

**“Oh the places you’ll go!”- Dr Seuss**

**A study of picture book booksonomies within LibraryThing.**

**Jenny Sarah Jones**

A dissertation submitted to the University of Wales in partial fulfilment of the requirements for the degree of Magister in Scientia Economica (MSc) under Alternative Regulations.

Department of Information Studies

Aberystwyth University

2016

## Abstract

The aim of this study was to investigate picture book booksonomies on LibraryThing with a particular focus on the ‘user warrant’ that might be revealed in the tags or tag categories that are discovered.

A quantitative research method was used to analyse a sample of ten sets of picture book tags (and their frequencies) using a categorisation model adapted from previous research yet influenced by the ‘warrant’ demanded by the tags themselves. 5,568 individual tags were applied 80,465 times within this sample- the distribution of the frequency with which the tags were applied obeyed the Zipfian Power Law suggesting that the booksonomies were a stable set of tags containing a relevant level of consensus.

Each set of tags was analysed within implicit context to the book that they related too, results were analysed both at the individual book level and across the sample. The largest single tag category across all of the booksonomies was the ‘*subject*’ category which accounted for 34.3% of the tags. 5% of the tags sampled were categorised as relating to ‘*use*’- this was further broken down into six distinct categories.

The tags from the ‘*subject*’ tag category were ‘mapped’ onto the relevant LCSH using an adapted model. 38.1% of the tags were matched at the 2<sup>nd</sup>, ‘almost syndetic’ level and 67.7% at the 3<sup>rd</sup> or ‘semantic level’. Tags not matched were grouped into semantic concepts and clear subject headings emerging from the booksonomies were observed.

The results revealed a clear ‘user warrant’ for the ability to apply ‘*use*’ descriptors to a picture book resource and evidence of subject headings emerging from the booksonomies both which have implications for designers and users of IRS that include substantial collection of picture books as well as for future researchers.

**Declaration**

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date .....

Statement 1

This work is the result of my own investigations, except where otherwise stated. Where correction services\* have been used, the extent and nature of the correction is clearly marked in a footnote(s). Other sources are acknowledged (e.g. by footnotes giving explicit references). A bibliography is appended.

Signed ..... (candidate)

Date .....

[\*this refers to the extent to which the text has been corrected by others]

Statement 2

I hereby give consent for my work, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....

<b>Contents</b>	
<b>Abstract</b> .....	<b>2</b>
<b>Declaration</b> .....	<b>3</b>
<b>Acknowledgements</b> .....	<b>9</b>
<b>1. Introduction</b> .....	<b>10</b>
<b>1.1 Overview</b> .....	<b>10</b>
<b>1.2 Aims</b> .....	<b>10</b>
<b>1.3 Objectives</b> .....	<b>10</b>
<b>1.4 Scope</b> .....	<b>11</b>
<b>1.5 Structure</b> .....	<b>12</b>
<b>1.6 Referencing</b> .....	<b>13</b>
<b>2. Literature Review</b> .....	<b>14</b>
<b>2.1 Introduction and definitions</b> .....	<b>14</b>
<b>2.2 Folksonomies vs traditional taxonomies</b> .....	<b>14</b>
<b>2.3 Research into folksonomies</b> .....	<b>15</b>
2.3.1 The Zipfian Power Law- frequencies of tags, convergence and emerging consensus..	15
2.3.2 The ‘long tail’ .....	16
<b>2.4 Mapping LCSH to folksonomy tags</b> .....	<b>16</b>
<b>2.5 Tag categories</b> .....	<b>17</b>
2.5.1 Studies utilising tag categories .....	17
2.5.2 ‘Aboutness’ as a tag category concept .....	18
2.5.3 The importance of categorising in context .....	19
<b>2.6 Tagger motivation and “warrant”</b> .....	<b>20</b>
2.6.1 User warrant .....	20
<b>2.7 Research into booksonomies</b> .....	<b>21</b>
<b>2.8 Picture book research</b> .....	<b>22</b>
<b>2.9 Summary</b> .....	<b>24</b>
<b>3. Methodology</b> .....	<b>26</b>
<b>3.1 Introduction</b> .....	<b>26</b>
<b>3.2 Choice of research method</b> .....	<b>26</b>
<b>3.3 Data collection</b> .....	<b>27</b>
3.3.1 Selecting the sample .....	27
3.3.2. LCSH collection .....	30
3.3.3 Harvesting and cleaning the data.....	31

<b>3.4</b>	<b>Categorising the tags .....</b>	<b>32</b>
3.4.1	The development of the category model .....	32
3.4.2	Categorisation in context .....	35
3.4.3	Trial book & modification of the categories .....	36
3.4.4	Second trial book .....	37
3.4.5	Selecting the sample range- whether or not to include the long tail .....	38
3.4.6	Issues with tag categorisation .....	40
3.4.7	Methods of analysis: tag categories .....	40
<b>3.5</b>	<b>Mapping the ‘subject’ category tags to the LCSHs.....</b>	<b>41</b>
3.5.1	Methods of analysis: LCSH & ‘subject’ tag matching.....	42
<b>3.6</b>	<b>Methods chapter summary and limitations .....</b>	<b>42</b>
3.6.1	Issues with the sample .....	42
3.6.2	Issues with the categorisation process .....	44
3.6.3	Issues with the mapping process .....	44
<b>4.</b>	<b>Results.....</b>	<b>45</b>
<b>4.1</b>	<b>Introduction .....</b>	<b>45</b>
<b>4.2</b>	<b>Data collection and sample .....</b>	<b>45</b>
4.2.1	The long tail and the sample range.....	45
<b>4.3</b>	<b>Categories.....</b>	<b>46</b>
4.3.1	Category groups.....	47
4.3.1.1	“Aboutness” group of tags analysis .....	48
4.3.1.1.1	‘Subject’ tag category analysis .....	48
4.3.1.1.2	‘Literary’ tag category analysis .....	51
4.3.1.1.3	‘Use’ tag category analysis .....	51
4.3.1.1.2	Remaining tags .....	53
<b>4. 4</b>	<b>Mapping .....</b>	<b>54</b>
<b>4.5</b>	<b>Summary of findings .....</b>	<b>56</b>
4.5.1	The long tail.....	56
4.5.2	Categories .....	57
4.5.3	Mapping.....	57
<b>4.6</b>	<b>Limitations .....</b>	<b>57</b>
<b>5.</b>	<b>Discussion .....</b>	<b>59</b>
<b>5.1</b>	<b>Introduction .....</b>	<b>59</b>
<b>5.2</b>	<b>Literature review and methodology.....</b>	<b>59</b>
5.2.1	Long tail.....	59

<b>5.3 The categorisation process.....</b>	<b>60</b>
5.3.1 "Aboutness" group of tags.....	61
5.3.2 'Literary' category tags .....	61
5.3.3 'Use' category .....	63
5.3.4 Implications of the categories discovered .....	64
5.3.5 Implications of the findings for cataloguers and IRS.....	65
<b>5.4 The mapping process.....</b>	<b>66</b>
5.4.1 Subject tags not matched and emerging consensus.....	69
5.4.2 Implications of the tag mapping findings for traditional taxonomies .....	70
<b>5.5 Summary of discussion .....</b>	<b>70</b>
<b>6. Conclusion .....</b>	<b>72</b>
<b>6.1 Introduction .....</b>	<b>72</b>
<b>6.2. A review of the aims and objectives.....</b>	<b>72</b>
6.2.1 Aims .....	72
6.2.2 Objectives.....	72
<b>References.....</b>	<b>76</b>
<b>Bibliography.....</b>	<b>83</b>
<b>Appendix.....</b>	<b>85</b>
<b>Appendix A: Final 'Aboutness'/connotative group categories used during data analysis85</b>	
Tag category: Subject .....	85
Tag category: Literary .....	85
Tag category: Use: Pedagogical: Formal.....	85
Tag category: Use : Pedagogical: Formal: Named product.....	85
Tag category: Use : Pedagogical: Formal: Levels.....	85
Tag category: Use: Pedagogical: Informal .....	85
Tag category: Use: Pedagogical: Socialisation.....	86
<b>Appendix B: Examples of 'difficult' tags and how they were interpreted.....</b>	<b>86</b>
<b>Appendix C: LCSH/LibraryThing three stage mapping process details.....</b>	<b>87</b>
<b>Appendix D: number of matches (Fr) .....</b>	<b>88</b>
<b>Appendix E: number of matches % (Fr).....</b>	<b>88</b>
<b>Appendix F: Hungry Caterpillar new subject heading suggestions .....</b>	<b>89</b>
<b>Appendix G: Corduroy new subject heading suggestions .....</b>	<b>91</b>
<b>Appendix H: Alexander new subject heading suggestions .....</b>	<b>91</b>
<b>Appendix I: Where the Wild Things Are new subject heading suggestions.....</b>	<b>93</b>

## List of tables

Table 3.3 Sample books ranked by frequency of “picture book” tag .....	28
Table 3.8 Screenshot of initial categories within the ‘Aboutness’ group .....	33
Table 3.9 Screenshot of initial categories within the ‘Isness’ group .....	34
Table 3.10 Screenshot of initial categories within the ‘Personal’ group .....	35
Table 3.11 Screenshot of initial category spread within ‘isness’ group .....	35
Table 3.12 Screenshot of final categories within the ‘Aboutness’ group .....	37
Table 3.15 Comparing the whole sample data across the trial books and the ‘aboutness’ categories .....	39
Table 3.18 Showing a sample of ‘subject’ tags prepared for Wordle.net.....	40
Table 3.17 Showing the three level matching process for LCSH ‘ <i>metamorphosis</i> ’ applied to The Very Hungry Caterpillar .....	42
Table 3.19 Listing some of the 139 ‘use’ tags in the “Pigs” long tail sample .....	43
Table 4.13 Showing the mapping data broken down by book.....	55
Table 4.14 Showing the % not matched in ranked order compared to the mean and median average. ....	55
Table 5.2 ‘ <i>Literary</i> ’ tags applied to “Corduroy” .....	63
Table 5.5 LCSH tags for each book compared to the % of tags not matched .....	68
Table 5.7 subject headings ‘suggested’ by the non-matched subject tags within the “Caterpillar” data .....	69

## List of figures

Figure 1.1 Overview diagram of the research structure including research questions ..	12
Figure 3.1 Screenshot of ‘ <i>picture book</i> ’ tags and its aliases.....	28
Figure 3.2 Screenshot of books sampled .....	28
Figure 3.4 Sample of a page from “Pigeon” .....	29
Figure 3.5 Screenshot of WorldCat entry for “The Very Hungry Caterpillar”.....	30
Figure 3.6 Screenshot of LibraryThing tags for “The Very Hungry Caterpillar” .....	31
Figure 3.7 Screenshot of raw data being cleaned within Word .....	31
Figure 3.13 Showing the Zipfian Power Law curve in the complete data sets for “G’night Moon” and “Pigs” .....	38

Figure 3.14 Showing the Zipfian Power Law curve in the selected sample range (tags applied twice or more) for “Goodnight Moon” and “Pigs” .....	39
Figure 3.16 Visualisation of LCSH ‘Conduct of Life’ .....	41
Figure 4.1 Showing the total average distribution between the tags applied once and more than once (Fr).....	46
Figure 4.2 Category distribution across all ten books using the mean average, .....	46
Figure 4.3 Category group distribution across all ten books using the mean average...	47
Figure 4.4 Total tag category distribution.....	48
Figure 4.5 ‘ <i>Subject</i> ’ category distribution (Fr) .....	49
Figure 4.6 Word cloud visualisation of the ‘ <i>subject</i> ’ category tags from the “Caterpillar” set weighted by frequency.....	50
Figure 4.7 ‘ <i>Literary</i> ’ category tag distribution (Fr).....	51
Figure 4.8 ‘ <i>Use</i> ’ category tag distribution (Fr).....	51
Figure 4.9 ‘ <i>Use</i> ’ category tags further analysed into sub categories (Fr).....	52
4.10 “ <i>Use: pedagogical</i> ” category distribution (Fr) .....	53
Figure 4.11 Tags not within the “Aboutness” group across all ten books (Fr).....	53
Figure 4.12 showing total % matched and not matched using the mean average (Fr) .	54
Figure 4.15 Plotting % of tag matches (Fr) at each stage .....	56
Figure 5.1 ‘ <i>Literary</i> ’ tags applied to “Brown Bear”, weighted according to frequency and visualised using Wordle. ....	62
Figure 5.3 ‘ <i>Use:pedagogical</i> ’ tags applied to “Alexander”, weighted according to frequency applied.....	63
Figure 5.4 % of LCSH matched at each stage using the LCSH as the ‘benchmark’ .....	67
Figure 5.6 LCSH tags for each book compared to the % of tags not matched.....	68

## Abbreviations

**IRS-** Information Retrieval System

**OCLC-** Online Computer Library Centre

**LCSH-** Library of Congress Subject Headings



## **Acknowledgements**

Firstly I would like to thank my husband Martin for all of the support and encouragement he has given throughout the course of my studies, I couldn't have done it without you. Thanks too must go to my sons Elliott and Sam for their patience, I hope I have not put either of you off distance/lifelong learning some day.

Thanks to my Mum for not putting me off distance/lifelong learning when she studied with the Open University while I was growing up- you were an inspiration. Thanks also to my Dad for his patience with my many questions about statistics and data and for his enthusiasm for my topic. It's about twenty two years ago since you introduced me to the 'world wide web' for the first time and you were right, it did change everything! You both provided a quiet space for me to write this dissertation up in and for that I am very grateful.

Finally thanks need to go to Pauline Rafferty for agreeing to supervise this dissertation and to the fabulous Sue Lithgow who always answered my questions quickly and in the most reassuring manner possible.

## **1. Introduction**

### **1.1 Overview**

Formal, authority controlled subject headings such as Library of Congress Subject Headings as used within traditional taxonomies are increasingly being compared and contrasted with the tags or tag categories sampled from socially constructed, non-hierarchical, uncontrolled folksonomies. The study of folksonomies allows the researcher to get a glimpse into the behaviour and motivation of the tagger and what they perceive the object they are tagging to be 'about' or what they perceive to be the relevant facts related to it.

This study takes a quantitative approach to the investigation of picture book booksonomies sampled from online site LibraryThing in order to gather insight into the particular practices, trends, and cultural and user warrants displayed by picture book taggers. Academic and picture book specialist Peter Hunt claims in his introduction to "*Understanding Children's Literature*" that "children's books are used for different purposes at different times- for more things than most books are" (1999, p. 11) this study will attempt to uncover evidence of this "use" of the children's book by focussing on a the picture book sub-group in particular.

