Tagging for Use: an analysis of use-centred resource description

Luanne Freund
School of Library, Archival and Information Studies
University of British Columbia
Vancouver, Canada
604 8220825
luanne.freund@ubc.ca

Richard Butterworth
Bridgeman Art Library
17—19 Garway Road
London, UK
44 (2) 20 7727 4065
richard.butterworth@bridgemanart.co.uk

ABSTRACT
This paper discusses use-centred resource description, a practice whereby an information object is assigned metadata that describes what it can be used for, as opposed to what it is. We look at precedents for this practice in the literature and present three cases of use-centred description in diverse information environments. Through analysis and comparison of the cases, we develop a preliminary framework for use-centred resource description.

Categories and Subject Descriptors
H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing

General Terms
Human Factors

Keywords and Conference Themes
Use-centred description, indexing, metadata, tagging; Other approaches to eliciting, identifying and expressing/capturing contextual information; Task-based interactive IR and seeking behaviour.

1. INTRODUCTION
The premise of use-centred description is simple: that there is value in representing an information object in terms of how it may be used in addition to the more traditional approach of describing what it is about. Use-centred description represents a shift away from thinking about information objects as containers filled with ideas, to thinking about them as resources or tools with which people interact, and which support human endeavour. The concept of “use” can be defined as, “the act or practice of employing or applying something”. It includes both the senses of engagement or interaction with a resource, and application of a resource to a particular purpose, task or context.

The notion of task- or use-centred description as a supplement to subject-based description is appealing because it offers an alternative point of access to information resources that is aligned with the information seekers’ tasks and intended uses of information, which are often more clearly defined than the specific subject-matter of the information they seek [1]. Describing information in this way is a commonplace practice among people when organizing and sharing information. Research in personal information management has shown that people frequently use folders with task designations to organize files on their personal computers [2]. A similar approach is often used when sharing information. For example, reviews for the book Information Architecture on the World Wide Web posted on the Amazon website included the following use-centred recommendations:

• “I found this to be a superb reference for self-study, one that breaks down complex issues…”
• “This is a great book to introduce business people to information architecture, for architects to reinforce their skills…”

Despite the fact that this use-centred approach to information is embedded in everyday information practices, formal resource description frameworks, such as Dublin Core, MARC, RDF, OAIS and FRBR make no explicit accommodation for this type of metadata. A recent review of digital library research notes that use-centred metadata, “does not yet seem to have a literature of its own” [3, p. 233]. Similar calls have been raised for the need to change resource description practices to include ways to describe: “what can be done with the document: buy, read, solve, calculate, download, play games, chat, sell, gamble, search, listen, watch, etc” [4, 5]. This obvious gap between informal and formal resource description practices highlights the need to determine how use-centred description may be used to support information seeking and retrieval (IS&R) in digital environments.

This paper presents an analysis of use-centred description based on a series of cases. The cases selected represent diverse examples of real-world use-centred description of various types. Our intent is to introduce this concept and show that the information environments in which use-centred description is applicable are heterogeneous and can be characterized along a number of dimensions.

1 Merriam-Webster Online, Retrieved April 30, 2008
The preliminary framework proposed in this paper lays out and discusses the different dimensions. Our longer term aims are both practical -- to develop this set of criteria into a collection of guidelines which may be considered when deciding which form of use-centred tagging may bring most benefit to a particular domain; and theoretical -- to develop a set of evidenced criteria by which the value of use-centred tagging may be discussed within the context of IS&R.

2. BACKGROUND

Theoretical approaches to resource description tend to be object-centric, and have traditionally focused on the intellectual content and the physical description of information objects [6]. Use-centred description, which is extrinsic to the information object, is therefore lacking a cohesive theoretical framework, although it is related to user- and domain -centred approaches in indexing [7,8], functional approaches to archival description [9] and emerging task-based approaches to IS&R [10,11].

