January 2016

Familiar Categories and Documentary Forms: Readers’ Perspectives

Amanda Cossham
Monash University, cossham00@xtra.co.nz

Please take a moment to share how this work helps you through this survey. Your feedback will be important as we plan further development of our repository.
Follow this and additional works at: https://ideaexchange.uakron.edu/docam

Part of the Library and Information Science Commons

Recommended Citation
DOI: https://doi.org/10.35492/docam/2/1/12
Available at: https://ideaexchange.uakron.edu/docam/vol2/iss1/12

This Conference Proceeding is brought to you for free and open access by University of Akron Press Managed at IdeaExchange@UAkron, the institutional repository of The University of Akron in Akron, Ohio, USA. It has been accepted for inclusion in Proceedings from the Document Academy by an authorized administrator of IdeaExchange@UAkron. For more information, please contact mjon@uakron.edu, uapress@uakron.edu.
Introduction
This paper presents an evaluation of the ways in which three different groups of readers (recordkeepers, teachers and secondary school students) categorise documents. It is part of an interpretive PhD that aims to explore and to understand the meaning that people ascribe to a socially constructed phenomenon: the bibliographic universe. This phenomenon is loosely defined as information entities that might be found in library collections. The research is comparing designed models (such as the functional requirements for bibliographic records (FRBR) model (IFLA, 1998/2009) or Murray and Tillett’s (2011) cultural heritage resource description), with readers’ mental models. The aim of the PhD research is to identify the ways in which library users (or readers) understand the bibliographic universe in a continuously changing information landscape of new media and communications, and to identify what their understanding might mean for the development of library catalogues. Drawing on Rosch’s (1978) work on the principles of categorisation, this paper explores the extent to which the changing information landscape is reflected in the categories developed by these participants.

Amy Devitt (2004) writes about “a shift from genre as defined by literary critics or rhetoricians to genre as defined by its users” (p. 3). This paper considers what kinds of understandings of genre, form, and type – and any other approach to categorising – are presented by a particular set of participants. Understanding of the forms and genres of documents is not the same for readers as it is for information professionals. Evaluating the way readers conceive of, model, or categorise documents and comparing their conceptions to pre-exiting knowledge organisation systems (such as library classification and subject headings), shows us what these library catalogue users think and expect. It can also indicate how well such systems match expectations. Library catalogues have long been considered hard to use (e.g., Borgman, 1986, 1996; Fast & Campbell, 2005; Novotny, 2004). There are questions about the continuing value of library catalogues when search engines and full text searching can do so much more. Hjørland (2012) examines the extent to which knowledge organisation such as classification is even necessary ‘after Google’. Various studies have demonstrated how the Internet and more specifically the World Wide Web have influenced mental models, expectations, behaviour and strategies of those using online library catalogues (e.g., Merčun & Žumer, 2008; Novotny, 2004; X. Zhang & Chignell, 2001; Y. Zhang, 2008a, 200b).

Activity and Participants
The chosen method to elicit participant categories and relationships between categories was an open card sort of 36 cards representing The Hobbit by J.R.R. Tolkien and associated documents in various versions, instantiations, and forms. These include translations, graphic novels, e-books, the sequel (The Lord of the Rings), Tolkien’s biography, The Hobbit manuscript, poems
from The Hobbit, works about Middle Earth, movie DVDs, a fan website, and a Wikipedia page. These are collectively termed documents rather than works in this paper, as ‘work’ has a particular meaning in the FRBR model and there are several different expressions of the work The Hobbit on the cards. Document is therefore the preferred term, according to the terminology framed by David Levy:

In this division of responsibilities [document, text, work], a document is a physical artifact bearing meaning- or information-bearing symbols; a work is the essential meaning or idea that is being communicated; and a text is that which mediates between document and work: a sequence of words which, as the expression of a work, can be realized or embodied in one or more documents. These notions thus typically form an abstraction hierarchy: from the fully concrete document to the abstract text to the even more abstract work. (Levy, 2003, section 3)

An ‘open’ sort indicates that category names were not pre-determined by the researcher (Hudson, 2013). Participants were free to determine the names and number of categories, along with any relationships they wished to express. They were asked to sort into categories that would help them to find the documents again later, and to give the categories names. There were no limits on the number of categories or the number of cards in each category. Each card had a picture of the document (see Figure 1) and a limited description of it. Actual cards were used for the activity, providing a consistent approach to both physical and electronic documents. Carlyle (1999), in her consideration of the categorisation of works for improved OPAC displays, determined that cards were as good as the objects themselves for such categorisation.