### **1.2 Aims**

*To investigate picture book booksonomies on LibraryThing with a particular focus on the 'user warrant' that might be revealed in the tags or tag categories that are discovered.*

### **1.3 Objectives**

*1. To conduct a literature review into folksonomies and social tagging with an emphasis on categorisation, mapping with authorised terms such as LCSH and tagger motivation*

*2. To categorise tags obtained from a sample of picture books on LibraryThing adapting existing models where relevant and to investigate the ways in which tags are applied to picture books by analysing the conceptual categories that they fall into.*

3. *To extract the 'subject' category tags and use these to 'map' against LCSH, investigating any conceptual groups of tags left for possible emerging consensus for additional subject headings.*
  
4. *To consider whether the booksonomy of a picture book might be a relevant source of data for further research into both social tagging and picture books.*
  
5. *To consider the practical implications of the research for cataloguers and designers of information retrieval systems that contain picture books*

#### **1.4 Scope**

This research was fundamentally limited by the scope of the study and the time scale involved. The sample was limited to ten books to ensure that a comprehensive, contextual analysis of the tags was carried out in order that the 'user warrant' of the picture book taggers was as accurately represented within the category headings and categorisation of tags as was possible. The study was also limited by the very subjective nature of the categorising and mapping processes as they were only carried out by the researcher with her own bias and subjective views.

The study of folksonomies within the information studies is a relatively new yet very dynamic field. As Library 2.0 technology develops and more users supply data within folksonomies via sites like LibraryThing there is an increasing amount of interest in what these informal, democratic and user-generated systems might reveal. As yet no previous specific research has been undertaken into either children's booksonomies or picture book taxonomies within a traditional cataloguing context so there are no directly comparable studies to build upon.

Picture book research is an emerging area of research, picture books and the academic field of study dedicated to them have in the words of Hunt been previously 'marginalised' (1998, p.1). A strong research field is developing however focused on social/political

interpretations of picture books and the relationships between adults who write, publish, buy, choose and read them and the children who they are intended for (Colomer, Kummerling-Meibauer, Silva-Diaz, 2010, p.1).

### 1.5 Structure

The structure of this study is loosely based around three main stages and this is reflected in the sections within the methodology, results and discussion chapters. Firstly the sample was chosen from LibraryThing and the tag data was collected (including frequency), then the tags were sorted into categories within a category model adapted from relevant research. Finally the tags in the ‘*subject*’ category were ‘mapped’ onto the LCSH applied to the same books collected from WorldCat.

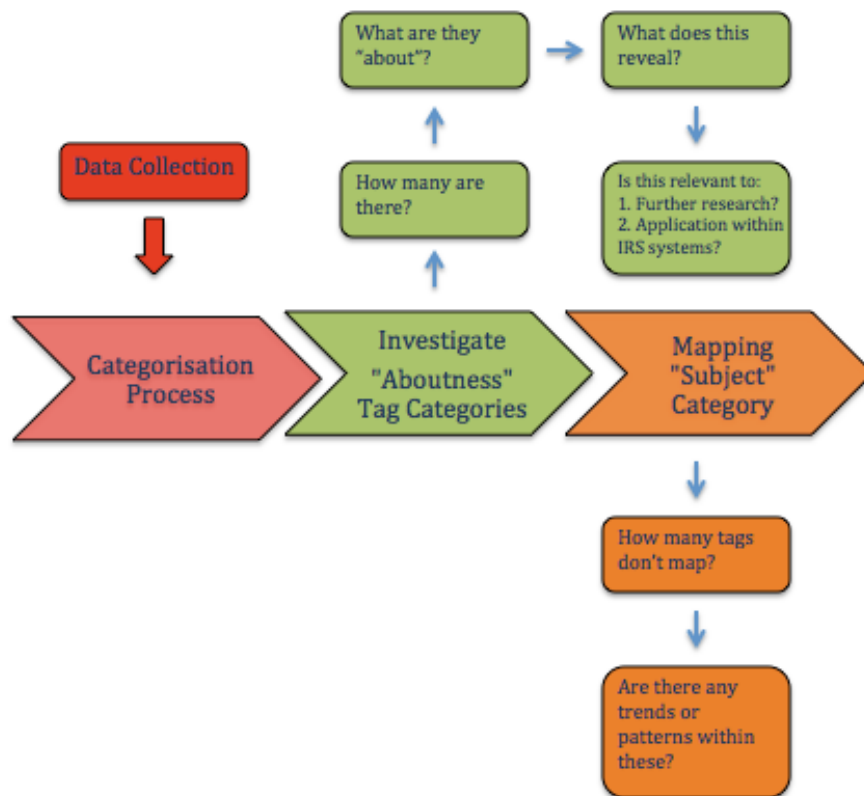


Figure 1.1 Overview diagram of the research structure including research questions

At every stage the results were analysed at both an individual booksonomy level and as a whole sample but the results obtained are contextual rather than aggregated. The

implications for both further research and practical application are discussed in the discussion chapter.

## **1.6 Referencing**

The referencing and citation style used throughout this study is Harvard (APA).

## **2. Literature Review**

### **2.1 Introduction and definitions**

Vander Wal is credited with creating and defining the term ‘folksonomy’ on a post he wrote in a closed Listserv list to describe what he later explained was “the user-created bottom-up categorical structure development with an emergent thesaurus” generated by social tagging (2007). Tags are the term given to “freely selectable keywords... which can be liberally attached to any information source” (Peters, 2009, p. 153). The process of applying these tags is called ‘tagging’ thus “collaborative tagging describes the process by which many users add metadata in the form of keywords to shared content” (Golder & Huberman, 2006, 198). As Peters explains “Folksonomy is a combination of the words ‘folk’ and ‘taxonomy’ and simply means ‘a taxonomy created by the people’” (2009, p.154). Research into ‘folksonomies’ therefore is research into the way that the resultant folksonomy behaves and functions, ‘booksonomies’ have been described as “a folksonomy specifically containing book tags” (Guyot, 2013, p.11) .

### **2.2 Folksonomies vs traditional taxonomies**

Library of Congress Subject Headings (LCSH) are an example of a traditional formal taxonomy system. LCSH are incorporated within the 650 field of MARC 21 data and subsequently they are used by libraries globally via their information retrieval and management systems. The LCSH form a ‘controlled vocabulary’ “designed to ensure uniformity and universality within and across library catalogues or other information retrieval systems so that locating information is predictable and precise.” (Adler, 2009, 313).

LCSH form part of a ‘syndetic’ system which connects synonyms, variants and related terms using cross-references but such a system is by its structural nature monologic as it relies upon an ‘authority’ to set these terms, the semantic relationships between them and the resultant hierarchies.

This hierarchical structure used in most modern formal taxonomies derives from Charles Cutter’s “Rules for a Printed Dictionary Catalog” (1904) (Wallace, 2007, p.177). Cutter

defined a syndetic system as:-

“...connective, applied to that kind of dictionary catalogue which binds its entries together by means of cross-references so as to form a whole, the references being made from the most comprehensive subject to those of the next lower degree of comprehensiveness and from these to their subordinate subjects, and vice versa”. (Cutter, 1904 , p.23)

In contrast within a folksonomy there is no ‘authority’ deciding terms used, the relationship between them and predetermined hierarchies, instead “you try to find ways that the individual sense-making can roll up to something which is of value in aggregate, but you do it without ontological goal” Shirky (2005). Critically, she says, “the semantics here are in the users, not in the system.” Distinguishing characteristics of this crowd-sourced “consensus of opinion” have been listed by some as being user-orientated; empowering; democratic; cheap; collaborative; distributed; dynamic and instructive (in that the tags can be analysed to reveal things about the taggers) (Furner, 2007, p.20.)

### **2.3 Research into folksonomies**

There is a growing body of research into folksonomies all of which face common issues concerned with the very nature of the tagging system and its dynamic, anarchic nature.

#### **2.3.1 The Zipfian Power Law- frequencies of tags, convergence and emerging consensus**

A central focus within the area of folksonomy research has been the observation of what is known as the Zipfian Power Law- the phenomenon within this context that the more popular tags will appear at a far higher frequency than the less popular tags within a population of tags (Mathes, 2004, p.11; Guy & Tonkin, 2006; Golder & Huberman, 2006; Halpin, Robu & Shepherd, 2007; Guyot, 2013). This means that within a folksonomy there is actually a degree of consensus that forms “given sufficient active users, over time a stable distribution with a limited number of stable tags and a much larger ‘long-tail’ of more idiosyncratic tags develops” (Halpin, Robu & Shepherd, 2007, p.220). This frequency with which more ‘stable’ tags are used is of vital relevance to researchers as Yi explains:-

“tags assigned by more people are believed to reflect the community consensus better than those assigned by fewer people- more frequently chosen tags for a resources are more valuable” (2009, p. 1658).

The ‘long tail’ contains not just ‘messy’ incomprehensible tags as complained of by many researchers (Cantador, Konstas & Joeman, 2011; Thomas, Caudle & Smith, 2010; Lu, Park & Hu, 2010) but also those revealing differing world views (Adler, 2009 ; Bates & Rowley 2009) however a certain number of tags will ultimately remain “ambiguous, overly personalised and inexact”(Guy & Tonkin, 2006).

### **2.3.2 The ‘long tail’**

Researchers wanting to design mapping methods to compare systems or anyone attempting to create ways in which traditional taxonomies and folksonomies could co-exist within a new style of information retrieval system have got to tackle the tricky question of what to do with this ‘long tail’ of infrequently applied idiosyncratic tags. Researchers looking at tag categories rather than individual tags and terms sometimes disregard any tags with a low frequency of application (Guyot, 2013; Golder & Huberman, 2006; Iyer & Bungo, 2011). Tourne & Godoy (2012) attempted an automatic analysis of tags applied to web resources but caused 12% of their tags to be disregarded after running spelling checks- most of these were abbreviated or non-English words. In “Trashy tags: problematic tags in LibraryThing” Thomas, Caudle & Smith list a variety of tag types that are unhelpful in terms of using the folksonomy as an augmentation to a more traditional taxonomy for example variations of existing tags, tags including numbers, misspellings and non-English words.

### **2.4 Mapping LCSH to folksonomy tags**

Heymann & Garcia-Molina (2009) contrasted LCSH and LibraryThing tags. They started with the LCSH then matched them to the tags creating lists of what they called ‘syndetic equivalents’ that matched either exactly or ‘almost exactly’ if the LCSH is “modified to remove ‘parenthetical remarks, swap the ordering of the words around a comma, stem or add or remove an ‘s’” (p.4). Using this method they found that 40% of the LCSH used in their study had ‘syntactic equivalents’. They then further matched the LCSH by using an



algorithm called “Wikipedia Explicit Semantic Analysis (ESA)” to create a list of “semantic equivalent” terms after which they claim “most of the remaining keywords” were matched (p.3). A similar mapping technique was employed by Yi in a study into predicting LCSH for social tags using “a semantic similarity approach” (Yi, 2009, p. 1658). Both studies used the LCSH as the benchmark set of data and used this as a starting point with which to search for the tags to match.

Another study into mapping a folksonomy from De.licio.us to LCSH was carried out by Yi and Chan (2009) who created what they called “LCSH trees” using syndetic terms to which they word-mapped tags. They claimed to have matched two-thirds of the tags that they sampled this way and concluded that “collaborative tagging commonly relies on post-coordination and presents a user-centred view; professional indexing with controlled vocabularies involves pre- or post-coordination and a system-centred view...the linking of the two such resources is valuable in that it can integrate the views of both the users and the systems in indexing and information organization”. (p.897)

## **2.5 Tag categories**

As folksonomies are not syndetic systems research that attempts to map or compare them with formal taxonomies using a sample can run into problems when attempts are made to match tags simply semantically. A more fruitful avenue taken by researchers has been to look instead at ‘categories’ of tags- to identify and explore the new ‘categories’ not previously included in taxonomies, to try to investigate tagging behaviour by grouping terms that reveal a similar ‘warrant’ and also to serve as a kind of ‘filtering’ system to identify those tags that relate conceptually between the systems- particularly the ‘subject’ headings.

### **2.5.1 Studies utilising tag categories**

Golder & Huberman, (2006) researched user activity on Del.icio.us and devised seven main tag categories that they further divided into two distinct sections. The first set of categories they claimed were ‘extrinsic’ to the tagger. These were:-

1. Identifying what the thing is about

2. Identifying what it is
3. Identifying who owns it
4. Refining categories- ones that qualify the categories above

The second group of categories they claim were 'intrinsic' to the tagger:-

5. Identifying qualities or characteristics e.g. 'scary', 'funny'
6. Self-reference e.g. tags that start with 'my stuff'
7. Task organising e.g. collecting information for a job search

Thomas, Caudle & Schmitz adapted Golder & Huberman's model, adding an extra category for foreign language tags- they concluded that "a hybrid catalog combining both LCSH and folksonomies would result in richer metadata and would be stronger than the sum of its parts, giving patrons the best of both worlds." (2009 p. 411) This study disregarded tag frequency and their sampling technique also resulted in some of the booksonomies studied having too few tags in them to provide a stable consensus of opinion (p. 422).

Ke & Chen (2012) proposed twenty-six categories, they observed that the Zipfian Power Law- "not just the tag distribution but the tag category distribution echoed a power law distribution". They, along with Heckner, Muhlbacker & Wolff (2008) observed how the tag categories differed with the objects being described. Heckner, Muhlbacker & Wolff noticed that tags on "Connotea" applied to photographic content tended to fall into categories related to content, location and device name whereas scientific articles were mainly tagged with time and task related tags.

### **2.5.2 'Aboutness' as a tag category concept**

Smith (2007) used Golder & Huberman's model as a basic model with which to develop her own categories but unlike previous studies based on their model she aligned the first two categories with concepts of 'aboutness' and 'isness' drawn from Sara Shatford's research into indexing photographs. Shatford used these concepts to distinguish the difference between what the photos were 'of' (a factual term) and what they were 'about' (a subjective opinion) (1986, p. 42-50). Judith Ranta (1992) drew upon Shatford's two terms when developing a fiction indexing system- she suggested that both 'denotative' or

factual elements (e.g. setting, factual elements of the plot- Shatford's "isness" concept) and 'connotative' or imaginative elements (e.g. the theme- Shatford's 'aboutness' concept) could be used to index works of fiction. Smith chose a very small sample of five books out of a sample of twenty five to focus on because she deemed them "particularly interesting and relevant to the overall subject analysis discussion", she also only used the LibraryThing tag cloud to sample the tags from rather than the whole data set (2007, p.5) so her research was very limited and subjective.

### **2.5.3 The importance of categorising in context**

Iyer & Bungo (2011) compared terms applied to forty popular medical books within MARC records on the OCLC Connections database with tags applied to the same books on LibraryThing. They examined the tags for 'conceptual meaning' and then assigned them to tag categories where there were elements of group commonality as "it is not effective to compare individual tags semantically with subject headings" because they vary widely and frequency alone doesn't provide enough depth of context as other researchers have similarly concluded (Guyot, 2013, p.61, Wichowski, 2009) "This context and understanding of the tags and their meanings thus facilitated the comparison of the subject headings with the tags and allowed the determination, on a deeper level, of the types of semantic relationships represented" (Iyer & Bungo, 2011, p.10-11) - the researchers call this the 'implicit context'.