Fidel [12] divides indexing into document-oriented and user-oriented approaches. She discusses user-oriented approaches mostly in terms of linking documents to user requests (cf. [6, chapter 13]), but argues that it is difficult to implement a priori request-oriented indexing because it is difficult to predict which documents match which requests until the requests have been made, particularly without a thorough understanding of user searching behaviour. Domain-based indexing goes beyond user-centred indexing, by considering the document within its context of use, i.e. the domain. This requires that the indexer have, “an understanding of the users and the domains goals, purposes, and activities to make an assertion about a document’s subject matter” [7, p. 600]. Examples of indexing frameworks developed using this approach include the Bookhouse project [13] and Nielsen’s [14] work-task based thesaurus. However, these and most other examples of domain-centred indexing employ subject rather than use descriptors, based on the idea that an understanding of the potential uses of a document will provide a useful lens “to extend the subject matter” [7].

Although some preliminary work has been done to assign task descriptors to documents to facilitate information retrieval [15], this pragmatic approach has not been widely adopted. Hjorland [16, p. 181] rejected the idea of assigning task or use descriptors as “short-sighted” and “superficial”; unlikely to provide any degree of stable long-term access due to the difficulty of predicting all possible future use scenarios.

This focus on the information object without reference to its end use is strongly rooted in traditions of archival and bibliographic description and is formalized in frameworks such as FRBR [17], MARC, and ISAD(G) [18]. It is interesting to note that in the FRBR report and in [19], which lay out the conceptual underpinnings of cataloguing standards, the importance of meeting users’ needs is emphasized, but implications of this with respect to use metadata are not expressed in the framework. The FRBR user model is limited to a set of four tasks: find, identify, select and obtain -- none of which relate to the actual use of the information object, but only to its acquisition. Furthermore, who exactly the proposed users are is never made explicit. Fidel [12], for example, makes the distinction between professional searchers and end users: in her work there is clearly an expectation that descriptive frameworks are used by information professionals (searchers) in order to find information for end users. It seems that this dated distinction between information professionals and end-users may be built into the cataloguing infrastructures to the extent that they exclude consideration of the ways in which ordinary people make use of information in their work and everyday lives.

In contrast, archival descriptive traditions have placed greater emphasis on the context and function of records and collections, in keeping with the nature of archival materials, which are the product of organic activity rooted in the context of creation [9]. Thus archival descriptive frameworks, such as the Encoded Archival Description (EAD), do accommodate functional metadata; however, this type of metadata is typically based on a functional analysis of the organization that created the materials, rather than an analysis of how those materials may be used in the future.

In practice, resource description is no longer an activity reserved for information intermediaries. In his study of folksonomies, Mathes [20] noted that information professionals, content creators and users of information all play a role in description in the Web context. Kipp [21] compared the indexing practices of these three groups (authors, intermediaries and users) using data from the CiteULike social tagging application and found that there were important differences, notably a much broader use of terminology by users. Given the more active role of authors and users in resource description in online digital environments, it seems possible that some of the aforementioned obstacles to the use-centred description may be diminishing, paving the way for it to play a more prominent role in information organization and retrieval.

We now turn to an examination of use-centred description in practice.

3. USE-CENTRED DESCRIPTION CASES

3.1 Case I: Archival Setting

Our first case describes the design and implementation of use-centred descriptions of print materials in an archival environment. The descriptions were created by information intermediaries to facilitate access and selection of archival materials by a distinct user group.

The Accessing our Archival and Manuscript Heritage project (AAMH) was undertaken at Senate House Library, University of London to establish a set of resources to facilitate broader access to the (primarily) archival resources held in the university [22]. Initial decisions were made that local and family historians (who were collectively referred to as ‘personal history’ researchers) would be targeted by the project, as there was evidence that the libraries held resources that would be very valuable to these researchers, if only they were aware of them. As the project progressed it became clear that within the time and financial constraints of the project the most effective approach would be to develop a body of secondary information about the archival resources available, rather than digitizing the primary resources themselves.

Through interviews with stakeholders, the project gathered information about the archival resources that would be of value to personal history researchers and developed a systematic schema for describing those resources. This schema was based on the initial observation that in the interviews, archivists described their
collections in a very different manner to the descriptions that were available in their finding aids and online catalogues. In essence the archivists verbally described what could be or had been done with their collections, whereas the catalogues and finding aids described what was in them. Also it was interesting to note that the archivists would spend time explaining what sorts of research questions a given collection was not useful for.