![Figure 1: Example of three cards](image-url)
The groups of participants were selected to enable a variety of perspectives. Recordkeepers have a professional understanding of documents and the relationships that exist between them, while teachers have a professional role in communicating how to use and create documents, and an understanding of learning styles. Secondary school students are taught to use documents, and taught through documents. Such students are under-represented in information behaviour research as compared to tertiary students (for example, De Rosa, et al., 2006; Head & Eisenberg, 2011; Hendry & Efthimiadis, 2008; Holman, 2009; Zhang, 2008a, 2008b, among many others). Thus, they provide a different perspective to what is already available in the published literature. The participants are characterised as library users and enthusiastic readers; the research does not consider the mental models of those who do not use libraries.

The card sorting activity formed the first part of interviews with participants. Students were interviewed in three groups, rather than individually as the recordkeepers and teachers were, to provide a more naturalistic approach to the research than one-on-one discussion with the researcher. Students were Year 10 (the second year of New Zealand’s secondary schooling) and around 14 years old. This age group is considered able to solve formal operational tasks, use scientific reasoning (to a point), and to think independently without relying on prior experience alone (Sigelman, Rider and De George-Walker, 2013, pp. 169-174). Each student group is reported as one participant as they jointly sorted the cards and developed category names. There were thus 22 participants but 15 sets of results.

The cards represented a limited set of documents and of necessity represented only a cross-section of the bibliographic universe. At this point in the research, which is ongoing, no attempt is made to suggest that results demonstrate average, general, or even widely held understandings; rather, it presents some of the possible ways these participants understand this cross-section through an examination of the categories they name.

**Findings**

A brief overview of the findings is provided in this section. Participants had 10 categories on average: the recordkeepers had the higher average of 11.8; the teachers 8.6; students were in the middle on 9.3. The smallest number of categories was four (a recordkeeper), with categories named: ‘Source, original’, ‘Reference’, ‘Work inspired by’, ‘Reference to work inspired by’. The recordkeeper noted:

“I’m a big bucket kind of a guy ... I really would like it to be just two groups, you know ... Source, original, and Ephemera.”

If outliers in each of the cohorts are removed (the recordkeeper with four categories and the teacher with 15 categories), the averages change dramatically to 13.4 for recordkeepers and 7.4 for teachers.
Unsurprisingly, the more categories named by participants, the more possibilities there were for precise category names and finely granular sorting rather than big bucket approaches. There were 10 participants who had at least one category with a single card in it. The card most commonly in a category by itself was the sheet music of a song from the movie (n=8), followed by the sequel The Lord of the Rings (n=6) and the Wikipedia page for The Hobbit (n=4).

Category names were sometimes single words (‘Music’, ‘Movies’, ‘Translations’); sometimes descriptive phrases (‘The Hobbit, all versions and languages’, ‘Info to help you study’, ‘Separate and different aspects’); and sometimes comparative descriptions where all category names provided by a participant needed to be considered before each was clear. For example, one recordkeeper named categories: ‘Main work’, ‘Further editions, later printings’, ‘Other formats, derivatives’, ‘Translations, different renditions’. There was a slight tendency for participants to use the same kind of category name (e.g., those that used descriptive phrases did so for most categories). The students were the least likely to use descriptive phrases and their category names were predominantly one or two word terms (see Table 1) developed by consensus.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movies</td>
<td>Movies</td>
<td></td>
</tr>
<tr>
<td>Graphic novels and pictures, art</td>
<td>Graphic novels</td>
<td>Graphic novels, other ways, including movies</td>
</tr>
<tr>
<td>Reviews</td>
<td>Online</td>
<td>Reviews, Internet</td>
</tr>
<tr>
<td>Fan and blog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>Non-fiction</td>
<td>Books about</td>
</tr>
<tr>
<td>Wikipedia/Information [the Wikipedia card]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Book in English</td>
<td>The book</td>
<td>The book, different languages and ways</td>
</tr>
<tr>
<td>Translations</td>
<td>Languages</td>
<td></td>
</tr>
<tr>
<td>Related to</td>
<td>Geography (or “maps and stuff”)</td>
<td>All Middle Earth</td>
</tr>
<tr>
<td>Music</td>
<td>Miscellaneous</td>
<td>Music</td>
</tr>
</tbody>
</table>