Iyer & Bungo make no mention of the 'long tail' or what they did with the incomprehensible tags in their sample but the development of the conceptual categorisation system applied with implicit context to each book is a useful method for adopting when investigating what the systems reveal about the user warrants, concepts of 'aboutness' and what this says about their implied searching behaviour. Heymann & Garcia-Molina's (2009) research mapping collection of LCSH with the same tags as applied within LibraryThing both term to term and semantically had issues with this lack of 'implicit context'-although they had a high match of tags and subjects the tags were applied to items that had very little in common due to a lack of context and the individual tags were applied in many different ways.

## **2.6 Tagger motivation and “warrant”**

One of the most exciting areas for researchers has been the access to this relatively new ‘user-centred view’. Tightly controlled, closed traditional taxonomies don’t allow for the ‘users’ of those systems to disagree or add to the description or classification of the elements but with the development of folksonomies and the open access nature of most of these social tags the opinions and behaviour of the ‘users’ are finally observable and measurable but as Furner remarks tagging is “used by multiple groups of people for multiple kinds of function” (2007, p.6).

The motivations of social taggers have been researched by many and have generally been divided into two main categories- firstly one of organisation and the second social-communication and description (Marlow, Naaman, Boyd & Davis, 2006, p.5; Ames & Naaman, 2007; Bartley, 2009; Korner, Grasl, Kem & Strohmaier, 2010; Guyot, 2013). Kipp (2007) looked at non-subject tags and broke these into “affective tags”, “emotional tags” and “time and relationships” she concluded that users have an emotional response to and desire to attach personal information to documents. Guy & Tonkin’s research into personal tag use by users of De.licio.us and Flickr revealed often tags are often serving ‘two masters at once; the personal collection, and the collective collection’ (2006). Morrison goes further to say that designers of a site that wants to develop a ‘useful’ tagging system should take into account the users’ motivational reasons for tagging “a folksonomy” is more likely to be successful when the goals of the website or information system intersect with the goal and motivation of users”. (2007)

### **2.6.1 User warrant**

The various tag category systems outlined above vary not just because of the various folksonomies they are being used to describe but they also exhibit the ‘warrants’ of both the researchers and the users of those tagging systems.

In the late 80s, before the advent of Web 2.0 and social tagging as we now know it, Beghtol wrote about the concept of warrant which she defined as “the authority a classificationist invokes first to justify and subsequently to verify decisions about what classes/concepts to include in the system” (1986, p. 110-111). She identified several basic kinds of what she called semantic warrants- literary, scientific, educational and cultural

which have since been explored and developed by various theorists.

More recently Beghtol wrote that:-

“Cultural warrant means that the personal and professional cultures of information seekers and information workers warrant the establishment of appropriate fields, terms, categories, or classes in a knowledge representation and organization system.” (Beghtol, 2005, p. 2)

A further concept of ‘user warrant’ was not discussed in depth by Beghtol but was defined by The National Information Standards Organisation (NISO, 2003) as “justification for the representation of a concept in a [thesaurus] or the selection of a preferred term because of frequent requests for information on the concept.”(National Information Standards Organization, 2003)

Modern librarians are increasingly seeking to put the ‘user warrant’ at the forefront of their retrieval systems. As Spiteri explains in her advocacy for user-generated metadata “The catalogue is a critical bridge between a library and members of its community...this link is becoming increasingly important in a wired world.” (2012, p.211) Linking user-generated content within OPACs or augmenting formal taxonomies and folksonomies within it e.g. LibraryThing for Libraries (Librarything.com, n.d.) is one way of maintaining this bridge between the users and the ‘authority’. Conversely by looking at the booksonomies already in use by our users the information professional may be able to identify strong ‘user’ or indeed ‘cultural’ warrants amongst taggers of particular subsets of books.

## **2.7 Research into booksonomies**

Much of the research on folksonomies focuses on ways in which they compare to more traditional taxonomies by developing their various mapping techniques as we have discovered. Early studies focussed on the first sites to use social tagging De.licio.us and Flickr (e.g. Overall, Sigubisson & Van, 2009; Suchanek, 2009) and as booksonomies such as those found on LibraryThing have developed there have been a number of studies looking into how these systems compare to LCSH as it is possible to compare data used to describe the same books (Heymann & Garcia-Molina, 2009; Lawson, 2009; Lu,

Park & Hu, 2010; Adler, 2009; Iyer & Bungo, 2011) LibraryThing has over 2 million users and tags applied within the booksonomies are openly available for a researcher to access.

In 2009 Bartley sent a sample of 98 self-selecting users of LibraryThing questionnaires asking about their motivation for using the site and 74% said that their motivation was 'collection management'. This may hardly be surprising however when 46% of those filling in the surveys worked in libraries or another information profession and 28% of those who answered had undertaken professional training in cataloguing! Melissa Adler in (2009) researched tags applied to fifty books that had been on at least one of the American Library Association's Gay, Lesbian, Bisexual, Transgendered Round Table's (GLBTRT) lists of recommended GLBT books, was included in a minimum of fifty LibraryThing member catalogues to get enough of a sample of tags and had one at least one book award or special mention. She then compared the tags applied within the folksonomy to their LCSHs. She concluded that firstly that there was a degree of consensus evident and secondly that the 'range of expressions of minority voices is highly visible and negotiable' (Adler, 2009, p. 18).

## **2.8 Picture book research**

Despite the many studies that have been undertaken into LibraryThing booksonomies and particular communities of taggers no one has yet looked at children's books or picture book booksonomies nor the behaviour of the taggers who describe them. The 'picture book' according to picture book researcher Perry Nodelman is "the one form of literature designed specifically for children" and remains "firmly connected to the idea of an implied child-reader/viewer" (1998, p. 11). Hunt argues that the picture book has a direct and indirect influence on "most adults and almost certainly the vast majority in positions of power and influence" (1998, p.1) and yet "the books...have been marginalised". Hunt suggests that the picture book is "overtly important educationally and commercially- with consequences across the culture from language to politics" (Hunt, 1998, p. 1)

Unusually for a booksonomy therefore the picture books being tagged on LibraryThing will not, we must assume, be being tagged by the main target audience for the books. Children don't write, publish or on the whole buy picture books yet they are the audience for the genre. The picture books tagged on LibraryThing will instead be tagged by the adult readers and rather more accurately in some cases perhaps 'users' of the books. In Hunt's introduction to "Understanding Children's Literature" he observes that:-

"children's literature is seen as the last repository of the *dulcis et utile* [sweet and useful] philosophy...children's books are used for different purposes at different times- for more things than most books are...some are 'good' time passers; others 'good' for acquiring literacy; others 'good' for expanding the imagination or 'good' for inculcating general (or specific) social attitudes, or 'good' for dealing with issues or coping with problems, or 'good' for reading in that 'literary' way which is a small part of adult culture, or 'good' for dealing with racism...and most books do several things" (Hunt, 1998, p. 11)

Hunt also points out that "children's books are part of the ideological structures of the cultures of the world" (1998, p.5) when even the definition of 'children' and 'childhood' changes from time to time place to place. This 'use' function applied to fiction titles is unusual and hasn't been explicitly mentioned in other booksomy research. Preliminary research into the 'task-based' categories used by researchers tends to refer to very personal acts e.g. 'to read' whereas the 'use' category, if it exists within the picture book booksonomy as Hunt's describes, it would be a social or extrinsic motivation.

In her essay "Picturebooks and changing values at the turn of the century" Colomer analysis the content of a selection of picture book published in the 1970s and those published in 2000 she notes that there is (in 2000) "particular emphasis on topics that deal explicitly with emotional education" and also on an education "focused on the more complex values of multicultural coexistence in society." (Colomer, 2010, p.41)

She puts this shift down to the changing economic and cultural demands of Western post-industrial society and the inevitable ways in which this effects the social construction of childhood, the socialisation process and formal and informal pedagogy and describes it as a shift towards the "social representation of childhood". (p.48)

## 2.9 Summary

The study of folksonomies is a growing area of research. Focus has tended to be on how formal taxonomies are reflected in the folksonomies- by mapping LCSH and tags for example but little research has been done into what the tags themselves display in terms of user and cultural warrant in a practical sense. Any research undertaken into folksonomies must take into account the much-documented issues with sample size, whether the tags are at a 'stable' state (by displaying the Zipfian Power Law in distribution) and the 'long tail'.

Category models used to create tag categories tend towards a certain rigid perception of 'subject' being factual and 'use' being purely personal, they have also on the whole been carried out in the abstract sense, divorced from the context of the books that they have been tagging. If the picture book booksonomies reflect the 'dulcis et utile' philosophy that Hunt writes about then a category system based on the trends within these tags themselves should portray this. A categorisation system based on the user warrant displayed in the patterns within tag categories based on 'implicit context' would best reflect the intricacies of this very particular group of booksonomies and the tagging behaviour applied to them. Mapping the 'subject' category tags and semantic tag groups to the relevant LCSH (and not the other way around) would also reveal any 'emergent' subject groups of tags, perhaps displaying 'user warrant' in the suggestion of new LCSH in progress.

Iyer & Bungo (2011) focussed on the difference between the way that the 'public' and non-professional users tag popular medical books and compared them with the cataloguers with their 'professional knowledge' and 'literary warrant'. Conversely for this research the 'professional users' of the book sample are more likely to be the 'public' users 'using' them with the children in their care or influence as opposed to the 'professional cataloguer' who may or may not have any experience of these books other than reading them when they were a child. Using 'aboutness' as a subject concept rather than a factual interpretation may help reveal the difference in these two 'user warrants', especially when the tags are matched against the LCSH and those not matched are investigated.



The existing research into booksonomies within LibraryThing hasn't yet investigated the interesting area of picture books and likewise modern research into picture books has yet to research booksonomies as a way of investigating this 'marginalised' yet politically and culturally powerful and influential publishing medium' (Hunt, 1998, p.1) and its position within society and our formal and informal curricula. If, as in other research, there is evidence of new emergent 'subject' tags then these too might be tracked over time to monitor changes in the 'values' held by society and the descriptors that they apply to picture books.

### **3. Methodology**

#### **3.1 Introduction**

This chapter provides an overview of the research strategy and methodology used within this study. It is divided into three sections as the research employed a three stage process- data collection, categorising the tags and mapping the ‘subject’ category tags to the LCSH.

#### **3.2 Choice of research method**

A literature review was performed into social tagging, folksonomies, tag categorisation, tagger motivation and ‘warrant’ theory. Separate searches for studies containing the keywords “folksonomies”, “booksonomies” and “picture books” were undertaken in the first instance using Proquest’s Library and Information Science Abstracts (LISA) and JISC’s Zetoc database with alerts set up to report new research on these topics over the course of the study. The literature review highlighted a lack of research into picture books in general and an absence of any research into picture book booksonomies in particular. Research into folksonomies is robust and wide ranging however with a growing number of studies into LibraryThing booksonomies in particular.

The literature review led to a quantitative research method being designed to investigate picture book folksonomies (booksonomies) within the online social cataloguing site LibraryThing. LibraryThing.com was chosen as the source of the booksonomy data as it’s the largest example of a book folksonomy available- there are just over 2 million members at the time this research and contains user generated data on 104,620,584 individual books. LibraryThing has been studied in many previous studies as the literature review revealed and the concept of matching Library of Congress Subject Headings with LibraryThing tags has also been the focus of several studies (Heymann & Garcia-Molina (2009); Yi (2009); Lu, Park & Hu (2010).

The aim of this study was to investigate the picture book booksonomies so a purposive sampling technique was devised to select ten books whose tags would be harvested along with their frequencies. Each book’s individual booksonomy was analysed on an

individual contextual basis and the tags sorted into categories within a model influenced by the research studied during the literature review. The data was then analysed at both an individual book level and at an amalgamated level across all ten books. The ‘subject’ tag category was then investigated further by comparing them to Library of Congress Subject Headings using a mapping technique influenced by previous researchers.

### **3.3 Data collection**

#### **3.3.1 Selecting the sample**

Other researchers choosing samples of books from LibraryThing have tended to use “most reviewed” lists on the Zeitgeist page as a source of data sets. Yi (2009) selected 23% titles at random from the top 500 most reviewed and Lapsa (2013) took a sample from the most popular (looking at ratings) and those most recommended. For this study it was important to choose a sample that reflected best what the picture book describing community on LibraryThing was tagging most frequently. As a sampling technique it was decided to search for the books that had been tagged with the tag ‘picture book’ the most number of times to get a sample of books most perceived to be ‘picture books’. ‘Picture books’ as a term can be used not just as a genre but also as a medium (Nodelman, 1998, p.11) and could be applied to anything from Shaun Tan’s moving tale of immigrants arriving in a new country “The Arrival” to the board book “The Hungry Caterpillar” which is aimed at preschool children.

As folksonomies don’t have a controlled vocabulary LibraryThing suggests creating a more stable term by providing a list of variants when you use the search option. There were sixty-six variants of “picture book” as defined by a pre-defined LibraryThing tag combination, these included “picturebook”; “picture books”; misspellings & foreign translations of the word (see Figure 3.1 below). This secured a sample of records with a high number of tags for analysis and the ones most relevant to the concept of “picture books” within this community.

## Tag: picture book

Includes: picture book, picture books, 絵本, pic book, bilderböcker, billedbok, bildebok, pictue book, pictue book, billedbøker, pictue book, picture bok, (picture book), bildebøker, pitue book, pictuer book, #picturebook, picture book, pitcure book, picture boook, picturbook, picutrebook, pcture book, pictur book, picture book, picture book, picturebook, picture book, pocture book, picture-book, prentenboek, pictuure book, picture boook, picturebook, picturebook, picture book, picture ook, picturee book, picture book, pictrebook, pic. book, picturebookk, picture book, picturebook", picture book, ppicture book, picturebok, picturebook, Picture BOok, Picture bOoks, Bilderbuch, PicBook, Bildebøker, Picturbook, Picture-books, Picture book, Bildebok, PICTURE BOOK, Pictuer book, Picturebook, Billedbøker, Picture Book, Picture Books, Pictue Book, Picture Book, g:Picture Book, . picture book, PICTURE BOOK, pICTURE bOOK, Picture\_book, Picture Boook, PICTUREBOOK, Picture Boo, # picture book, Picture boo, bilderbuch, Picture ook, Piture Book, Pictur Book, picture boo, Picture books, f:picture book, Picture Book, Pictue books, PICTURE BOOKS, Pictue Books, picture book, Ppicture book, Pictue book, Picture book, Picture BookS, picture book, picture Books, Pictuebook, Pictue Book, Pic book, Picture Bok, Picture Book, Bilderböcker, PictureBook, PictureBooks, Picture book., Picturebok, Pictire book, picture Book, g:picture book, Bilderbuch, Pictue book, Picturebooks, Picture-Books, #PictureBook, Picture\_Book, Picture-Book, Picture book, Pictire Book, Pictue book, Pictue Book, Picture Book, PicBook, Pcture book, picture book., Picture book, format - picture book, Picture bok, Pictrebook, picture-books, Picture Book, Picturebookk, Picture Book, Picture boook, picture book..., Picture Book, picture book;, ^picture book, Pictue book, Picture Book, Pic Book, picbook, (Picture Book), picture book\_\_, picturebook, Picture Book., Picture book;, Picture Book, Pictuer Book, pictureBook, Picture Book, Picture Book, (Picture book), Pictur book, Prentenboek, Picture-book, PICTURE book, picture bk., PICTURE BOOK., picture\_book, Billedbok, Pictrebook, PictureBok, ^Picture Book, picture book, Picbook, Picture Book, Pictre Book, picturebooks (what?)