After interviewing several personal history researchers it was established that they had surprisingly well-defined and established research questions, but lacked the skills in identifying information resources to help them address those questions. Furthermore, because they were not typically academics with ready access to research libraries they did not have access to the intermediation services that librarians provide to help them link up their research questions with promising archives and resources. Given these observations and analyses the main focus of the project became the development and publication of ”use-centred” descriptions, which were in effect an encoding of archivists’ and other information experts’ knowledge of what a collection could be used for by a personal history researcher. An example of use-centred description is shown in Figure 1.

The use-centred description schema consisted of five main sections:

**Detailed usage description** — this is the central section of the description. It gives a brief overview of the contents of the collection and then describes the research questions that the collection could be used to address.

**How to tell if the collection is useful** — many of the collections included various archival finding aids so a brief description of them and how to use them is given here. In most cases the best way to find out whether a collection does indeed hold useful information is to email or phone a query to the archivist. If this is the case then the minimum information the archivist would need to answer a query was described here.

**Examples of the holdings** — if possible a few annotated scans of holdings are included in this section showing typical items in the collection. When holding institutions have scans of documents online, links to these surrogates are given.

**Access arrangements** — this section describes the holding institutions’ rules of access, with a particular emphasis on access for researchers who are not academics or students.

These five sections together are meant to provide enough information for a user to make a decision about whether they should select a resource and how to obtain access to that resource.

After interviewing several personal history researchers it was established that they had surprisingly well-defined and established research questions, but lacked the skills in identifying information resources to help them address those questions. Furthermore, because they were not typically academics with ready access to research libraries they did not have access to the intermediation services that librarians provide to help them link up their research questions with promising archives and resources. Given these observations and analyses the main focus of the project became the development and publication of ”use-centred” descriptions, which were in effect an encoding of archivists’ and other information experts’ knowledge of what a collection could be used for by a personal history researcher. An example of use-centred description is shown in Figure 1.

The use-centred description schema consisted of five main sections:

**Detailed usage description** — this is the central section of the description. It gives a brief overview of the contents of the collection and then describes the research questions that the collection could be used to address.

**How to tell if the collection is useful** — many of the collections included various archival finding aids so a brief description of them and how to use them is given here. In most cases the best way to find out whether a collection does indeed hold useful information is to email or phone a query to the archivist. If this is the case then the minimum information the archivist would need to answer a query was described here.

**Examples of the holdings** — if possible a few annotated scans of holdings are included in this section showing typical items in the collection. When holding institutions have scans of documents online, links to these surrogates are given.

**Access arrangements** — this section describes the holding institutions’ rules of access, with a particular emphasis on access for researchers who are not academics or students.

These five sections together are meant to provide enough information for a user to make a decision about whether they should select a resource and how to obtain access to that resource. There was also a collection of abstracting sections, such as an overview, dates, geographic coverage and so on, which were designed to allow users to more easily browse or search the published collection of use-centred descriptions. Descriptions were written discursively, as we believed that using terse, controlled vocabularies could be off-putting to the target user population.

The International Law Association (ILA) was established in 1873 to promote the harmonisation of laws that crossed international boundaries. Its membership consisted mostly of those in legal profession, but also politicians and economists. This collection holds the complete membership details until 1938.

Name, occupation and address are recorded. Unfortunately the address given for each member may be only specific to their town; this is particularly true for the earlier membership lists. In some cases other administrative details are recorded such as when fees were due, whether the individuals were honorary members, or in some cases their decease date.

Therefore this collection is only likely to be of use in breaking a research dead end: if you have very little information about a relative you know to have been in the legal profession then this collection may point you to an address or decease date which will allow you to move your research on. That said, the 1920s and 30s membership lists include many members from mainland Europe with accurate addresses, and with the destruction of many of these sorts of records during the Second World War, this collection may contain one of the few remaining records of your relatives.

Figure 1: The detailed usage description for the International Law Association Membership Records

A set of use-centred descriptions of archival collections was developed after liaison with the archivists responsible for the collections, individual personal history researchers, personal history societies and a professional genealogist. Collectively this analysis provided not only the descriptions themselves, but a thorough understanding of the search behaviour and information needs of personal history researchers. This understanding was most usefully expressed as a checklist of things to look for in collections, which meant that they would be particularly valuable to personal history researchers. For example, any resource that contains names and addresses from mainland Europe in the 1930s was seen as very valuable as many such records were lost in World War II. (See figure 1 as an example.)

