Introduction: the continuing evolution of social tagging

Louise F. Spiteri and Diane Rasmussen Pennington

The genesis of an idea: Louise's perspective

I was introduced to the concept of social tagging when I was asked by Library and Archives Canada to speak about folksonomies at a metadata conference in Ottawa in 2005. Although I had heard the term, which was coined by Thomas Vander Wal (2007) in 2004, I did not know much about it, but I was certainly interested in the opportunity to learn more about this concept. As with most scholars in this field, my first in-depth exposure to the concept of social tagging was Adam Mathes' now classic article on folksonomies (2004). My area of expertise was in the areas of cataloguing, classification and thesaurus construction, all areas where language and descriptors are carefully chosen and controlled by professional information managers. I became intrigued at the possibilities that social tagging could provide to our carefully curated metadata records in libraries, which was the basis for my first article on the topic on social tagging (Spiteri, 2006) and which opened a new area of research interest that has continued to grow over the years.

For several years, I have studied the contributions of social tagging to library discovery systems (Spiteri, 2006; 2007; 2009); my interest in this particular topic was inspired by courses I teach in the areas of the organization of information, cataloguing and classification, as well as my involvement in social reading sites such as LibraryThing and Goodreads. I was struck by the dynamic and interactive nature of these reading sites: readers voluntarily edited metadata records for books, added social tags to describe content, created and shared reading lists, engaged in discussions with other readers, wrote reviews of items they had read and responded to reviews written by others. I was struck also by the difference between these dynamic sites and the static nature of the public library catalogues that I used, and used as exemplars for my students. These catalogues contained carefully constructed metadata records, using established and standardised metadata standards such as Anglo-American Cataloguing Rules and, more recently, Resource Description and Access, codified via the MARC (MAchine Readable Cataloging) framework and standard Library of Congress Subject Headings to describe the content and genre of a work. These practices are what I taught – and continue to teach – my students. I have never questioned the importance of these standardised records to describe the collection of a library; in fact, I continue to promote them actively to my students and colleagues. In comparison to the social reading sites, however, the library catalogues I used, both personally and professionally, struck me as somewhat

sterile: To use social media language, they *pushed* information, but did not allow for any interaction with the users of the catalogue. From this observation emerged my interest in using social tagging as a means to help library discovery systems become social spaces – a concept I explored closely with another colleague, Laurel Tarulli (Spiteri and Tarulli, 2012; Tarulli and Spiteri, 2012) – where users could input and interact with content, in much the same way as they can in social reading sites.

In Spiteri (2006) I suggested a research agenda for social tagging in the following areas: what is the tagging behaviour of people who use folksonomies? Why do people choose the tags they use; what motivates them to modify these tags; how often do they modify them? How are folksonomies used communally? How do folksonomies foster consensus in the use of tags? How does the community affect which tags are used and how? To my delight, these questions have been explored comprehensively over the years and have produced a rich corpus of knowledge in the field of social tagging. The internet is moving rapidly from the social web embodied in Web 2.0, to the semantic web (Web 3.0), where information resources are linked in such a way as to make them comprehensible to both machines and humans. The Web 3.0 environment provides us with the opportunity to explore the evolving role of social tagging (including hashtags, geotags and the like) in the semantic web. Hashtags, for example, have expanded beyond the scope of Twitter to include many other platforms, such as Facebook, Instagram, Pinterest, WordPress, Tumblr, YouTube and so forth; one single hashtag can thus link information resources from a variety of platforms, as will be discussed later in this chapter. This book is an exploration of the role that social tagging can play in helping to link people and information resources in a linked data environment. I am pleased to share this journey with my co-editor, Dr Diane Rasmussen Pennington, whose research focuses on information engagement and includes social media, digital consumer health information, digital photograph representation and online education pedagogies.

The genesis of an idea: Diane's perspective

I was thrilled when Dr Louise Spiteri invited me to co-edit this book with her. As we both teach information organization and cataloguing and we overlap in certain areas of research such as social tagging and linked data, it has been the perfect opportunity to assemble a group of authors together who are also thinking in similar terms towards this emerging area of research and practice in LIS and on the internet more generally.