Tag and its aliases used 380,578 times by 10,992 members. 

Figure 3.1 screenshot of ‘picture book’ tags and its aliases applied within LibraryThing

The sample was chosen by selecting the ten books most frequently tagged with the tag of “picture book” or one of its aliases on 19<sup>th</sup> July 2015. This sample method resulted in a list of books generally aimed at young primary school age children so in this study the term “picture book” can accurately be used to refer to a genre.

### Most often tagged picture book

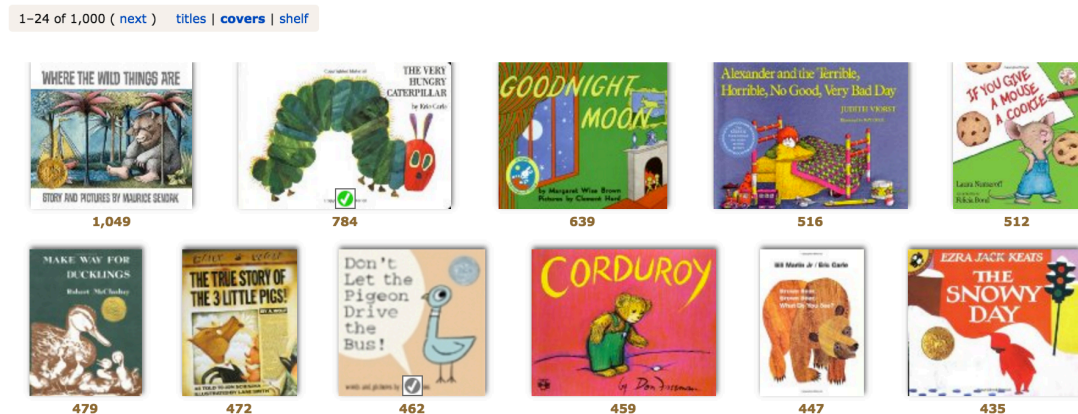


Figure 3.2 Screenshot of books sampled

Title	Author	Illustrator	Pub. Date	Pic. B Tag (Fr)
Where The Wild Things Are	Maurice Sendak	Maurice Sendak	1963	1,049
The Very Hungry Caterpillar	Eric Carle	Eric Carle	1969	784
Goodnight Moon	Margaret Wise Brown	Clement Hurd	1947	639
Alexander & the Terrible, Horrible, No Good, Very Bad Day	Judith Viorst	Ray Cruz	1972	516
If You Give A Mouse A Cookie	Laura Numeroff	Felicia Bond	1985	512
Make Way for Ducklings	Robert McCloskey	Robert McCloskey	1941	479
The True Story of the Three Little Pigs	Jon Scieszka	Jon Scieszka & Lane Smith	1989	472
Don't Let The Pigeon Drive The Bus!	Mo Willems	Mo Willems	2003	462
Corderoy	Don Freeman	Don Freeman	1968	459
Brown Bear, Brown Bear, What Do You See?	Bill Martin Jr.	Eric Carle	1967	447

Table 3.3 Sample books ranked by frequency of “picture book” tag

As some of the book titles within this sample are long they have shortened when referred to within this study. All of the books in the sample were published in the USA apart from “Alexander” which is Australian- this probably due in large part to the fact that LibraryThing users are overwhelmingly from the USA. The most recently published book on the list is “Pigeon” by Mo Willems so there is a heavy bias within this sample towards the ‘classic, traditional’ end of the picture book market. The books in this sample range from the minimalist board book “Brown Bear” to the short story-length “Corduroy”. “Pigeon” and “Pigs” are more non-traditional narratives, “Pigs” is a post-modern retelling of ‘The Three Little Pigs’ from the wolf’s point of view for audiences who understand the original well enough to get all of the jokes and references. “Pigeon” is written as a direct first person dialogue with ‘the audience’ where he tries to persuade them to allow him to drive a bus using lots of funny arguments- prompts within the text encourage and imply the children’s negative responses.



Figure 3.4 Sample of a page from “Pigeon”

The diverse variation in form, intended audience, use of language, narrative style and content was wide enough within this small sample of ten picture books for a small research project to provide a representative sample of the ‘picture book’ genre.

All ten books were read and kept available throughout the categorisation process so that they could be analysed in context. Most of the picture books in the sample were very short- containing only a couple of hundred words at most- yet had thousands of tags applied to them so it was clear that it was very important to have contextual knowledge of

what was implicit in the text, pictures or actual lexical space of the book (denotative) and what was perceived to be “in the book” (connotative) (Ransley, 1987, p. 20). Unlike previous research into booksonomies the short length of the books in the sample meant that each book could be read and easily referred to for clarity throughout the process although as only one person undertook the categorisation this was still very subjective.

### 3.3.2. LCSH collection

On the same day as the sample group was selected from LibraryThing the LSCH from the ten books were collected from OCLC’s WorldCat online catalogue. LCSH were chosen as the example of the subject headings to represent a formal taxonomy because it is the standard vocabulary used by the majority of libraries world wide and also the largest general indexing vocabulary in the English language (Yi, 2009, p.1659). WorldCat was chosen as the source of the LCSH as it contains the bibliographic data from over 10,000 libraries globally. The headings were taken purely from the 650 field in the Marc 21 code and any LC subject headings for children’s literature were included although none of these differed to the main headings in practice.

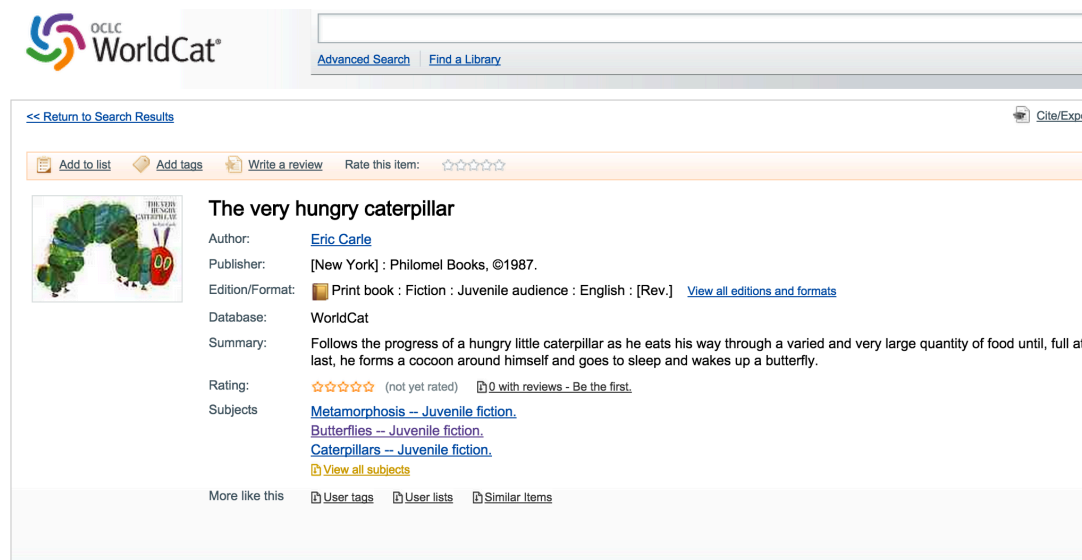


Figure 3.5 Screenshot of WorldCat entry for “The Very Hungry Caterpillar”

### 3.3.3 Harvesting and cleaning the data

The tags were imported using an adapted version of Guyot’s method of extracting tags and frequencies from LibraryThing (Guyot, 2013, p. 23-26). First the tags were copied and pasted (maintaining their HTML format) then they were pasted into Microsoft Word.



Figure 3.6 Screenshot of LibraryThing tags for “The Very Hungry Caterpillar” on the sample day

Guyot’s method of using the ‘Replace All’ function was then used to replace all closing parentheses with a paragraph using the special string character ‘^p’ to leave the data as shown in Figure 3.7 below.

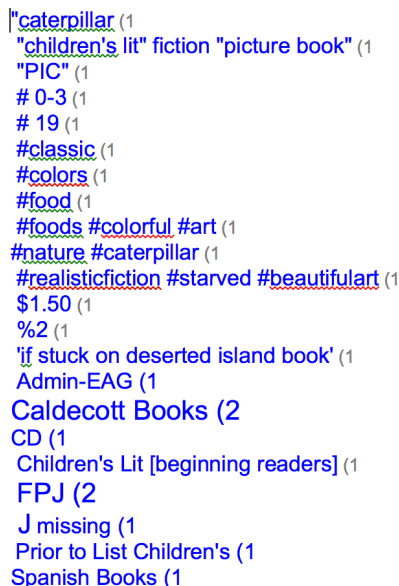


Figure 3.7 Screenshot of raw data being cleaned within Word

Any tags that had included parenthesis within the tag itself (rather than just belonging to the frequency information) had to be manually cleaned up by exchanging a dash for the parenthesis because of a stage further on. Likewise any ‘tagmashes’ where people had put more than one tag and sometimes even a set frequency together were removed e.g. #foods #colourful #art. The data was then copied and pasted into a Microsoft Excel spreadsheet and using a wizard under the ‘set to columns’ function everything after the opening parenthesis was put in a second column thus maintaining the frequency values for each individual tag. The ‘messy’ tags were not removed at this point as the decision was made to categorise two booksonomies entirely to examine the tag frequency distribution before deciding what to do with the ‘long tail’.

### **3.4 Categorising the tags**

#### **3.4.1 The development of the category model**

Having researched various different approaches to tag categorisation a model similar to that employed by Golder & Huberman (2006) was examined to begin with- one based upon ‘extrinsic’ and ‘intrinsic’ warrants as this seemed closest to providing the categories that were needed to be distinguished in order to meet the aims and objectives of this study. Deeper consideration of these categories and the way that researchers have adapted them uncovered some issues in terms of the picture book booksonomies however.

Thomas, Caudle & Schmitz (2009) for example used Golder & Huberman’s categories to examine booksonomies and it is clear that there is no place for the kind of ‘*use*’ category that isn’t ‘*personal task-based*’ within their first two ‘subject’ categories. Describing ‘subject’ categories as ‘factually’ based has inherent issues when referring to fictional or artistic work. Shatford, when researching photographs aligned the ‘subject’ of the photos with the ‘subjective’ rather than the factual (1986, p. 42-50) which is echoed by Ranta’s literal versus figurative levels within her fiction indexing system (1992, p.5) because some of the ‘subject’ tags will be abstract as well as concrete. The ‘subject’ of the picture book is not necessarily ‘extrinsic’ to the tagger, the perception of the ‘subject’ relies on personal interpretation.

An initial list of categories was therefore created separated into two broad sections based on Shatford’s ‘aboutness’ or subjective category and ‘isness’ or factual category



(Shatford, 1986). It meant that tags related to ownership, location, and physical state of the book for example ('Isness') could be separated from those relating to the comprehension of 'aboutness' which contains both '*subject*' and '*use*' that isn't personal task-based but implies a wider application. This distinction is however somewhat blurred when it comes to the '*literary*' category as this includes descriptions of the language and literary techniques within the text but it could be argued that this is more similar to a subjective tag than a factual one. After investigation many of these tags implied a 'reading' of the book rather than just a description of the physical object so it was felt to be more suitable for the '*literary*' category to fit within the 'Aboutness' group.

Aboutness				
Sub.Theme		Sub.Pedagogical use		(Is this use?)
Nouns	Verbs	Specific 'Use'	Lit Analysis/Literacy	Socialisation
dog	running	vet topic	rhyming words	happiness
			phonemes	emotional literacy
				emotions

Table 3.8 Screenshot of initial categories within the 'Aboutness' group

Table 3.8 shows the initial category ideas at the development stage with some fake examples. 'Happiness' eventually would have ended up either in the final '*subject*' or '*affect*' category (if there was a contextual semantic meaning that the book would 'cause' happiness rather than it being about 'happiness'). The 'pedagogical' categories evolved throughout the categorisation process but at this initial stage '*specific use*' was applied as more general term.

At this stage nouns and verbs were separated but there were few verbs in the final results and as these two categories were better summarised as '*subject*' meaning direct narrative or visual content they were merged at a late stage. Interestingly '*phonemes*' and similar terms usually used by teachers teaching children to read via the phonics method were common in the samples. A decision was made to include these in the '*literary*' category

rather than ‘Use: Pedagogical: Formal’ because strictly phonemes come from an analysis of the lexical space of the book even though it would be hard to write anything but a wordless picture book without them!

<b>Isness</b>					
<b>Medium</b>	<b>Illustrator</b>	<b>Author</b>	<b>Format- ext</b>	<b>Location</b>	<b>Physical state</b>
picture book	name	name	board book	F JUN	Falling apart
			with toy	Red box	Torn
				Classroom	New

<b>Audience</b>	<b>Ownership</b>	<b>Visual element?</b>	<b>Awards/List</b>	<b>Language</b>
kids	Elliott's	Collage	Greenaway	French
family		Photos	Top 100	In Spanish
Yr 1		Bright colours		

Table 3.9 Screenshot of initial categories within the ‘Isness’ group

Some of the categories within the ‘Isness’ group reveal an implied ‘use’ of the picture book as well- especially those naming classrooms or ‘teaching collections’ but as these tags were interpreted to relate to the physical book rather than the content of the book they were not included in the analysis of categories implying ‘use’.

<b>Personal</b>	
<b>Value judg</b>	<b>Task- not ed</b>
funny	read
classic	

Table 3.10 Screenshot of initial categories within the ‘Personal’ group

Aside from the ‘Aboutness’ and ‘Isness’ groups of categories there were two further groups- ‘Personal’ including value judgements (including ‘classic’) and task-based tags if they were obviously personal e.g. ‘to read’ would be ‘personal task based’ whereas ‘use to teach alliteration’ would be put in ‘Use: Pedagogical: Formal’ in the final version.

<b>Non categories</b>	
<b>Description</b>	<b>Unknown/Undecipherable</b>
A dog went for a walk and then	
Book about a pigeon who	

Table 3.11 Screenshot of initial category spread within ‘isness’ group

The final grouping of categories were those ‘trashy’ or non-categories including long descriptions of the narrative and anything undecipherable. Foreign language terms were looked up using an online translating site and were included in the relevant categories rather than being put in the ‘undecipherable’ category.