*Table 1: Student group categories*

Phrases were used by students where consensus on a single term was not reached. For example, Group 3’s ‘Graphic novels, other ways, including movies’ encompassed each student’s ideas. Group 1’s ‘Wikipedia/Information’ was both the precise name of the single card in that
category (a Wikipedia page) and the description of the kind of thing that they thought it was (information).

**Familiar Categories**

There were documents and documentary forms that seemed easy for participants recognise and categorise, that did not cause any hesitation or confusion. These were put in categories that are recognisable outside the context of this particular card sorting activity (music, poems, quiz, movies, graphic novels, and translations) and can be considered to be rhetorical genres (among other things). In contrast, other categories were named as a way of describing subsets of this particular set of cards or, in the case of the students, to express a collective understanding (e.g., ‘Fictional information about the books’; ‘Studying and notes and biography’, ‘Graphic novels and pictures/art’, and, ‘Descriptions without criticism’). The easy to recognise documents and documentary forms are being termed *cultural markers*: they appear to indicate culturally familiar categories of types of document that are distinctive for many participants regardless of the context. In particular, these are poems, (sheet) music, movies, graphic novels, and translations.

Eleven participants had a category that mentioned translation; two others had a category that mentioned language (‘The Hobbit, all versions and languages’, ‘The book, different languages and ways’); two did not make any mention of them (the recordkeeper and teacher who had the least categories). Category names were ‘Translations, different renditions’, Translations’ (n=4), ‘Foreign language/s’ (n=2), Versions in other languages’, ‘Editions in different languages’, ‘Languages’, and ‘Other translations’. ‘Other translations’ included the movies and the graphic novels in English as well as in German, so here translation was being used in a wider sense. One recordkeeper put the movies with ‘Versions in other languages’, half-seriously telling me they were “Hollywood language”, while a teacher categorised the card labelled ‘English language edition’ with ‘Editions in different languages’ despite The Hobbit itself being in English, noting:

> “I put in the English one as it does say English language edition because it might apply to other countries or if you were in another country you might want an English language edition.”

It is possible that participants saw the documents in other languages as something they were less interested in and thus easy to separate out. However, the translations were also positioned by some participants as arising out of the original work, or a subset of that work.

‘Graphic novels’ was used as or in a category name by all three student groups, but only by one adult (a recordkeeper), indicating the students’ greater awareness of a literary genre or format often associated with teenagers. Most of the 10 students had read at least one graphic novel, and for them, graphic novels can be considered a cultural marker. Other adults spread the graphic novels between translation categories (n=7; one graphic novel was in German), ‘Adaptations’, ‘Versions of historical importance’, ‘Further editions,
later printings’, ‘What Tolkien actually wrote’, and other more general categories such as ‘The book’. It was not possible to ask about the rationale for every single card’s placement in the time available, so it remains unclear why one participant thought they were ‘Versions of historical importance’. One disadvantage of the card sort is that participants did not always appear to read the information on the cards carefully, and may have misunderstood or misinterpreted what a card stood for. In support of this suggestion, some participants moved cards to different categories when explaining their categories to me and realising what the cards actually were.

Apart from identifying the familiar categories or cultural markers, one of the important findings was that there is not necessarily a correlation between a particular named category and the documents included within that category by different participants. That is, the same cards were not always included in the similarly named categories. Similarly named categories are those judged to have the same intention or purpose, such as with the translations. For example, when categorising the Tolkien biography, only two participants used the word ‘Biography’, two others used ‘Tolkien himself’ and ‘About the author’ while the rest included the biography in a broader category with other documents including those that would be considered non-fiction in a library and those that arose as a direct response or tribute by fans to Tolkien’s writings (termed ‘Geek stuff’ by one participant). There is a general sense of what participants thought about what a biography is and where it should be categorised, but a good deal of flexibility in the actual terminology and the emphasis (or otherwise) that they placed on the biography. Rosch (1978, p. 10) notes that “Most if not all categories do not have clear-cut boundaries”, even when the features of that category are clearly understood and identifiable. Here, category names were flexible and boundaries were not fixed, although there seemed to be a core group of documents that was generally agreed upon as part of a category. The least variation occurred with the students’ categories, and similarly named categories have been presented on the same or adjacent rows in Table 1 to illustrate this lack of variation.

