cogenz

Cogenz Ltd was founded in 2006 to provide the Cogenz social bookmarking service to companies looking to harness social software for collaboration and knowledge management. A private beta was made available to selected test users in July 2006, followed in September 2006 by a corporate beta program. The first version of Cogenz was released in April 2007.

Cogenz is a commercial, hosted social bookmarking service for companies wishing to harness the collective intelligence of their employees using social software. Knowledge workers in an organisation can use Cogenz to:

- store the online resources - internet or intranet - they use to perform their jobs
- share them with colleagues across functions and geographies
- browse, search and track collective intelligence relevant to their needs
- identify experts and communities of interest

Unlike public social bookmarking services, this is all done through a private branded installation, controlled by the enterprise.

1.4.3.3 Connectbeam

Connectbeam is a commercial integrated social software platform for the enterprise, combining social bookmarking and tagging, social networking, expertise location and live profiles, along with social search. It aims to increase collective intelligence and make all employees more productive.

Users can bookmark intranet or Internet pages, websites, and documents as they work, grouping their bookmarks in 'topics' and applying keywords ('tags') that help them organize and identify information. Search results incorporate the information bookmarked and tagged by colleagues. Search results link users to the profiles of colleagues who have searched for similar information, thus identifying people with similar expertise and interests. Users can increase their social presence in the enterprise as their Connectbeam profiles show their interests, knowledge and skills, updated dynamically. Levels of sharing and privacy can be set. Groups can be developed and group topics created for a workgroup, project team, department, or invited list of participants to bookmark, tag and share information securely.

See http://www.connectbeam.com/solution.html for details.

1.4.3.4 Raytheon

Christine Connors has described how her company, Raytheon, successfully uses social tagging in a hybrid approach. At Raytheon, people submit suggestions of URLs together with recommended tags which are subsequently evaluated and approved by librarians. "We only rarely disapprove of a user-submitted term; overly general, vague or completely off-base terms are those that get deleted. We occasionally call to clarify a submission." She explained:

"In search, we insert Suggested Sites in a 'feature' box to the right of the regularly ranked results. We do not allow these suggestions to affect the ranking determined by the algorithm. Our surveys show that the sites submitted via this process repeatedly rank as the result deemed 'best' for the user's query. It is the single best thing we've done. I can't tell you how much we've spent on formal taxonomies, but suffice it to say that it's enough for me to wonder why I haven't gone into business for myself!

Why does it work? Chiefly because the sites submitted are specific to a group or discipline, and no matter how hard we try, having a degree in library science does not give you a degree in engineering (insert discipline here). We do not speak their vernacular. We do well enough to add value with controlled terms, but these folk tags have a life of their own. These tags are a fantastic resource - user warrant - for keeping the controlled vocabularies up-to-date. They provide us feedback we could get no other way. Given the ease with which people can tag things – and yes, we could argue about whether there should be some cognitive burden for quality's sake - we gain a unique insight via this process." (Lemieux, 2007)

1.4.3.5 Notorious

Notorious is a commercial application for enterprise portals. It provides a platform for users to tag and share their browser bookmarks with those of their team and company. It is possible to download five full licenses free.

Notorious is a system for organising employees' personal bookmarks, and sharing these with others in their company, team, country etc. When users see a page of interest to them and others, they "Note" the page, by listing keywords or "Tags" which are relevant. Users can filter and sort their "Tags" and bookmarks, in alphabetical order, most popular, and most recent order. It is therefore easy to see the most popular bookmarks or topics being bookmarked. Notorious is designed for use specifically within a company, meaning that the content is relevant to the organisation.

See http://www.notorious-software.com/ for details, including FAQs.

1.4.4 Some issues for social tagging in the enterprise

Gartner's and others' discussions of social tagging suggest that it will become an engine for innovation and that enterprises will need to master it. Issues that will need consideration, many mentioned above already, include:

Identity, transparency, privacy: will the system require all users to be identified, that is, to use their real names? Will lurking be allowed? Will there be some private groups?

In a collaborative and sharing environment, how will confidentiality and different levels of permission and access be managed?