I was first exposed to the concept of social tagging when I read my PhD co-advisor's book *Explorations in Indexing and Abstracting: Pointing, Virtue and Power* (O'Connor, 1996). At this time, social tagging as a term did not exist, but he explained his vision of a 'community memory interface' that would address the difficulty of searching for and describing the aboutness of pictures using words (O'Connor, 1996, 151). It would allow library patrons to contribute their own 'functional or adjectival descriptors' to a digital collection of pictures and the system would develop user profiles over time through these descriptors, both of which would lead to relevant searching and browsing results based on similar user profiles. His empirical research into this idea, in which he elicited descriptors about a set of images from a group of people, found that people struggled with assigning Library of Congress Subject Heading types of terms to the pictures. Instead, they enjoyed assigning non-restricted captions, such as 'That's what friends are for!', many of which moved beyond basic topicality, including emotions, metaphors and so on.

This research, as well as Brian O'Connor's follow-up related work (Greisdorf and O'Connor, 2002; O'Connor and Wyatt, 2004), inspired me to write my PhD dissertation on photojournalism professionals' preferences for description and representation within their online photograph archives, published under my former name, Diane Rasmussen Neal (Neal, 2006; Neal, 2008). Among other interesting results, I found that they needed and wanted control over the descriptors used to index and search for their images, but this control was not readily available in their systems. I wrote my dissertation around the time when social tagging was becoming a phenomenon, as Louise mentioned above (Mathes, 2004; Vander Wal, 2007), but was not yet widespread. Therefore, I used the term 'user-assigned descriptors' (UADs) in my dissertation research, rather than 'social tagging', to label photojournalism professionals' assignment of their own terms to their own pictures. I argued for the need to combine UADs and vocabulary control, such as automatic suggestions of UADs, to allow for the best possible mix of freedom and reliability.

My further research in this area has continued to explore the use of social tagging for non-textual documents – not only for images, but also music and video – especially the potential for describing, searching and browsing by users' emotions (Neal et al., 2009; Lee and Neal, 2010; Knautz et al., 2011; Neal, 2012, Pennington, 2016). Through these studies, I am finding that social tagging needs to be collective as well as individual; if a user wants to find a song that makes them feel happy on a Friday afternoon, some elements may be somewhat universal within a Western context (major key, fast tempo, uplifting lyrics) but may also be personal (individual tastes in and associations with, music differ among individuals).

I had first heard the terms linked data and semantic web discussed generally at the 2004 Association for Information Science & Technology conference, when Sir Tim Berners-Lee described his vision for the semantic web in his remarkable plenary session. My deeper explorations into it began when I was asked to present about linked data to Library and Archives Canada, when I was still working at Western University in London, Ontario, Canada, just like Louise's early introductions to social tagging. Developing a substantial understanding of linked data required considerable effort, but eventually I realised the similar goals of social tagging and linked data: while the practices and approaches are different, each one has the power to establish meaningfully unique connections between online documents, whether that document is an image, a social media post, a written text or anything else we might find online (Neal, 2010). When used together, their socio-technological power will be even stronger.

I have been working in collaboration with my students to explore the barriers and opportunities associated with implementing linked data in library and information settings. I introduced this in my 2016 *CILIP Update* article (Pennington, 2016) and I am actively writing and presenting in this area (Pennington and Cagnazzo, 2018). Some obstacles are institutional in nature, such as lack of staff and funding. Other issues involve a mismatch in technical implementations at different sites, which makes the semantic sharing of data envisioned difficult. I am, however, optimistic about the ability of social tagging within linked data, based on my own reflections as well as the work presented in this book. This is because the true power inherent in social tagging lies within the multitude of users and they therefore control the rich semantic connections made possible through the technology of linked data. This is ultimately Brian O'Connor's 'community memory interface becoming alive', although in a much different implementation than he could have imagined in the mid-1990s. I am privileged to have trained under his vision and to be a researcher in the area now during this rapid evolution.