### 3.4.2 Categorisation in context

Using the wider concept of “aboutness” as a subsection of categories rather than the narrower, denotative “subject” concept the categories needed to be assigned in context to each individual book keeping the the “user” warrant in mind – what Iyer & Bungo (2011) refer to as ‘implicit context’. For example “Alexander” is specifically about having a difficult and emotionally stressful day and getting pulled up by adults about bad behaviour and even the LSCH reflected this in having ‘attitude’ as their only heading. As the emotional issues were implicit in the text these and tags such as ‘moods’, ‘feelings’ and ‘behaviour’ were categorised as ‘subjects’ within the ‘Aboutness’ group. “If You Give a Mouse a Cookie” isn’t implicitly about manners at all, there’s little emotional depth to it and no one tells the mouse off for his behaviour within the story. ‘Manners’ are not an explicit subject within the text or the pictures of this book but the tags applied reveal that some readers and users of the book interpret or imply that the book is about having manners or a lack of them therefore the ‘manners’ tag in that instance was put in

the 'Use: Pedagogical: Socialisation' category because of the specific context of that application to the whole book. Heckner, Muhlbacker & Wolf (2008) researched tags on 'Connotea' that they concluded "considerably add to the lexical space of the target resource" and with such short books and such a wealth of different tags applied to them within the folksonomy the study of these picture books tags reveals a similar phenomena.

As the data was still in an HTML format the tags were maintained as hotlinks to the tag as applied within LibraryThing- this meant that the 'context' of the tag applied could also be checked when there was some question over meaning. Many of the tags that would have otherwise have been undecipherable were understood using this function e.g. FIAR relating to an American Christian home schooling curriculum resource called "Five In a Row". Searching for FIAR via Google doesn't explain the tag but by investigating the tag using the hotlink it is clearly being used by many different taggers so by searching "FIAR books" into Google the tag meaning was revealed.

### **3.4.3 Trial book & modification of the categories**

In order to test these initial categories "Goodnight Moon" was randomly selected as a trial book for the categorisation process. After each of the 2,404 individual tags had been sorted into one of the categories a new column was calculated using the frequency value to obtain a total number of tags (Fr) per category and within the sections. Lists of particular tags sorted by category were created by copying and pasting data into another worksheet and applying filters and the categories were then examined for issues. '*Calming*', '*lulling*' and '*soothing*' were three tags that were initially put into the '*use*' category but as this seemed to imply using the picture book to cause 'an affect' a new category was created and the '*use*' category was expanded to separate the different kinds of 'use'.

Investigation of these first tags also revealed a number of tags referring to graded reading schemes so a list was drawn up of the acronyms for these and if there was a significant number of tags claiming that a book was '*AR 3.7*' for example (Accelerated Reading score) then any '*3.7*' tags were interpreted as relating to the same thing.

		Aboutness	
		Sub.Theme	
Tag Name	Frequency	Nouns	Verbs

Sub.Pedagogical use				
Specific 'Use'	Lit Analysis/Literacy	Socialisation	AR/recog level	Affect

Table 3.12 Screenshot of final categories within the ‘Aboutness’ group

The final categories which were then used to sort the tags for the final eight books were further modified at the data analysis stage. ‘*Specific Use*’ was further separated into formal and informal pedagogical use and ‘*named pedagogical product or programme*’ and ‘*nouns*’ and ‘*verbs*’ were combined into ‘*subject*’. See Appendix A for the final definitions of the “Aboutness” group categories.

### 3.4.4 Second trial book

To test that the first set of results and categories wasn’t only relevant to the pared back, basic style “Goodnight Moon” which is for the very young “Pigs” was chosen as a second trial title as it’s a more recently published book occupying a different appeal and lexical space. Both of these sets of tags were categorised entirely so that there were two complete examples of data sets including the ‘long tail’ of tags applied infrequently to compare.

No category changes needed to be made after the second trial book data had been categorised so the data was then transferred to individual spread sheets where the tags were extracted into lists by category and then analysed to calculate the percentage

distribution between categories and within the smaller ‘main body of the tags’ (tags applied by a frequency of two or more) and including the ‘long tail’.

### 3.4.5 Selecting the sample range- whether or not to include the long tail

As figures 3.13 and 3.14 below show the percentage distribution between the overall frequency of ‘long tail’ tags as this research has defined it and the main body tags are fairly consistently distributed if you compare the two booksonomies. The data sample and distribution of tag frequencies also obey the Zipfian Power Law as has been observed in other research (Mathes, 2004, p.11; Guy & Tonkin, 2006; Golder & Huberman, 2006; Guyot, 2013).

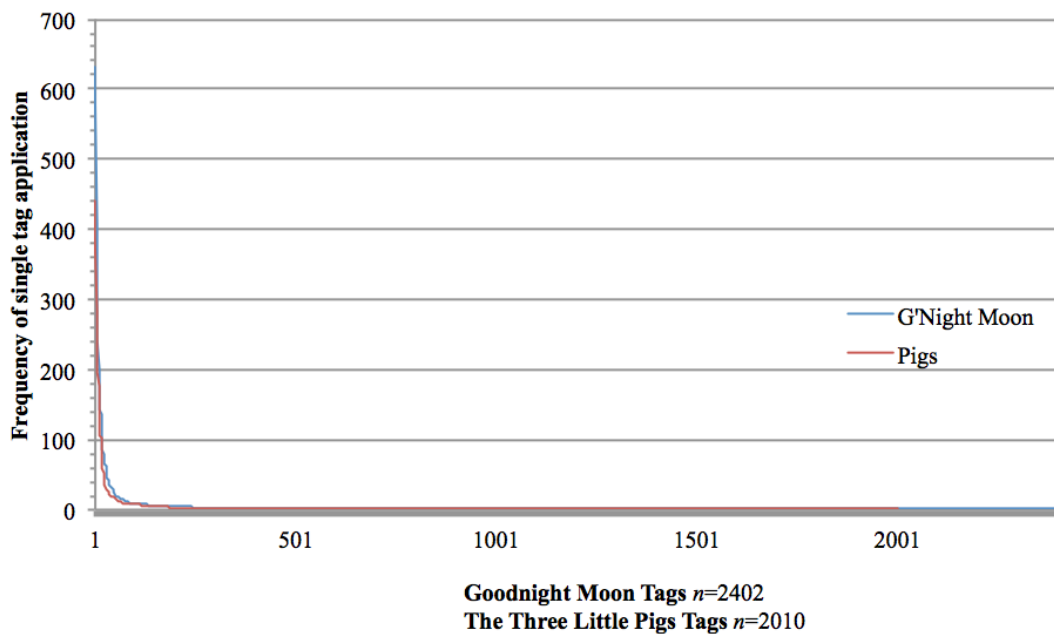


Figure 3.13 Showing the Zipfian Power Law curve in the complete data sets for “Goodnight Moon” and “Pigs”

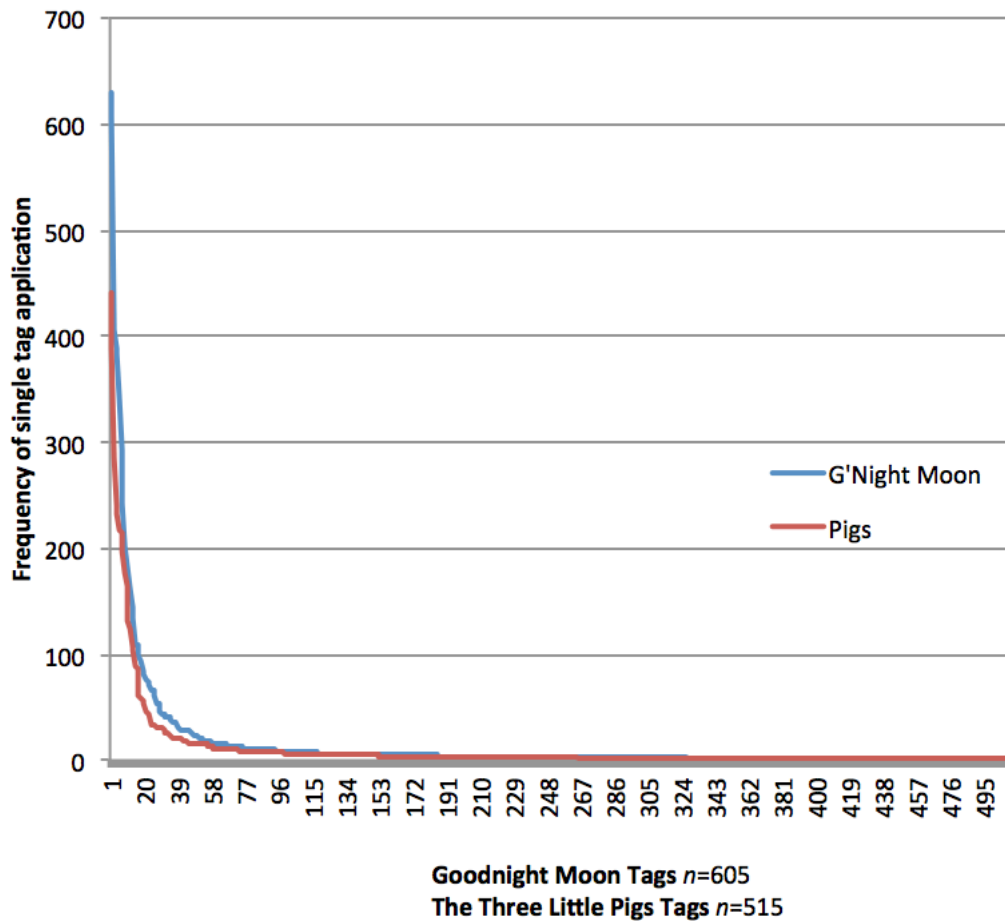


Figure 3.14 Showing the Zipfian Power Law curve in the selected sample range (tags applied twice or more) for “Goodnight Moon” and “Pigs”

4,412 tags were sorted into categories during this trial stage and stacked column charts were produced within Excel to display the comparative overall application of a tag category between the long and shorter samples to investigate whether removing the ‘long tail’ would significantly affect the research proposal- this is discussed further in the results and discussion chapters.

<b>Pigs</b>	<b>Subject</b>	<b>Use</b>	<b>Lit</b>	<b>Social</b>
% Share of tags in this category overall	19.91%	4.23%	8.96%	0.63%
% Share of tags in this category in the long tail only	8.23%	9.30%	8.63%	2.61%
<b>Goodnight Moon</b>				
% Share of tags in this category overall	69.55%	12.37%	15.41%	2.67%
% Share of tags in this category in the long tail only	36.47%	32.16%	16.47%	14.90%

Table 3.15 Comparing the whole sample data across the trial books and the ‘aboutness’ categories

### 3.4.6 Issues with tag categorisation

The main issue with applying any system of categorisation is that there will always be tags that don't fit easily within the 'system' that you create. The process in this research was only carried out by one person so such decisions were very subjective. Examples of how particularly difficult tags were interpreted are outlined in Appendix B.

### 3.4.7 Methods of analysis: tag categories

Microsoft Excel spread sheets were used to create individual work books for each booksonomy then the data for the whole sample was collected in a central spread sheet before being analysed, results were then presented in the most appropriate chart forms.

Word clouds were created using Wordle.net to give an insight into the individual tags within a selection of categories. Subject tags were combined with their frequencies using the following algorithm to produce a list compatible with Wordle's Java script programme so that the word clouds would visually represent the frequencies of tags applied in the size of font used:-

=CONCATENATE(A2," : ",B2)

However there was such a range in frequency of tags within the lists that the first word cloud produced for the 'subject' category for "Hungry Caterpillar" contained too many tiny, illegible words. To solve this issue the following formula was applied to the weighted data set in order to reduce the range of word sizes whilst maintaining the ratio:-

=CONCATENATE(A2," : ",10\*log(B2))

=CONCATENATE(A2," : ",B2)			
A	B	C	D
<b>Hungry Cat. Subject Tags</b>	<b>Frequency</b>	<b>Prepared for Wordle</b>	<b>Reduced range for clarity</b>
<a href="#">butterflies</a>	653	butterflies : 653	butterflies : 28.1491318127507
<a href="#">counting</a>	480	counting : 480	counting : 26.8124123737559
<a href="#">caterpillar</a>	474	caterpillar : 474	caterpillar : 26.7577834167409
<a href="#">food</a>	435	food : 435	food : 26.3848925695464
<a href="#">insects</a>	379	insects : 379	insects : 25.7863920996807

Table 3.18 Showing a sample of 'subject' tags prepared for Wordle.net



### 3.5 Mapping the ‘subject’ category tags to the LCSHs

The mapping process was developed to create a syndetic model with which to match the tags and was an adaptation of the ‘LCSH tree’ model used by Yi (2009) and Yi & Chan (2009) and that used by Heymann & Garcia-Molina (2009) in that there was a three stage method. See Appendix C for details of the three stages with examples.

To ensure that the concept represented by the LCSH was being correctly interpreted in order to create the list of matches the Library of Congress website was used to check definitions when there was a lack of clarity in their meaning. One such example of a more esoteric LCSH was the ‘Conduct of Life’ heading given to “Wild Things”, a heading which was created in 1999 and is given the variants ‘ethical behavior’, ‘deportment’ and ‘behavior, ethical’ on the Library of Congress website (Library of Congress, 2015).

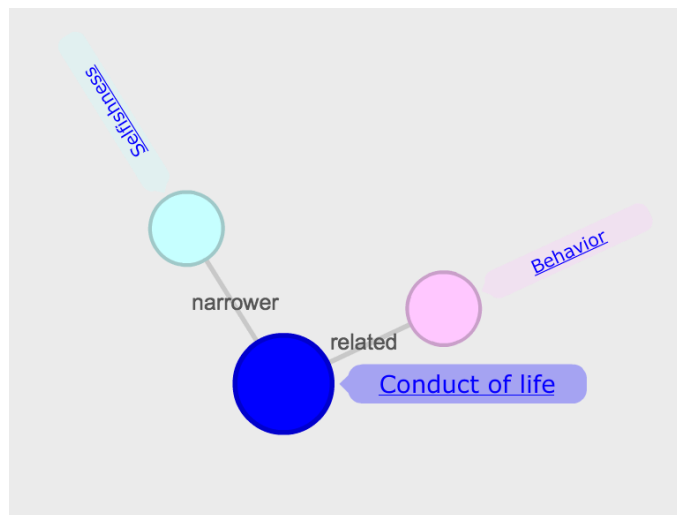


Figure 3.16 Visualisation of LCSH ‘Conduct of Life’ (Library of Congress, n.d.)

Throughout the matching process a tag and its frequencies was used twice if it matched to more than one of the LCSH therefore the final sum used to calculate the results included these duplicate applications.