Will there be a need for role-based or job-based collections?

Will private bookmarks and tags be included?

Is a critical mass of users needed before the system becomes effective?

Alerting and discovery mechanisms are important and should be included in a tagging system.

Although tagging systems are designed to be easy and intuitive to use, there will be employees who are not familiar with them – they may require some training or encouragement to use an unfamiliar system.

Systems should be designed for extensibility and remixing of content – that is, the system should integrate mesh with others in the organisation so that all corporate content and resources can be drawn together.

Management of the tagging will still be required, especially if one of the hybrid approaches discussed above is taken. For example:

> "immediately showing related existing tags while a user enters a new one can contribute to a more coherent way of adding tags to content, as typical spelling variations including morphological variants (tag, tags) can be caught early and users might be tempted to reuse more specific existing tags instead of adding the one they had in mind. Changing tags later on usually is possible, which also opens the way to tag merging and consolidation." (Hoppenbrouwers, 2007)

There is a challenge in gathering information about who is doing the tagging and allowing that to filer the results – using an organisational network analysis approach.

Finally, this is an emerging field and it is still, as many writers say, perhaps too early to be sure about the usefulness of tagging and how best to manage it. Although there are some systems available for organisations to use, and some models to follow, there are

not a huge number and they have not been tested for long. It will be vital to monitor developments and explore the huge amount of current information that is available and the many online discussions, in order to take advantage of the potential of this very powerful new development in information management.

1.5 References

Anderson, Paul, What is Web 2.0? Ideas, technologies and implications for education, JISC, 2007 (http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf)

Andrews, Whit, 2007, What Social Search Means to Your Enterprise, Gartner, ID Number: G00148270

Ann Arbor District Library (http://www.aadl.org/catalog) (Accessed April, 2007)

Beckett, Dave, 2006, "Semantics through the tag", XTech 2006: "Building Web 2.0" — 16-19 May 2006, Amsterdam, The Netherlands (http://xtech06.usefulinc.com/schedule/paper/135)

Berg, Oscar, 2007, "Insights about enterprise folksonomies", post to blog, The Content Economy, April 10, 2007, http://www.thecontenteconomy.com/2007/04/insights-about-enterprise-folksonomies.html

Blyberg, John, "AADL.org Goes Social" post to blyberg.net (http://www.blyberg.net/2007/01/21/aadlorg-goes-social/) (January, 2007)

Cook, Niall, 2007, Fifteen Uses of Corporate Bookmarking, June 3, 2007, post to Collective intelligence: a blog about social bookmarking, (http://blog.cogenz.com/index.php/2007/06/03/fifteen-uses-of-corporate-bookmarking/)

Cormier, Patrick, 2006, "Social tagging for the enterprise" post to Information management now, 25 February, 2006, http://imbok.blogspot.com/2006/02/social-tagging-for-enterprise.html

Cotton, Tom, "The really shiny part of my job - Proof of Concepts" post to Tom Cotton's 8109, March 2007, (http://blogs.educationau.edu.au/tcotton/)

Damianos, Laurie, Griffith, John and Cuomo, Donna, (2006), Onomi: Social Bookmarking on a Corporate Intranet, MITRE Coproration, (http://www.mitre.org/work/tech_papers/tech_papers_06/06_0352/06_0352.pdf)

The MITRE Corporation

Fienberg, Jonathan, Poundbang blog comment, 2007, (http://poundbang.in/2007/02/24/measuring-efficiency-of-tagging-with-entropy/#comments) (Accessed April, 2007)

Flint, Keely, 2007, "BUPA talk about Cogenz trial at Blogs and Social Media Conference 2.0", June 6, 2007, post to Collective intelligence: a blog about social bookmarking, (http://blog.cogenz.com/index.php/2007/06/06/bupa-talk-about-cogenz-trial-at-blogs-and-social-media-conference-20/)

Golder, Scott A. and Huberman, Bernard A, The structure of collaborative tagging systems, 2006 (http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf)