Related works

Other excellent books have been published about social tagging. Gene Smith's 2007 publication, *Tagging: people-powered metadata for the social web*, introduces the concept of social tagging, how it could be used to improve the user experience and its role in information architecture and online communities. *Folksonomies, Indexing and Retrieval in Web 2.0* (Peters and Becker, 2009) discusses the applications, strengths and weaknesses of social tagging in collaborative information services and examines how established methods

of knowledge representation and models of information retrieval could be translated to this new format. *Recommender Systems for Social Tagging Systems* by Marinho et al. (2012) examines the three recommendation modes in social tagging systems: users, resources or tags and surveys recommender systems built to serve social tagging systems. *Recommender Systems and the Social Web: leveraging tagging data for recommender systems* (Gedikli, 2013) looks more broadly at recommender systems designed to provide personalised recommendations of products or services to users and how social tagging data can be used to improve these systems. *Folksonomies Social Tagging: a clear and comprehensive guide* (Blokdyk, 2017) provides practical suggestions for how to make the best and most efficient use of social tagging to organise business and project activities and processes.

<A>Themes covered in this book

The books discussed above have provided valuable insight into the role of social tagging in information discovery. The emphasis of social tagging has often tended to focus on discrete applications, such as social bookmarking sites, library discovery systems, blogs and so forth. This book extends the scope of social tagging to examine its contribution to the semantic web as a form of linked data. 'The Web has evolved from a global information space of linked documents to one where both documents and data are linked. Underpinning this evolution is a set of best practices for publishing and connecting structured data on the Web known as Linked Data' (Bizer, Heath and Berners-Lee, 2009, 1–2). Web documents often contain data that cannot be understood easily by machines. The semantic web is about facilitating access to web data by making it available in machine-readable formats that allow both people and machines to collect this data. Linked data is a way of creating links between data from different sources across different platforms. Berners-Lee (2006) proposed the following linked data principles for publishing web data to enable a single global data space:

1 Use uniform resource identifiers (URIs) as names for things.

2 Use URIs so that people can look up those names.

3 When someone looks up a URI, provide useful information, using standards such as Resource Description Framework (RDF) and SPARQL.

4 Include links to other URIs so that they can discover more things.

There are several examples of linked data repositories, such as DBpedia (https://en.wikipedia.org/wiki/DBpedia), which extracts structured content from the Wikipedia sites; the FOAF (Friend Of A Friend) ontology (www.foaf-project.org), which describes persons, their activities and their relations to other people and objects; and

GeoNames (www.geonames.org), which contains over 10,000,000 geographical names. These datasets contain discrete units of information such as names, locations, music albums, film titles and so forth. This book explores social tagging as a potential form of linked data; hashtags, for example, can already link content across a variety of platforms, such as Twitter, Facebook, Tumblr, WordPress, Instagram, YouTube and Pinterest. So, for example, a hashtag on a specific topic such as #PreventingType2 (preventing Type 2 diabetes) can link us to information from the following resources:

- Twitter (http://bit.ly/2jfWsJZ)
- YouTube playlist (http://bit.ly/2jk6Osz)
- Instagram (http://bit.ly/2rbfllC)
- Google Image results (http://bit.ly/2jfWRMv), which lead to several other results
- individual articles (e.g. http://bit.ly/2r5XEEf and http://bit.ly/2jfXGVB)
- a variety of Facebook pages (http://bit.ly/2jht9H9)
- Storify (http://bit.ly/2rbKl52).

In the next two chapters, Laura Cagnazzo and Sue Yeo Syn look more broadly at the role of social tagging in a linked data environment. These chapters examine the main features of the semantic web and linked data and on the relationship between the semantic web and Web 2.0. Cagnazzo examines a series of frameworks designed to enhance social tagging and to overcome some of its limitations through linked data. Syn explores efforts to format social tags as RDF triples and to define the semantic meanings and relationships of tags. Although these efforts are still limited, they successfully demonstrate that formatting tags with RDF-based models can allow tags to contribute to linked data in the semantic web environment.