These descriptions are neither comprehensive (about sixty collections were described, and there are several thousand collections held) nor in any way objective. They are intended to be an encoding of the archivists’ collective knowledge of what a collection can be used for to serve as a useful finding aid aimed at a class of users who may not have the skills to make full use of the more traditional content-based finding aids that are available.

### 3.2 Case II: Organizational Setting

The second case illustrates use-centred description in a workplace information environment. The descriptions are in the form of metadata assigned to digital documents stored in a shared repository. The authors and users of the documents are members of a large team of software engineers working for a particular product line in a large hi-tech firm.

As part of a larger research project, a study of metatagging practices with a shared document repository was undertaken [23]. The repository was heavily used by the team members as a resource to support their work activities. The repository is password controlled, available only to members of the team, and contains several thousand documents written by and/or recommended by the team members for use by their colleagues. The repository plays an important role in the knowledge management strategy of the company, which places heavy emphasis on the re-use of documents written by employees.

According to this strategy, the value of information assets can only be realized when they are “locatable, assessable and usable”
This strong emphasis on utility and reuse of information is reflected in the indexing framework for the repository, which includes about thirty metatags, including one called purpose, which represents how the document is intended to be used. Metadata is assigned by the team members themselves, as part of the process of submitting documents authored or recommended by them, to the repository.

To better understand how this community interpreted and used the purpose metatag we conducted a content analysis of the assigned values. Our sample consisted of 3414 tagged documents, which is the set of all documents submitted to the repository for a period of about two years (June 2002 to July 2004). The system allows multiple purpose tags to be assigned to each document, so the data set contained 4844 purpose values. These were assigned by a total of 402 unique users.

The number of purpose tags assigned to each document ranged from 0 to 13, with an average of 1.4 tags per document. Very few documents were left untagged (<1%) and the vast majority (80%) had a single purpose tag. Although a list of preferred purpose tags was available to submitters, data entry in this field was not controlled, so users were able to add new tags. The result is that purpose was expressed using 126 different terms.

Within this set of 126 purpose tags, we identified five subcategories, which contain different types of use-centred information to the user. The categories and summary data on the tags are shown in Figure 2, and described in more detail below.

**Information Uses:** The first category we identified represents the broad and generic information uses associated with documents. The most common information uses in this set of tags are related to education, documentation, and demonstration. This category is the most prevalent, making up 44% of all the purpose tags, and is fairly cohesive, with only 18 different values.

**Work Tasks:** The second most prevalent category of purpose tags (36%) represents the work tasks which can be supported by documents. The range of tasks in this category is very broad, including technical software engineering tasks such as deployment, administration and migration, as well as consulting tasks such as mentoring and project review.

**Information Genres:** A relatively small number of purpose tags (11%) describe the genre of documents. This is surprising, given that a separate metadata field exists to record genre. However, it seems that some genres, such as use case, course listing, presentation, etc. are strongly associated with particular uses and purposes.

**Use Situations:** A small sub-category of purpose tags (7%) represent situations or events for which documents were created or in which they are likely to be of use. Tags describe different types of consulting projects, classes, and specific workshops.

**Other (Topical):** The final catch-all category includes a small number of tags that describe topics covered in the documents, rather than task or use contexts.

---

<table>
<thead>
<tr>
<th>Subcategory</th>
<th># unique values</th>
<th>% of tags</th>
<th>Example values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Uses</td>
<td>18</td>
<td>44%</td>
<td>educate, demonstrate, document, compare</td>
</tr>
<tr>
<td>Work Tasks</td>
<td>44</td>
<td>36%</td>
<td>deployment, marketing, costing, architecture</td>
</tr>
<tr>
<td>Information Genres</td>
<td>33</td>
<td>11%</td>
<td>course listing, contracts, FAQ support, use cases</td>
</tr>
<tr>
<td>Use Situations</td>
<td>10</td>
<td>7%</td>
<td>class, workshop</td>
</tr>
<tr>
<td>Other: Topical</td>
<td>16</td>
<td>1%</td>
<td>Eclipse v2.0 Plugin, Portlet, zSeries</td>
</tr>
</tbody>
</table>

Figure 2: Breakdown of Purpose Metadata Tags by Subcategory

The first four subcategories of tags represent distinct interpretations of purpose, although they differ in degree of specificity and context-dependence. Information uses such as education and documentation are likely to apply in many different information environments; whereas work tasks are specific to the fields of software engineering and consulting. Use situations and events in which documents can be used are even more specific and are likely to be meaningful only within this team. This last type of purpose tag was not heavily used, perhaps because it is too specific to promote re-use, or because few documents are associated with such specific uses. Finally, the practice of using genre as a stand-in for purpose suggests that document genres are strongly associated with particular information intents, work tasks and/or use situations, and thus genre tagging acts as one type of use-tagging.