Discussion
Three aspects arising out of this stage of the research are discussed here.

No Single Approach
There is no single consistent approach to categorising this particular set of cards used by any participant. The documents were categorised according to different literary genres, rhetorical genres, reasons for using, format, accessibility, and form. As well, some participants indicated relationships between categories that implied an awareness of intertextuality (cf Seely Brown & Duguid, 1996); some had sub-categories; others produced a flat set of categories with no expressed relationships between them. While I consider participants’ categories to be a mixture of genre, form, and type, with accessibility and possible use as additional determinants, from the
participants’ point of view their categories were optimal (at least on the day) and they did not see it as uneven, inconsistent or a mixture. For each participant, it was the categories they thought were necessary.

Rosch (1978) notes that “maximum information with least cognitive effort is achieved if categories map the perceived world structure as closely as possible” (p. 3). Here, categories are the ones that participants thought would help them to re-locate the documents at a later date; their own perceived world structure. As well, a range of granularity occurred within a participant’s approach. One teacher categorised together all versions of The Hobbit including graphic novels and the movies, but had a separate category for translations. However, in the subsequent interview he demonstrated a strong awareness of the granularity to be found in detective fiction, his favourite literary genre.

A point of comparison with the apparent mixture of approaches is the BISAC subject headings and codes. These are developed by the Book Industry Study Group (BISG) as part of their effort to standardise best practice:

BISG develops and maintains a number of classification systems for both physical and digital products. The systems can be used individually or together to help determine where the work is shelved in a bricks-and-mortar store or the genre(s) under which it can be searched for in an online database. (BISG, 2014a)

These encompass a range of approaches as varied as those used by the participants, mixing literary genre, rhetorical genre, format, and form, such as reference, music (subdivided into genres and styles, and printed music among other things), poetry, drama, Bibles, graphic and comic novels, fiction, juvenile fiction, juvenile non-fiction, literary collections, literary criticism, and study aids (BISG, 2014b). Given the extent of BISG’s research, it is a reasonable assumption that these headings reflect the book purchasing public’s understanding. The current research supports this assumption.

**Domain Knowledge**

Participants were selected because they were considered – personally and occupationally, by themselves and by the researcher – to use and have a good understanding of documents and their organisation in various contexts. Three different communities of practice (cf Bowker and Star, 2000) or discipline were selected to see whether there was any consistent understanding of the bibliographic universe that could be attributed to their domain knowledge. Differences exist between the participant groups, and within the participant groups, but it is difficult to determine how significant these are (or whether they are significant at all). As noted, recordkeepers had on average more categories than the students and teachers, meaning their category names were more varied and specific, and they showed a greater awareness of the range and difference between different types of document. They also did the card
sort faster than the teachers, suggesting that the process of categorisation was a more familiar one. Five of the recordkeepers indicated relationships between the categories, which could be linked to an awareness of the contexts and documentary relationships that is found in organisational recordkeeping. Only two of the teachers indicated relationships and none of the students did so.

The question of whether communities of practice have different understandings of the bibliographic universe is complicated by the fact that New Zealand’s recordkeepers frequently have a background in other areas before coming to recordkeeping (see Cossham, 2004, p. 46). Here, recordkeepers had backgrounds in primary teaching (two), academia, ‘nothing professional’, librarianship, and theology. In contrast, four of the teachers had teaching as their only career. Personal interests also played a part for all participants (including the students), with genealogy and family history, political engagement, hobbies, sport, private research, and current study potentially having an impact on their understanding, although these aspects were discussed during the interview rather than in the card sorting itself. As a general tendency, the lenses of recordkeeping and teaching could be said to, respectively, sharpen and make more granular, and blur and broaden, the categories identified.