Ryan Deschamps examines the connection between public policy and hashtags via three Canadian case studies. Deschamps shows the close connection between social tagging and Canada's political and social context and highlights the need for a more comprehensive framework for inclusion of online interactions to social change.

Louise Spiteri explores the potential contributions of hashtags to library discovery systems via an examination of three hashtags and their equivalent Library of Congress Subject Headings. Spiteri suggests that hashtags can serve as an important way to link library resource discovery systems to information resources in a variety of social media services, such as Twitter, Facebook, Instagram, Pinterest and YouTube.

Laurie Bonnici and Jinxuan Ma analyse how effectively hashtags have been used within two special information-based interest groups on Facebook and develop user instruction and intervention strategies for use in Facebook. Bonnici and Ma suggest that

linking works is an important way to consolidate information relevance and currency in the process of contextualising discoverability of information across social dimensions.

Max Dobson examines the use of tagging as a form of linked data in an online fan community. Fandoms have created functioning online communities and relationships around particular tags and use searchable tags and descriptive tags to make the content in fandom spaces more easily searchable.

Diane Rasmussen Pennington uses dementia as a case study to demonstrate how user-generated hashtags, or other forms of surrogate representation, could be applied in a linked data environment in order to improve access to care, resources, people and other needs. This could enable people to make more informed decisions about treatment and lifestyle options.

Sanjay Khanna explores how social tags can serve to link content within enterprises. Social tags can contribute to greater information discovery in the workplace and can be an important way to link employees through shared expertise and interests. Khanna examines also the role of social tagging in linking communities of practice within an enterprise.

Given the continued importance of recommender systems in the creation of tagged content, Kishor John discusses social tagging recommender systems. As information resources on the web continue to grow and particularly those that involve collaborative – or social – input, recommender systems can play an increasingly important role in helping people to tag resources by reducing the cognitive burden that this task may involve. Further, by suggesting tags based on the analysis of user input, recommender systems can help create more structured tagging vocabularies that reduce the drawbacks with which tags are often associated, such as polysemy, synonymy and homonymy. John examines the different types of recommender systems, highlighting their strengths and weaknesses:

- collaborative filtering recommender systems
- content-based recommender systems
- context-based recommender systems
- demographic recommender systems
- knowledge-based recommender systems
- hybrid recommender systems.

This book examines the themes above through the lens of academic researchers and practitioners. The authors reflect different geographic perspectives from the United Kingdom, India, Canada, and the United States. Intended readers include practicing library and information professionals who implement electronic access to collections, such as cataloguers and systems developers, as well as information architects and web developers. Of value to

researchers and practitioners is the potential to link social tags, hashtags, or geotags to the RDF data model. So, for example, social tags could form an ontology that could be used in semantic web applications, which would allow different web-based resources to be linked to a stable URI for the social tags. This use could have practical implications as well to practitioners who wish to link resources from different platforms via social tags or hashtags. The chapters could lead to an increased understanding of user behaviour about how social tags, hashtags, or geotags could assist in the design of better and more intuitive user interfaces. Instructors and students in different academic disciplines, such as library and information science, computer science, informatics, and information management, could apply the themes of this book to courses, particularly in the areas of metadata, taxonomies, ontologies, information architecture, records and information management and bibliographic description.

Since its genesis as a concept in 2004, social tagging continues to provide a wealth and variety of exciting research avenues. With the continued growth of the Web 3.0 semantic web, social tagging can provide an increasingly important way to categorise and store information resources to make them understandable to both humans and computers.

References

Berners-Lee, T. (2006) *Linked Data – Design Issues*, www.w3.org/DesignIssues/LinkedData.html.

Bizer, C., Heath, T. and Berners-Lee, T. (2009) Linked Data – The Story So Far, *International Journal on Semantic Web and Information Systems*, **5**, 1–22.

Blokdyk, G. (2017) *Folksonomies Social Tagging: a clear and comprehensive guide*, CreateSpace Independent Publishing Platform.