1st Match		2nd Match		3rd Match	
LCSH	No. of tags (Fr)	Plurals, variations etc	No. of tags (Fr)	Basic level variation & related terms	No. of tags (Fr)
<u>metamorphosis</u>	173	<u>Metamorphosis--Fiction</u>	2	<u>transformation</u>	38
		<u>metamorphois</u>	2	<u>change</u>	32
		<u>metamorphism</u>	2	<u>growth</u>	31
				<u>growing</u>	27
				<u>changing</u>	12
				<u>ThingsThatChange</u>	6
				<u>transformations</u>	5
				<u>development</u>	4
				<u>evolution</u>	4
				<u>Caterpillars transforming into butterflies after</u>	2
				<u>transformatie ('transformation' in French)</u>	3
				<u>changes</u>	2
				<u>groeien ('grow' in Dutch)</u>	4

Table 3.17 Showing the three level matching process for LCSH ‘*metamorphosis*’ applied to The Very Hungry Caterpillar

### 3.5.1 Methods of analysis: LCSH & ‘subject’ tag matching

A Microsoft Excel spread sheet was created with a different sheet for the data set from each book. The tags were then mapped onto the three levels and produced as lists with frequencies attached (see Table 3.17 above). Total percentages were produced for the number of tags matched at each level of the model and these were used to produce charts and tables in styles most appropriate to clear visual analysis. Where groups of the non-matched tags shared a semantic similarity they were rearranged into tables and their frequencies calculated to produce the suggested new user generated subject headings discussed in the discussion chapter- extra examples of these tag groups are provided in the appendix.

### 3.6 Methods chapter summary and limitations

Ten picture books were chosen from LibraryThing using a purposive sampling process, their tag data and frequency values were then collected. The relative LCSH headings from the ten sample books were collected on the same day from WorldCat online.

#### 3.6.1 Issues with the sample

Ten books is a small sample to draw data from compared to other booksonomy research e.g. Guyot (2013) or Yi (2009) but considering the time-consuming contextual analysis of each tag within each booksonomy required for the categorisation process it was

considered to be sufficient for this research project. Tag context analysis arose both from the entire picture book text and illustrations but also sometimes from the LibraryThing tags within the context of their application within the booksonomy.

Trial categorisations of the tags from two books showed evidence of the Zipfian Power Law within the distribution of tag application meaning that the booksonomies showed evidence of a stable distribution and a significant degree of consensus forming within them. Sampling techniques that don't take this into account can encounter issues with low numbers of relevant tags (Thomas, Caudle & Schmitz, 2009, p. 42). The research is limited in scope by the size of the data sets within each booksonomy and the small sample of picture books titles chosen.

Deeper analysis of the percentage ratio of each category within the subsection of the 'long tail' compared to the overall percentage ratio showed that there was a significant number of tags that came under the "use" category within the 'long tail' that was discarded. The figures amounted to 32% share within the 'long tail' subsection for the "Goodnight Moon" booksonomy and 9.3% in that of the "Pigs". Closer inspection of these tags with only one application revealed that the majority had more in common with the 'description' tag than with those in the 'use' one as they tended to be sentences describing what use taggers were going to put the books to.

Tag Name
and have the students compare the two in a venn diagram
3 little pigs unit
6-Trait Writing
a fun book to discuss point of view
abstract thinking
and he has quite a story to tell that just might prove his innocence. Writing Trait: Organization
and order of events
and see what they come up with to share!
and sequencing.
and telling their tales from one of these characters' voices. <a href="http://www.writingfix.com/PictureBookPrompts/True_St">http://www.writingfix.com/PictureBookPrompts/True_St</a>
awesome for a picture walk and prediction

Table 3.19 Listing some of the 139 'use' tags in the "Pigs" long tail sample

Analysing all ten booksonomies entirely would have been a sample too large for this current study so for the purposes of this work the scope was narrowed to those tags applied more than once. If further research was undertaken into the ‘use’ categories within picture book booksonomies the ‘long tail’ would be a rich source of individual tags and would provide further an insight into the motivation of the tagger.

### **3.6.2 Issues with the categorisation process**

The categories chosen will ultimately reflect the ‘user warrant’ of the researcher herself as will be discussed in the discussion chapter and are therefore subjective. The process was carried out by a single person and the categorisation of each tag was at times based on a high level of semantic interpretation so is in turn highly subjective. Defining rigid rules for each category and having a group of people categorising the books would achieve a less subjective results pattern.

### **3.6.3 Issues with the mapping process**

‘Subject’ category tags from each booksonomy were mapped onto the LCSH using a model adapted from research discovered during the literature review. The interpretation of the second and third level of matches (the ‘almost equivalent’ and the ‘semantically similar’ groups including broader and narrower terms) was a subjective one and likewise would have benefited from a wider number of people matching the tags in order to produce a less biased set of results.

## 4. Results

### 4.1 Introduction

The results have been divided into three main sections in this chapter. The first section contains the data connected to the sampling and sample range including the issue of the ‘long tail’ and contains results across the entire range of categories. The second section covers the results of the categorisation process. The results here are given for the whole sample and are broken down for each booksonomy to provide context and a richer understanding of the data. The final section contains the results of the mapping process between the LCSH and the LibraryThing tags that were categorised into the ‘*subject*’ category. This data is also provided for both the whole sample and for the individual booksonomies.

### 4.2 Data collection and sample

After the ten books had been sampled from LibraryThing and their tags collected and cleaned **8,860** individual tags were categorised into the model that had been developed. This figure included the entire tag sample from “Goodnight Moon” and “Pigs” including the ‘long tail’.

Consideration of the two complete samples and investigation of the tag distribution across the sets led to the decision to reduce the sample range over the rest of the eight books to all of the tags applied more than once. The remaining sample covered ten books and **5,568** individual tags. As it is the frequency of the tag that is of interest to this study the frequencies were calculated for each of these tags resulting on a tag frequency of **80,465** over the sample.

#### 4.2.1 The long tail and the sample range

The Zipfian Power Law was observed in the distribution of the tags over the two trial book samples as was seen in Figure 3.13 and 3.14. In the tag data from “Goodnight Moon” 82% of the tags applied fell within the two or more range and for “Pigs” it was 79.7%- the mean average over both sample was 81.2% as Figure 4.1 below shows.

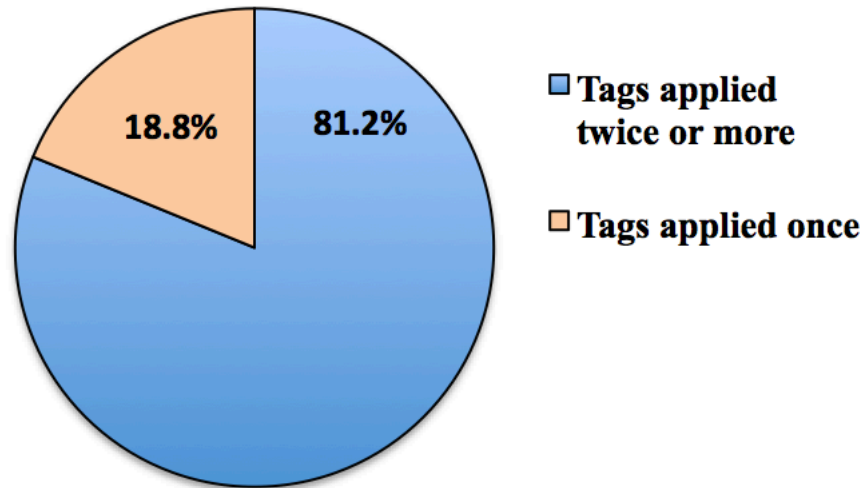


Figure 4.1 Total average distribution between the tags applied once and more than once (Fr)

### 4.3 Categories

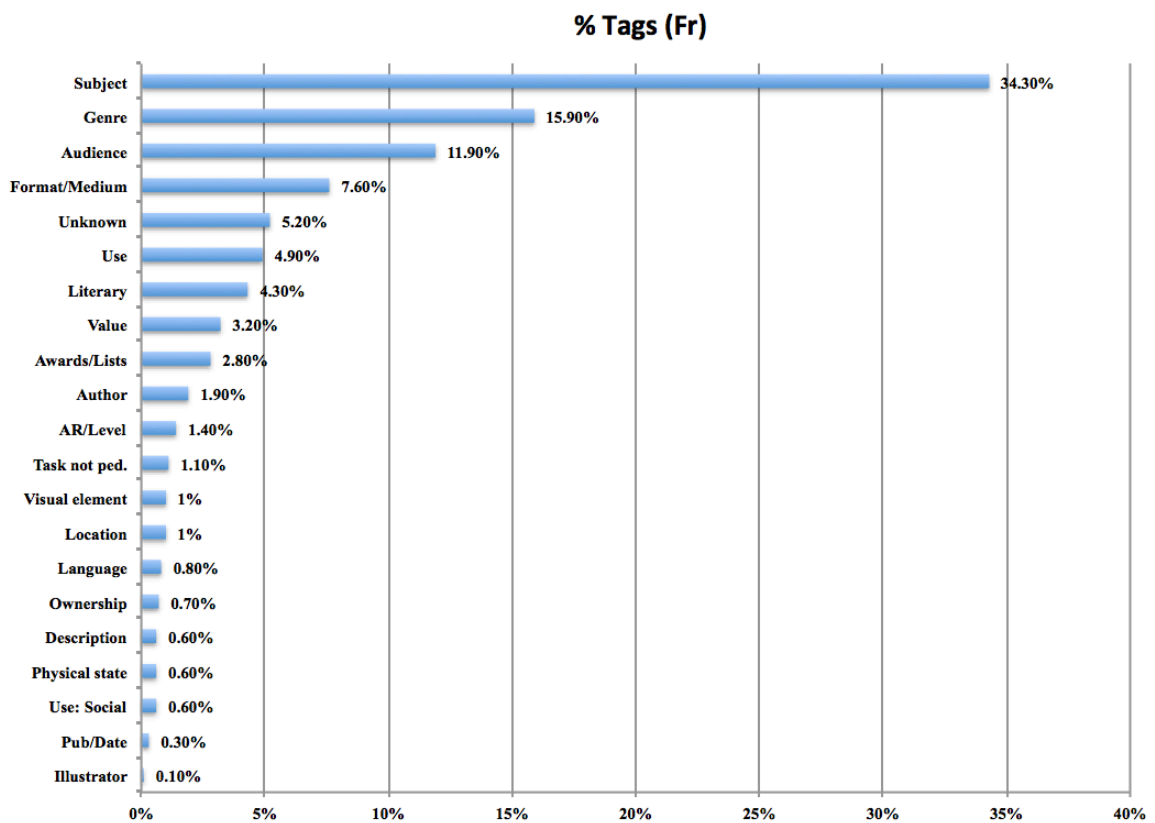


Figure 4.2 Category distribution across all ten books using the mean average, n= 80,465  
 Once the data sets from each of the ten books had been sorted into the category model the results for each category were calculated both individually so that any contextual differences could be investigated and then as a whole set of data. (See Figure 4.2 above)

The spread of tag frequency also shows the Zipfian Power Law curve- 34.3 % of the tags referred to the ‘*subject*’ with the next popular tag category being ‘*genre*’. Only 5.2% of the tags in the sample were categorised at ‘*unknown*’.

### 4.3.1 Category groups

As the research objectives focus on the ‘user warrant’ the initial categories were grouped into four category groups as shown in Figure 4.3 below. The “Personal” tag group consists of ‘value judgements’ and ‘tasks not educational or pedagogical’ e.g. ‘*to read*’, “trashy” tags (to borrow the term from Thomas, Caudle & Schmitz (2010)) consists of everything from the “unknown” and “description” categories.

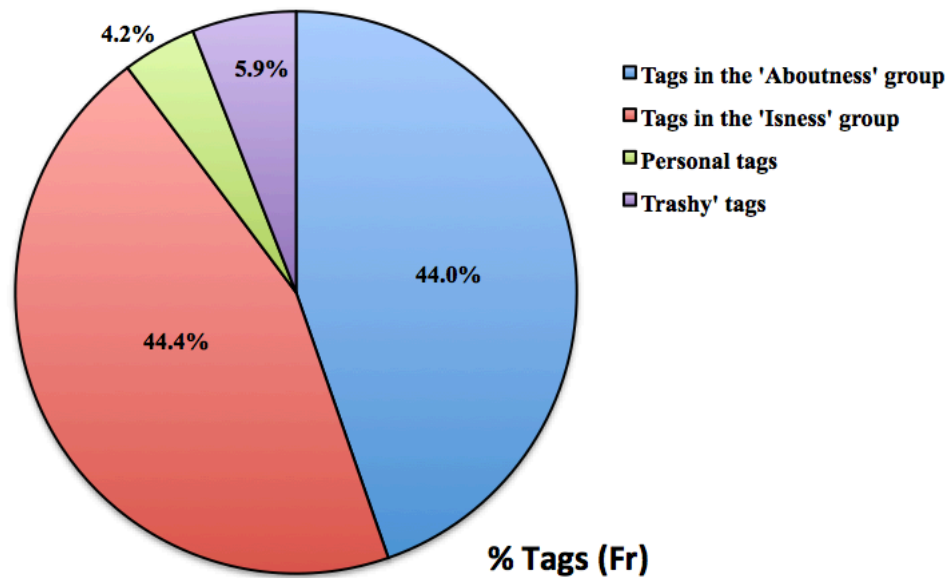


Figure 4.3 Category group distribution across all ten books using the mean average,  $n=80,465$

As Figure 4.3 shows 44% of the tags applied fell within the “Aboutness” group.

#### 4.3.1.1 “Aboutness” group of tags analysis

The 44% of tags within the “Aboutness” group can be broken down further into ‘*subject*’, ‘*literary*’ and ‘*use*’ tags with the largest share of this falling within the ‘*subject*’ category.

##### 4.3.1.1.1 ‘*Subject*’ tag category analysis

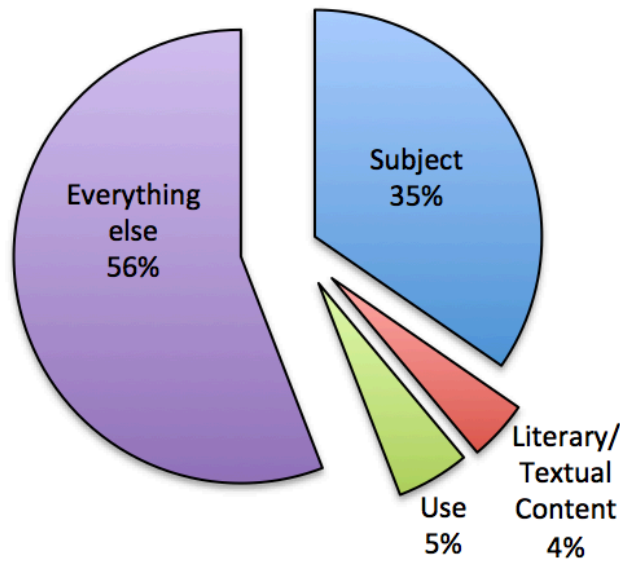


Figure 4.4 Total tag category distribution (Fr)  $n = 80,464$

Figures were obtained from the data sets within the context of each individual book- the mean average of the breakdown of the percentage distribution within the individual book rather than across the whole data sample is shown below with the mean average shown across the chart. There is a narrow range of differentiation within the ‘*subject*’ category.

“The Very Hungry Caterpillar” has the highest number of tags within the ‘*subject*’ category- 47.3%. The matching process between LCSH and LibraryThing tags revealed a high number of tags that could be grouped together semantically within the ‘*subject*’ tags from “Caterpillar” in particular as will be discussed later on in the study.