The secondary school students had a good understanding of most of the types of documents on the cards, and strong awareness of the quality of the information. Their categorisation was remarkably similar between the groups emphasising a common understanding. They do not have discipline knowledge as such; rather, they have an educational system framework that serves as a default community of practice. Information literacy practices are embedded in the New Zealand school curriculum. The students’ understanding of document forms and genre was fostered by the school’s thorough approach to information literacy, and strongly supported by the school librarian, who was mentioned enthusiastically by all students as a source of information about the resources they needed for their study.

All Year 9 and 10 students are taught about different documentary forms (including literary and rhetorical genres) and information search processes using guided inquiry (for discussion of guided inquiry generally, see Kuhlthau, 2010; Kuhlthau, Maniotes and Caspari 2007; Kuhlthau, Heinström and Todd, 2008). They conduct a guided inquiry or project, a sustained piece of work on a particular topic running across all of their different subject classes. For the Year 10 topic, students were instructed to:

“Do some background research to form a good research question”; solve the question using “a range of sources, primary and secondary”; check that they are valid sources and use the best ones – “Call experts, email experts, read books, read online, watch documentaries etc…”; write down relevant information from sources; “select the best information” from their notes. (Summarised and quoted from the school’s Year 10 guided inquiry booklet).
Their information sources must be assessed for usefulness and reliability (guidance is provided on these aspects). Students complete an Evidence collection sheet for each “piece of evidence” and must submit three to six of these sheets. A checklist ensures they have:

“Filled in the bibliographic information, written short notes, made sure [the] notes are relevant to the question, taken notes on both sides of the issue, taken notes from at least 3 sources, tried to take notes from a wide range of sources”. (Summarised and quoted from the school’s Year 10 guided inquiry booklet).

In the interviews, all students emphasised the quality of information as a factor in categorisation, and were the only participants to do so. Their card sorting was slower than the recordkeepers, partly due to doing the activity in small groups, and some of the documents, such as the dissertation, were unfamiliar to them, but in each group at least one student knew about each type of document. They were close to the average of all participants for the number of categories, and as has been noted, had similar cards in similarly named categories.

**Perceived Attributes and Cultural Markers**

Categories are based on perceived attributes, and in this set of cards, participants had to rely on limited information about each document along with any pre-existing knowledge they had of document types and these particular documents (Rosch, 1978, p. 42). There is little doubt that participants had pre-existing knowledge and indicated familiarity with the cultural markers but they struggled to categorise less familiar genres or documentary forms such as a dissertation, a quiz book, and reviews of the movie. Some of the categories including only one document were a default ‘I don’t know what to do with this’ categorisation, and several participants had what amounted to miscellaneous categories. There was definitely a sense from some participants that they would not be interested in all of the documents in this set, and therefore it didn’t matter quite as much where they were put: big buckets were suitable. Rosch (1978, p. 8) notes that “inseparable from the perceived attributes of objects are the ways in which humans habitually use or interact with those objects”. The context of likely use, or purpose, is a variable that has an impact on categorisation and may affect the level of granularity and abstraction applied. Recordkeepers (apart from one) and students seemed less affected by likely use than the teachers, possibly because recordkeepers are used to organising documents that they themselves are not likely to use, while students simply accepted this as an activity around a well-known book and movie.

The category that stands out is translations. These were unerringly identified by 13 of 15 participants, including the students, and can be
considered a familiar category. The word ‘translation’ was not printed on any of the cards, but the language of the document was specified. This does raise the question: what are translations? They are not a literary genre, nor are they a form or format. Although a case could perhaps be made for considering them to be a rhetorical genre, arising out of a particular social action, this is a little tenuous. The Library of Congress Genre/Form Terms for Library and Archival Materials (Library of Congress, 2015) does not include any suitable terminology.

Here, then, is a kind of way of thinking about documents that is familiar to participants but may not be well catered-to by existing library approaches. The translator and the process of translation are frequently invisible, yet against that can be set the clear identification of translations by the participants. Hofstadter, in his book on the art of translation, notes the lack of conscious awareness of a translator when we read a text in our own language:

… translation tends to be one of those “out of sight, out of mind” kinds of things. Most readers take translators and translations for granted. … We basically are taught – both by omission and by commission – to ignore, forget about, and disrespect translators. (Hofstadter, 1997, p. 355).