Gedikli, F. (2013) *Recommender Systems and the Social Web: leveraging tagging data for recommender systems*, Springer.

Greisdorf, H. and O'Connor, B. (2002) Modelling What Users See When They Look at Images: a cognitive viewpoint, *Journal of Documentation*, **58**, 6–29.

Knautz, K., Neal, D. R., Schmidt, S., Siebenlist, T. and Stock, W. G. (2011) Finding Emotional-laden Resources on the World Wide Web, *Information*, **2**, 217–46.

Lee, H-J. and Neal, D. (2010) A New Model for Semantic Photograph Description, *Journal of Information Science*, **36**, 547–65.

Marinho, L. B., Hotho, A., Jäschke, R., Nanopoulos, A., Rendle, S., Schmidt-Thieme, L., Stumme, G. and Symeonidis, P. (2012) *Recommender Systems for Social Tagging Systems*, Springer.

Mathes, A. (2004) *Folksonomies – Cooperative Classification and Communication Through Shared Metadata*, http://adammathes.com/academic/computer-mediated-communication/folksonomies.html.

Neal, D. R. (2006) *News Photography Image Retrieval Practices: locus of control in two contexts*, University of North Texas.

Neal, D. R. (2008) News Photographers, Librarians, Tags and Controlled Vocabularies: balancing the forces, *Journal of Library Metadata*, **8**, 199–219.

Neal, D. R., Campbell, A., Neal, J., Little, C., Stroud-Mathews, A., Hill, S. and Bouknight-Lyons, C. (2009) 'Musical Facets, Tags and Emotion: Can We Agree?' In *Proceedings of the 2009 iConference, Chapel Hill, NC*.

Neal, D. R. (2010) Emotion-Based Tags in Photographic Documents: the interplay of text, image and social influence, *Canadian Journal of Information and Library Science*, **34**, 329–53.

Neal, D. R. (ed.) (2012) *Indexing and Retrieval of Non-Text Information*, De Gruyter Saur.

O'Connor, B. C. (1996) *Explorations in Indexing and Abstracting: Pointing, Virtue and Power*, Libraries Unlimited.

O'Connor, B. C. and Wyatt, R. B. (2004) *Photo Provocations: thinking in, with and about photographs*, Scarecrow Press.

Pennington, D. R. (2016, July/August). Demystifying linked data: are you ready for what's next?, *CILIP Update*, 34-6.

Pennington, D. R. (2016) 'The most passionate cover I've seen;': emotional information in fan-created U2 music videos, *Journal of Documentation*, **72**, 569–90.

Pennington, D. R. and Cagnazzo, L. (2018) Relationship Status: libraries and linked data in Europe. In *Proceedings of the 15th International ISKO Conference, Porto, Portugal*, ISKO.

Peters, I. and Becker, P. (2009) *Folksonomies, Indexing and Retrieval in Web 2.0*, De Gruyter Saur.

Smith, G. (2007) *Tagging: people-powered metadata for the social web*, New Riders.

Spiteri, L. F. (2006) The Use of Folksonomies in Public Library Catalogues, *The Serials Librarian*, **51**, 75–89.

Spiteri, L. F. (2007) The Structure and Form of Folksonomy Tags: the road to the public library catalog, *Information Technology and Libraries*, **27**, 13–25.

Spiteri, L. F. (2009) The Impact of Social Cataloguing Sites on the Construction of Bibliographic Records in the Public Library Catalogue, *Cataloging & Classification Quarterly*, **47**, 52–73.

Spiteri, L. F. and Tarulli, L. (2012) Social Discovery Systems in Public Libraries: if we build them, will they come?, *Library Trends*, **61**, 132–47.

Tarulli, L. and Spiteri, L. F. (2012) 'Library Catalogues of the Future: a social space and collaborative tool?, *Library Trends*, **61**, 107–181

Vander Wal, T. (2007) *Folksonomy Coinage and Definition*, www.vanderwal.net/folksonomy.html.