Guy, Marieke and Tonkin, Emma, "Folksonomies: tidying up tags", D-Lib magazine, January 2006, v. 12, no. 1 (http://www.dlib.org/dlib/january06/guy/01guy.html)

Hoppenbrouwers, Jeroen, 2007, *Social Tagging/indexing*, Workshop at the 31st ELAG conference, Barcelona, May 9-11, 2007, (http://elag2007.upf.edu/workshops/hoppenbrouwers.pdf

Kipp, Margaret E. I. and Campbell, D. Grant, 2006, "Patterns and inconsistencies in collaborative tagging systems: an examination of tagging practices", in Proceedings Annual General Meeting of the American Society for Information Science and Technology, Austin, Texas (US) (http://eprints.rclis.org/archive/00008315/01/KippCampbellASIST.pdf)

Lemieux, Stephanie, 2007, "The Value of Social Tagging in a Corporate Setting", post to blog, Taxonomy: community of practice, (http://taxocop.wikispaces.com/Social+tagging)

Lothian, Nick, 2007, Nick @ Education.au: Nick's Education.au Weblog (http://blogs.educationau.edu.au/nlothian/)

Maness, Jack, 2006, "Library 2.0 Theory: Web 2.0 and Its Implications for Libraries", Webology, Volume 3, Number 2, June, 2006 (http://www.webology.ir/2006/v3n2/a25.html)

Millen, David, Feinberg, Jonathan and Kerr, Bernard, 2005, "Social Bookmarking in the Enterprise", Social computing, vol. 3, no. 9, November 2005 (http://acmqueue.com/modules.php?name=Content&pa=printer_friendly&pid=344)

National Information Standards Organization, 2005, ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (http://www.niso.org/standards/standard_detail.cfm?std_id=814)

Porter, Joshua, “The Del.icio.us lesson”, post to Bokardo.com (http://bokardo.com/archives/the-delicious-lesson/) (May 2, 2006)

Rainie, Lee, 2007, 28% of Online Americans Have Used the Internet to Tag Content, Pew Internet and American Life Project (http://www.pewinternet.org/pdfs/PIP_Tagging.pdf)

Schwatrz, Ephraim, 2007, “ConnectBeam: Social bookmarks behind the firewall: Taking collaboration to the next level by tapping into your employees' knowledge”, Infoworld, (http://www.infoworld.com/article/07/05/30/MOES-connectbeam_1.html)

Shirky, Clay, 2005, “Ontology is Overrated: Categories, Links, and Tags”, in Clay Shirky's Writings About the Internet, (http://www.shirky.com/writings/ontology_overrated.html)

Spalding, Tim, 2007, “When tags work and when they don't: Amazon and LibraryThing” post to Librarything, (http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php) (February, 2007)

TNN -Tags / Networks / Narrative, blog (http://www.hum.dmu.ac.uk/blogs/tnn/) (Accessed April, 2007)

University of Pennsylvania, Penntags, (http://tags.library.upenn.edu/) (accessed April, 2007)

Vander Wal, Thomas, vanderwal.net (http://vanderwal.net/)

Vuorikari, Riina, 2007, Audit Report on MELT content - version 2 Part II: Folksonomies – state-of-the-art, requirements and use cases for the MELT social tagging tool, MELT Project (http://eunbrux09.eun.org/shared/data/melt/MELT_D5P2_Part2_final.pdf

Weinberger, David, 2006, comment to Joho the Blog, (http://www.hyperorg.com/blogger/mtarchive/penntags_when_card_catalogs_me.html)

WWW2006, Collaborative Web Tagging Workshop: papers http://www.ibiblio.org/www_tagging/2006/

Zeldman, Jeffrey, “Remove Forebrain and Serve: Tag Clouds II”, (http://www.zeldman.com/daily/0505a.shtml) (4 May, 2005)

Zurek, Bob, 2006, “It's a tag, tag, tag World! Enterprise Next” post to The Innovation Capitalist blog, May 4, 2006 (http://www-03.ibm.com/developerworks/blogs/page/BobZurek?entry=it_s_a_tag_tag)