### 3.3 Case III: World Wide Web Setting

The third case considers use-centred description in the framework of Delicious (http://del.icio.us), a social bookmarking application for Web documents. In this environment, document URLs are described using free-form tags and annotations, which are assigned by document users for personal and shared use.

Delicious is one of the best known collaborative bookmarking applications on the Web. It is a free tool, which allows registered users to save bookmarked URLs in their accounts, to tag them with one or more tags, and to add a brief annotation. Users of the system can access and search through their own bookmarks as well the bookmarks of other users. Although some discussions of tagging equate it with subject indexing [25], the tagging instructions on Delicious support a broader interpretation: “Tags are one-word descriptors that you can assign to your bookmarks on del.icio.us to help you organize and remember them. … You can use tags describing an article or website's subject, location, name, category, people, places, ideas -- pretty much anything you can think of.”

Golder and Huberman [26] analyzed two sets of Delicious data (by user and by popular tags), and identified a number of functions that tags perform: identifying what documents are

---

about, what types of object they are (article, blog), and who owns the document content, refining other tags, identifying qualities or characteristics (“stupid”), relating the document to the tagger (“mystuff”), and, finally, as a means to organize information for a task (“jobsearch”, “research”). With respect to the last category Kipp and Campbell [27] note that when users tag documents in relation to their own specific interests or tasks, it adds a new temporal and situational dimension to description that is fundamentally distinct from traditional classification.

A cursory analysis of the most popular Delicious tags shows that use-centred tagging, while not easy to identify from the tags themselves, is a common practice, perhaps as common as subject-based tagging. The 20 most popular tags posted on the site on are: design, blog, webdesign, web2.0, tools, software, programming, music, art, video, reference, web, css, tutorial, photography, linux, inspiration, education, free, and howto. Among this group, reference, how to, inspiration, and tools clearly express how a page can be used rather than what it is about. Among the 742,888 pages tagged with reference are a broad range of general reference resources, such as the Internet Public Library and WorldCat, as well as specific sites, such as How to be Comfortable on a Long Plane Ride. The set of pages tagged with inspiration (293,355) includes examples of well-designed webpages as well as other inspirational resources, such as the TED: Ideas Worth Spreading website.

What is more interesting, is that even tags that seem to be subject-based: webdesign, programming, photography, education etc., are frequently used in a functional manner. For example, some of the pages tagged with webdesign are about webdesign, but many of them are pages that would be used to do web design, such as design templates, CSS tutorials, and tools for Web analytics. Another indication of the functional nature of many of these tags, are those which represent document genres, such as tutorial. Just as genres were used in the previous case to represent the purpose or intended use of documents, they show up in Delicious, as tags that are more indicative of the function than of the semantics of information objects.

4. DISCUSSION

The three brief cases presented above vary along a number of dimensions, but they all involve some type of use-centred description of information resources as a means of facilitating access to and use of those resources.

The cases share a number of common features. They are all user-centred as well as use-centred, as they are either designed specifically for a defined user group or users are directly involved in the resource description. They are all situated in heterogeneous information environments, made up of information objects of various types and formats, created for various purposes, and used in a range of different ways, unlike, for example, databases consisting solely of scholarly articles or of books. A third similarity is that in all cases, use-centred descriptions are combined with other types of description, including subject-based. Although subject and use tag types are not differentiated in Delicious, most pages are assigned multiple tags, so use-centred tags are seldom used independently.

These three common features: user-centredness, heterogeneity and the combination with subject-based description seem to be logical pre-requisites for getting benefit from use-centred description. It is hard to imagine the value or practicality of this practice without these conditions being met. However, two additional common features, the lack of controlled vocabularies and use of manual rather than automatic methods of description seem less likely to be intrinsic features of use-centred description, and more a function of the relative novelty of this approach. The tools and methods for more systematic and efficient use-centred description have not yet been developed.