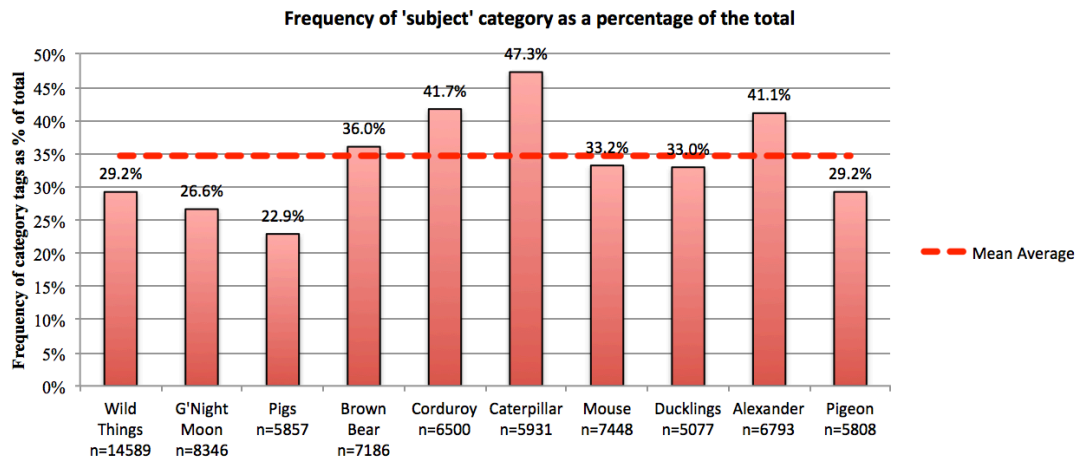


Figure 4.5 ‘Subject’ category distribution (Fr) n= 80,464

“The Very Hungry Caterpillar” has the highest number of tags within the ‘subject’ category- 47.3%. The matching process between LCSH and LibraryThing tags revealed a high number of tags that could be grouped together semantically within the ‘subject’ tags from “Caterpillar” in particular as will be discussed later on in the study. (See Figure 4.6 below for a Word Cloud showing all ‘subject’ terms)

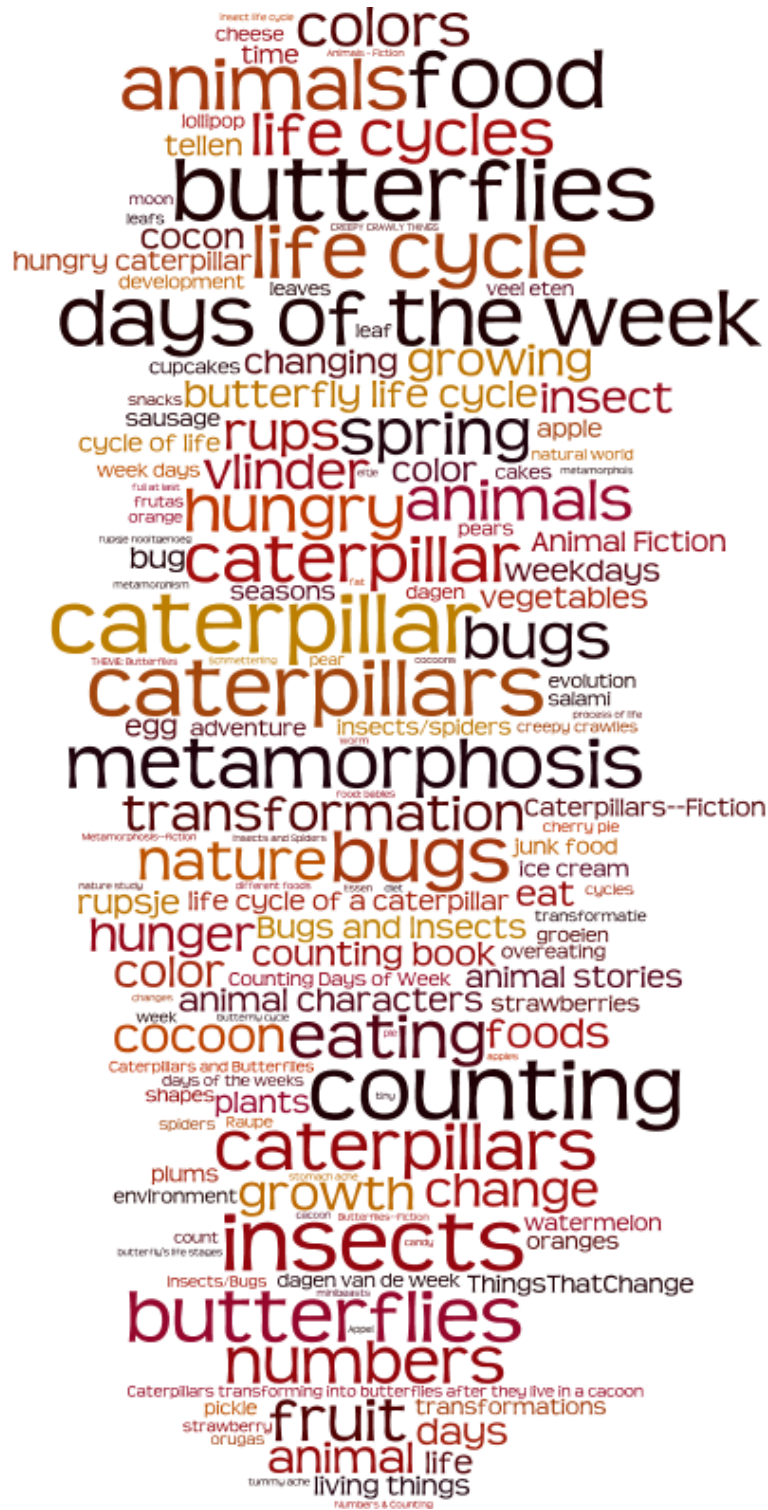


Figure 4.6 Word cloud visualisation of the ‘subject’ category tags from the “Caterpillar” set weighted by frequency (Wordle.net, 2016)

#### 4.3.1.1.2 ‘Literary’ tag category analysis

4% of the tags in the “Aboutness” group refer to ‘*literary*’ or textual tags. There is more of a range in percentage distribution across the ten books within this category- ranging from “Corduroy” with 0.3% to “Brown Bear” which has 13.5%.

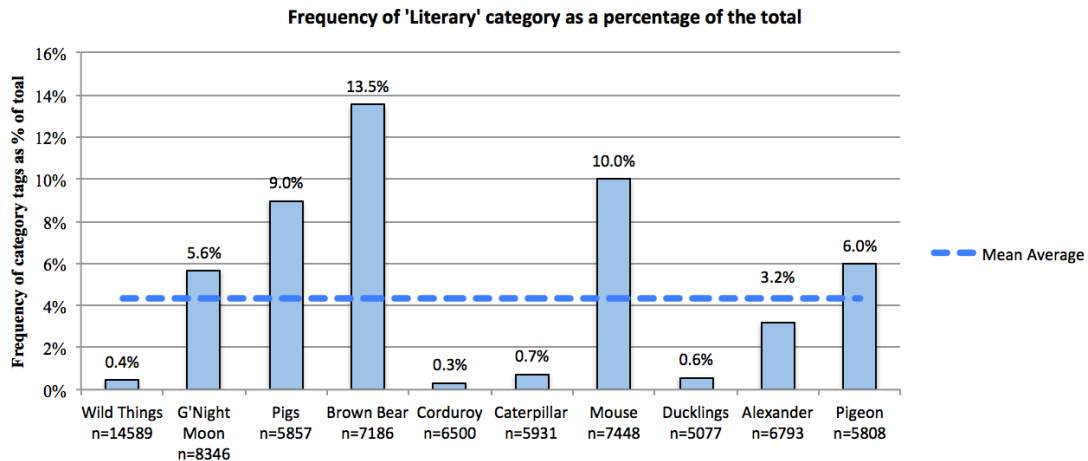


Figure 4.7 ‘Literary’ category tag distribution (Fr)  $n = 80,464$

#### 4.3.1.1.3 ‘Use’ tag category analysis

5% of the tags within the “Aboutness” group were categorised as implying “use”. The spread of percentage distribution across the ten books shows two outlier sets (“Wild Things” 1.9% and “Pigeon” 9.2%) but the other results are more consistent with the mean average.

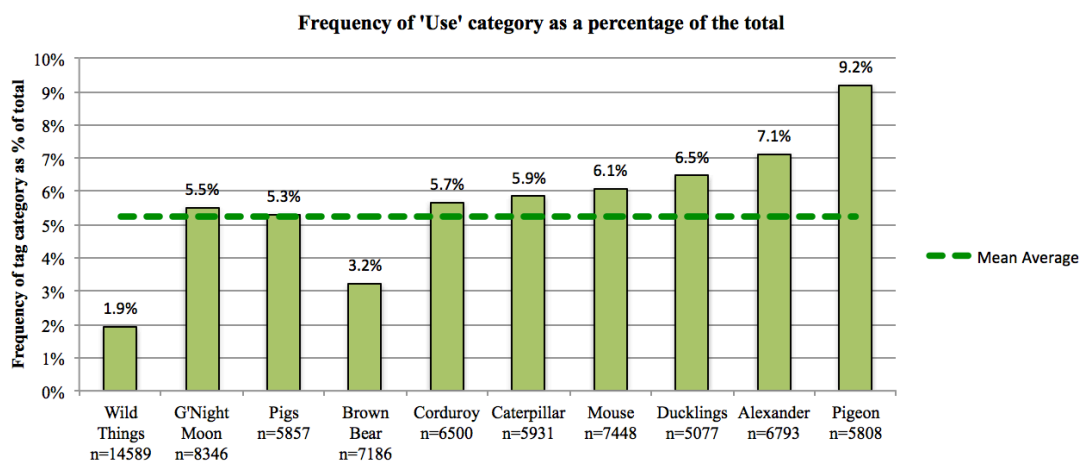


Figure 4.8 ‘Use’ category tag distribution (Fr)  $n = 80,464$

The ‘*use*’ category is of particular focus within this study because investigation of these tags may support Hunt’s “dulcis et utile” theory about how picture books are used for

many things (Hunt, 1998, p.11). The category was further broken down into tags with more specific purposes- ‘*affect*’, ‘*social*’ (referring to those relating to socialisation and semantically implying their ‘use’ in this process), ‘*informal*’ e.g. ‘share with children’ and the largest category- ‘*pedagogical*’ tags or those that imply use within a formal educational context whether at home or at school.

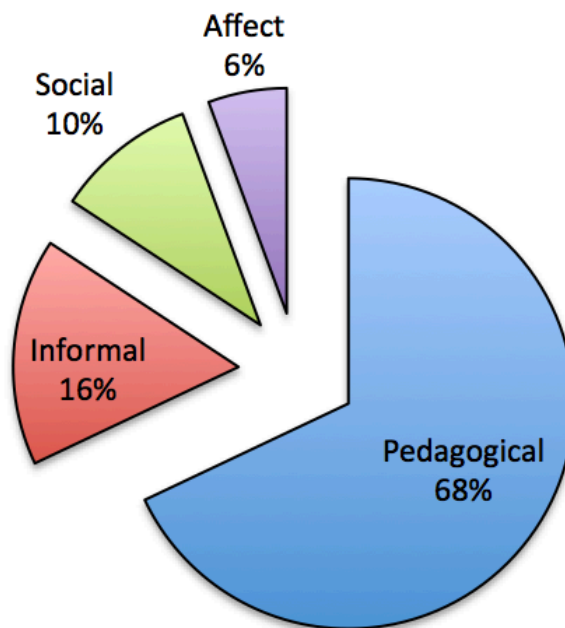
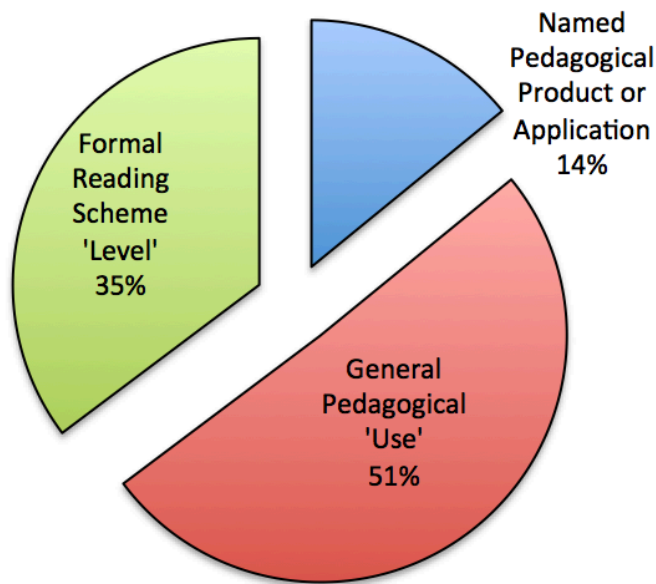


Figure 4.9 ‘Use’ category tags further analysed into sub categories (Fr)  $n= 4,212$

Of the 68 % ‘*pedagogical*’ sub category tags these were then broken down into a further level of sub categories as shown in Figure 4.10. The largest group at this level was ‘*general pedagogical*’ tags but at 35% the formal reading scheme level group is a significant proportion of these tags- 975 tags applied over the ten books in total were in this category.



4.10 “Use: pedagogical” category distribution (Fr)  $n = 2,766$

#### 4.3.1.2 Remaining tags

56% of the tags were not in the “Aboutness” group but fell within the “Personal”, “Trashy” or “Isness” groups. The analysis of how this was distributed within each of the ten individual booksonomies is shown in Figure 4.11. “Wild Thing” had the highest proportion of tags in this group at 68.4% but with the lowest being “Caterpillar” at 46.1% the range in results is fairly narrow.