A translation is usually a subsequent instantiation of a work following its success in the original language. That is, the document that is a translation must be based on another document, not on the original work or expression (text), both of which are considered to be abstractions in the FRBR model. Simultaneous translation and publication is much less common, although there are authors (e.g. Italo Calvino) and situations (e.g., official publications in bilingual countries) for which this is standard practice. Pisanski and Žumer (2012, p. 584-5) noted that while their participants identified with a FRBR model, there was a tendency for them to place an expression (text) in the original language higher up in a model’s hierarchy than subsequent expressions in other languages, and in some cases, to see subsequent expressions arising from the original manifestation (document) rather than from the work itself. The FRBR model does not express the process of creation; it is a static model.

**Conclusion**

While acknowledging that the range of documents was limited for practical reasons (even 36 cards was considered a lot to sort by some participants), the outcomes of this card sorting activity provide an indication of how these participants think about the categories that are significant to them in the current information landscape. Their perspectives, individually and collectively, are not generally like the models that are used by libraries, nor do they match any other particular knowledge organisation system exactly.
Participants identified form and accessibility as characteristics for categorising as much as they did literary or rhetorical genre, or a fiction/non-fiction split. Bowker and Star (2000) note that “As the information systems of the world expand and flow into each other, and more kinds of people use them for more different things, it becomes harder to hold to pure or universal ideas about representation or information” (p. 301), while Mazzocchi suggests that:

The possibility (and the actual existence) of multiple descriptions and classifications is also due to the fact that they have offered and still offer a meaning to the lives of those people using them, and a basis for survival and co-adapting in their world-environment. (Mazzocchi, 2013, p. 371).

Deodato (2014) has suggested that a participatory culture could help create space in libraries for marginalised discourses, and expand conventional library systems to better represent a broader range of discourses. While the participants in this research cannot be considered marginalised, their particular way of categorising is not well-reflected in current library catalogues, and a participatory, Web 2.0 approach, might help mitigate this.

What implications might this have for library catalogues? There is no doubt that the complex and highly structured data in bibliographic records is an important knowledge organisation scheme. As they stand, however, catalogues do not reflect very well the kinds of categories that participants expect. There is a tension between professional understanding and expertise developed over decades if not centuries and that leads to effective knowledge organisation, and the changing information and new media landscape. This does not obviate the need for knowledge organisation but expands greatly the range and type of documents that exist, and the ways in which these documents can be accessed, used, re-used, and read. The easy access that the Internet offers in finding and retrieving documents has an impact on how people categorise, and on the models they have of knowledge organisation systems. Library cataloguing does not take into consideration aspects such as genre theory, translation theory, paratext or documentalism. The needs of the individual do not seem to be adequately met by library catalogues, even when they meet the needs of society by having knowledge organised in this way. Smiraglia (2009, 2015) reminds us that “The online [library] catalog of the digital age is just one of many retrieval systems making up a rich complex of tools for resource discovery” (2015, p. 1).

Svenonius (2000) emphasised that “The role of the bibliographic record in a digital environment is not yet clear. Especially unclear is what exactly a bibliographic record should describe” (p. 64). Over the 15 years since 2000 this has not become much clearer. Merčun et al. (2013) point out that we still need to work out what bibliographic data users actually need, and that different user groups have different requirements. There is a (relatively) new set of cataloguing guidelines, RDA, which was explicitly intended to
describe digital entities, but the context in which those digital entities are being created and in which library catalogues must function continues to change rapidly. The FRBR model that provides a conceptual framework for RDA needs further consideration in the light of the changing information landscape and the consequential change in users’ expectations and understanding (Cossham, 2013). Cain has suggested that:

the real value, which prompts individuals, organizations and governments to provide funds for libraries, archives and kindred cultural resources, is seen to be in the resources themselves; the metadata, and value found there, are secondary. … We have to devise ways, not too expensive, to live and operate in a bibliographic universe where inconsistencies and discrepancies are normal. (H. Cain, personal communication, June 25, 2015)

Documentary forms are changing, digital formats are proliferating, and the library catalogue has not yet fully risen to the challenges these bring. We need to see library catalogues as one of many possible information retrieval systems, and we need to incorporate approaches that better reflect the understanding – or lack of understanding – of users.

References