The cases also vary in a number of interesting ways. The concept of use is interpreted in a variety of ways across the cases. In the archival case, use descriptions focused on the types of research questions a collection could be used to address, as well as how to access and use a collection. In the organizational case, metadata was used to indicate what documents were intended to do (i.e. educate, document), the work tasks they could support, the situations in which they were created or meant to be used, and the genre of documents. In Delicious, tags are unstructured and uncontrolled, but based on the most popular tags, use-centred description seems to indicate work tasks and activities that bookmarked documents can support, such as programming, and types of information uses to which they are suited, such as reference or education.

The roles played by participants are another source of variation across the cases. In the archival setting, descriptions were created by intermediaries, based on their understanding of the collections and the users. In the organizational setting, descriptions were assigned by the authors and contributors of documents for the users who were their colleagues and themselves. Delicious tags are assigned by users for themselves and other public users. So, it seems that under different conditions, use description can be performed by any or all of the actors involved.

However, there is a distinction between the author’s and the user’s perspectives on use. The author may intend that a document be used in a particular manner or situation. These intended uses are tied to the context in which a document is created, and may be expressed in the information object by the author either directly through metadata or keywords, or indirectly, through choices of style or genre. However, the actual uses of documents may have been neither intended nor anticipated by the author, and can only be identified from within the context of use, either by studying users, or through the users’ own experience. The contexts of document creation and use may be one and the same, or may be separated by time, place, domain or community.

It is certainly true of the archival materials in the first case, that the actual uses are hardly ever the same as the uses for which the materials were created or collected. This makes information access particularly challenging, because it requires that users make a creative leap to guess how resources created in one context can be used in another. The value of use descriptions provided by intermediaries who have some understanding both of the contexts of creation and use, is that they are able to bridge that gap and support resource discovery.

In the case of the workplace community of practice, the opposite situation exists. Here, the contexts of creation and use almost complete overlap, as the authors and the users are members of the same community. As a result, authors are so familiar with the
ways in which the information is likely to be used, that they are able to provide highly relevant use metadata to help colleagues filter out information poorly suited to their needs. So, the particular value of use description in this environment is in supporting filtering and resource selection.

The social bookmarking case is different yet again, as it is situated almost entirely in the context of use, with very few ties to the context of creation or the intended uses of documents. In this context, documents are tagged by users according to how they have actually used or plan to use documents. On a personal level, these use-tags support the management and re-finding of documents, as familiar tasks and situations seem to provide a stronger framework for memory than do topics [2]. On a social level, use-centred tags are excellent tools for serendipitous resource discovery because tags such as inspiration, tripplanning or cooking are often assigned to pages that do not necessarily contain those keywords, but which other people used for those purposes.

Thus, a preliminary framework for use-centred description drawn from these case studies includes the following dimensions:

- **Actors and roles**: authors, compilers, intermediaries, users, taggers/indexers;
- **Orientation of descriptions**: personal, shared within community or public;
- **Context of description**: context of document creation or context of document use
- **Uses**: intended or actual; questions, problems, information uses, tasks, activities, situations
- **Vocabulary**: controlled or uncontrolled
- **Method**: manual or automatic
- **Purpose**: resource organization, re-finding, discovery, selection

The identification of these dimensions serves starting point for further study of use-centred description. Much work remains to be done in this area, both to test and validate the framework, and to determine whether or not use-centred description makes a substantial contribution to information access and use. The challenge in evaluating this approach is that information use is domain-specific, situated, and may take place outside the context of the information system. Therefore, this will require an extension of traditional evaluation frameworks in the manner suggested by Järvelin and Ingwersen [10, p 321-333].

5. CONCLUSIONS

The notion of use-centred description is not new, but has had little impact on the design of resource description frameworks or information systems. Based on the cases presented here, use-centered description does seem to offer the possibility of supplementing and extending subject-based description to support resource discovery, selection and organization in user-centered systems operating in heterogeneous information environments. Given that these types of systems and environments increasingly are becoming the norm, we would argue that we can no longer afford to ignore this approach. The preliminary framework for use-centred description outlined here should provide the basis for further work in this area.

6. REFERENCES