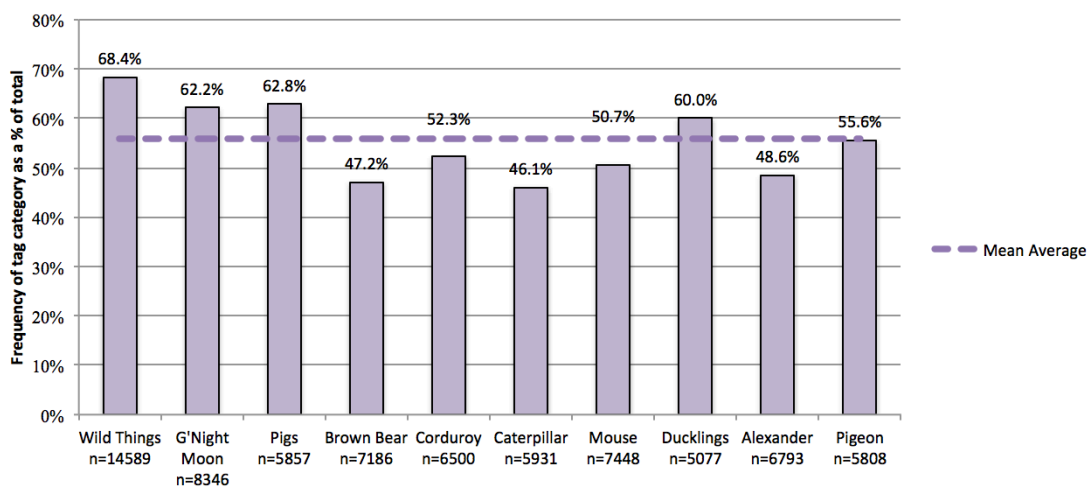


Figure 4.11 Tags not within the “Aboutness” group across all ten books (Fr)  $n = 80,464$

## 4.4 Mapping

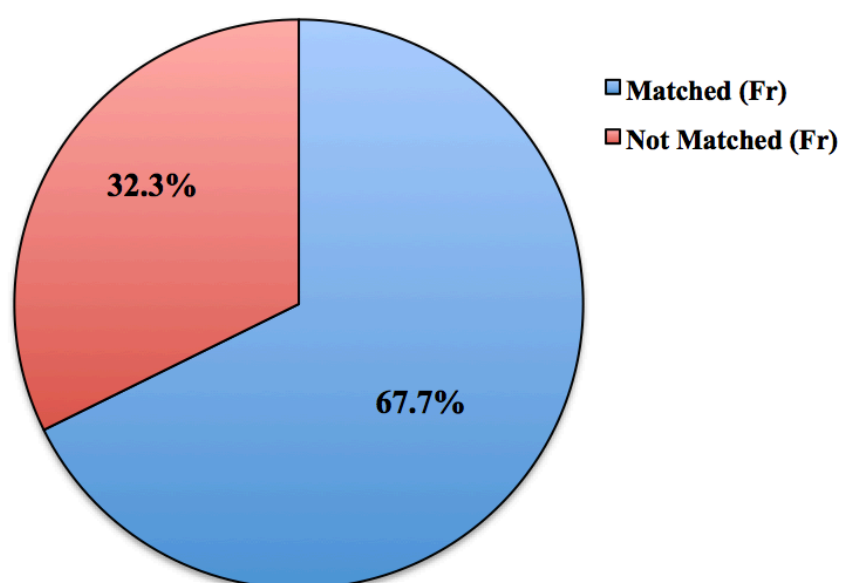


Figure 4.12 showing total % matched and not matched using the mean average (Fr),  $n=34,408$

The LCSH to LibraryThing tag mapping process resulted in 67.7% of the LCSH being matched by the ‘3<sup>rd</sup> Level’ of tags.

The tags from the folksonomy were used more than once if appropriate and the frequency was used within the calculations so the figures used refer to the % of LibraryThing ‘*subject*’ category tags that were related to the LCSH headings either directly (1<sup>st</sup> Level), semantically- misspellings, plurals, foreign language tags that matched the LCSH (2<sup>nd</sup> Level) or by the 3<sup>rd</sup> Level which included related terms within a basic level variation both superordinate and subordinate (Golder & Huberman, 2006).

Book/Sample	Total at 1st	Total by 2nd	Total by 3rd	Not matched
Alexander <i>n</i> =2916	2.9%	3.1%	43.2%	56.8%
Mouse <i>n</i> =2185	13.0%	31.5%	46.0%	54.1%
Caterpillar <i>n</i> =9597	15.7%	22.4%	47.5%	52.5%
Corduroy <i>n</i> =2838	19.0%	20.9%	50.7%	49.3%
G'Moon <i>n</i> =2543	31.9%	35.9%	59.3%	40.7%
Wild Things <i>n</i> =5616	28.8%	31.1%	70.7%	29.3%
Pigs <i>n</i> =1371	7.5%	63.3%	83.6%	16.4%
Pigeon <i>n</i> =1913	15.7%	41.9%	87.1%	12.9%
Brown Bear <i>n</i> =2634	37.1%	68.6%	91.9%	8.1%
Ducks <i>n</i> =2795	26.4%	62.7%	97.4%	2.7%
<b>Mean Average</b>	<b>19.8%</b>	<b>38.1%</b>	<b>67.7%</b>	<b>32.3%</b>
<b>Median Average</b>	<b>17.4%</b>	<b>33.7%</b>	<b>65.0%</b>	<b>35.0%</b>

Table 4.13 Showing the mapping data broken down by book

As the result for the matches for “Ducklings” appeared to be a significant outlier within the range the median average of tags not matched was calculated to see whether it would get a more relevant benchmark but as can be seen in Figure 4.14 below there was only a 2.7% different between the mean and median average so on balance the mean average is the more relevant measure of matching rates.

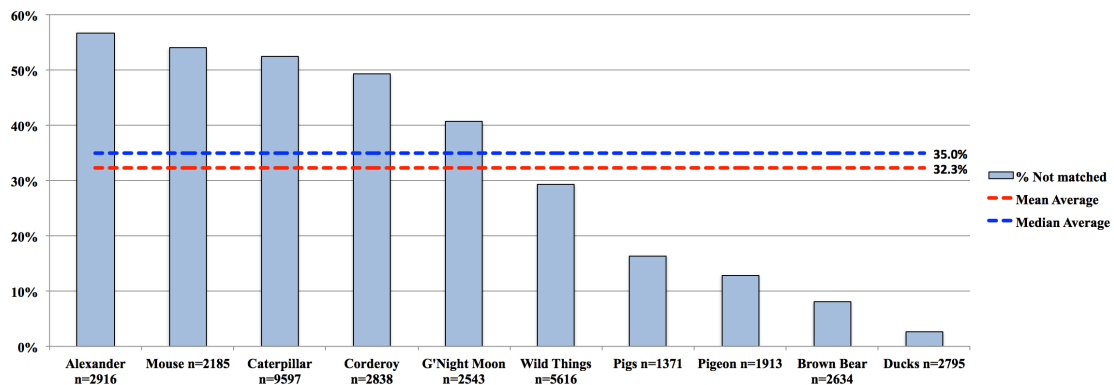


Table 4.14 Showing the % not matched in ranked order compared to the mean and median average.

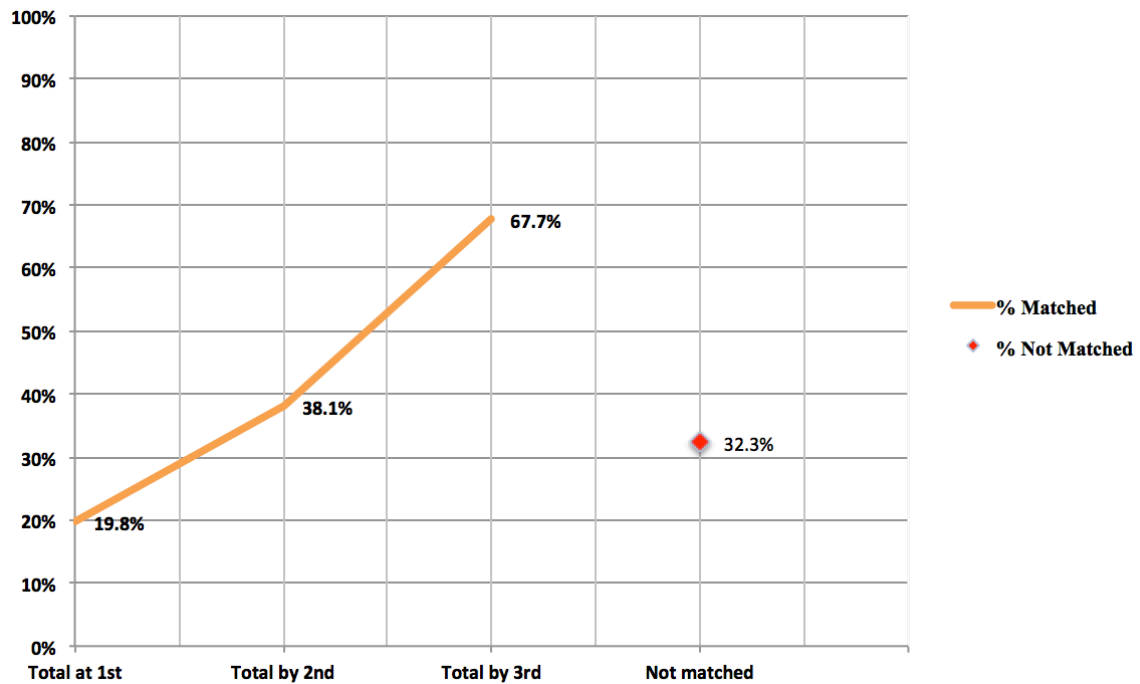


Figure 4.15 Plotting % of tag matches (Fr) at each stage (mean average across the sample  $n= 34,408$ )

By the 2<sup>nd</sup> Level of the mapping process 38.1% of the ‘*subject*’ tags had been matched and only 32.3% remained unmatched by the end of the process. There was a certain amount of variation across the ten individual booksonomies and this is represented in Figure 4.16 at the end of this chapter where the stages are broken down for each book.

#### 4.5 Summary of findings

Tags from ten picture book booksonomies were analysed after discarding the long tail within the sample. **5,568** individual tags were applied **80,465** times within the sample and it was the frequency of the tag that has been used within this study in order to get a clearer view of tagging behaviour.

##### 4.5.1 The long tail

81.2% of the tags within the two trial booksonomies categorised were applied twice or more leaving a long tail of just 18.8%. This long tail did however contain a high percentage of tags that fell within the ‘*use*’ category but it is beyond the scope of this survey to investigate all ten books completely.



### 4.5.2 Categories

The largest single category across all of the booksonomies was the '*subject*' category with accounted for 34.3% of the tags (Fr). When the category groups were analysed 44.4% fell into the "Isness" or denotative group, 44% into the "Aboutness" group and only 5.9% into the "Trashy" tags group using the model developed. The low level of "Trashy" tags may be due in part to the subjective technique employed which involved actively seeking reasons to interpret the tag rather than obeying strict rules about misspellings and acronyms for example. The fact that each book was read and referred to by the researcher and that the tags themselves were consulted via their hotlinks in order to investigate a 'meaning' perhaps also accounts for the low number of indecipherable tags.

The distribution of categories within the "Aboutness" group was fairly consistent across the booksonomies except in the case of the '*literary*' category where there was some variation. The '*use*' category accounted for 5% of the sample and was investigated further as it is one of the objectives of this study to look at the different "uses" of the picture book as a medium implied by the tags. The main '*use*' category was further broken down into subcategories drilling down into the pedagogical category in particular.

### 4.5.3 Mapping

The mapping of LSCH to LibraryThing tags described in the methodology resulted in 67.7% match of tags across the systems. 38.1% of these occurred by the 2<sup>nd</sup> Level of the mapping process which included those tags not literally but semantically the same as the LCSH.

Further investigation will be done within the discussion chapter into the subject tags not matched to see whether there is any evidence of consensus within groupings of tags that might imply the 'need' for a new LCSH.

## 4.6 Limitations

The results are limited by scope- both in the small number of books within the sample and in the reduced data set that discounted the long tail. The categorisation and mapping processes were carried out by one person so they carry a high level of subjectivity.

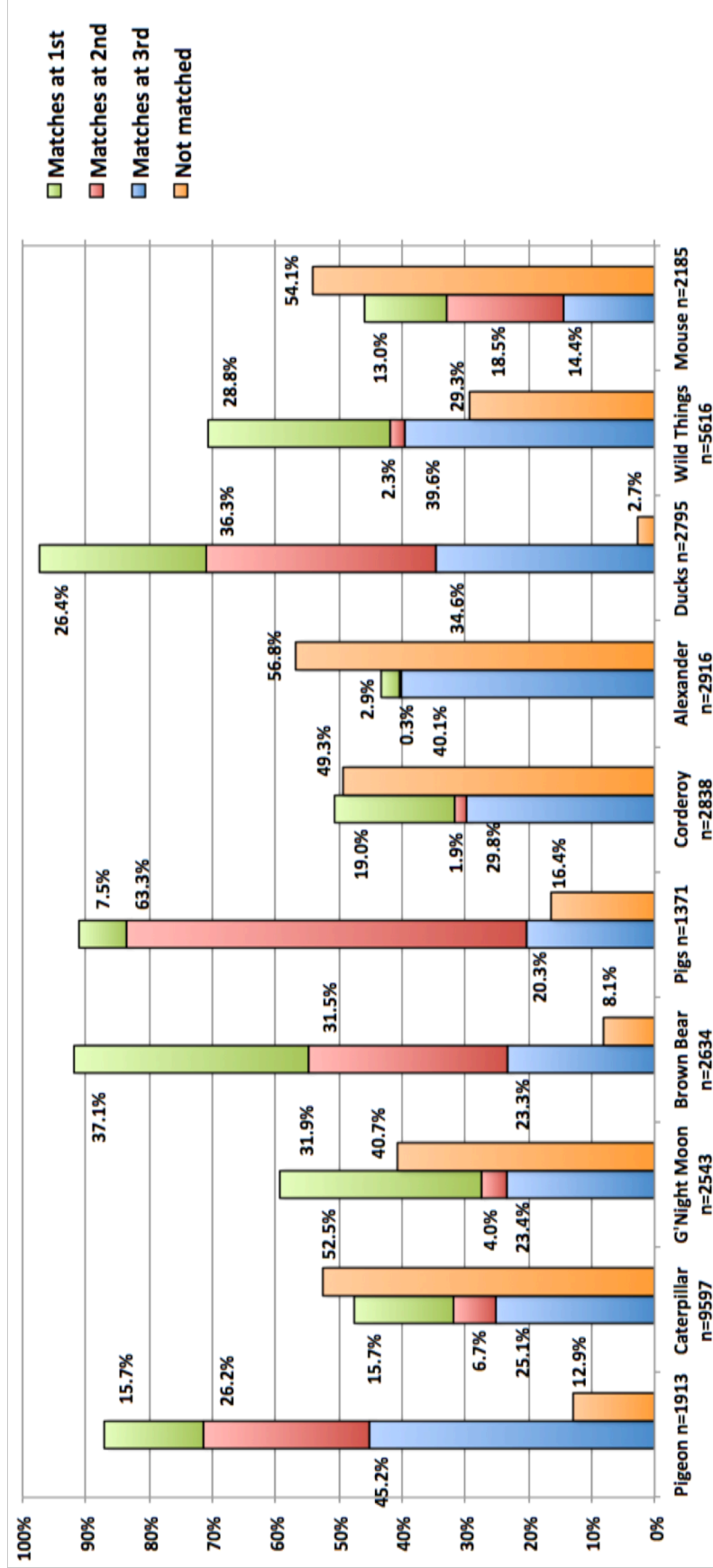


Figure 4.16 Matches broken down by book at each level (Fr)

## **5. Discussion**

### **5.1 Introduction**

In this chapter the results will be discussed using the same sections as in the previous results chapter and in reference to the overall research aims and objectives. Firstly the sampling and methodology will be reviewed and discussed, then the categorisation process and results, the mapping of the '*subject*' tags to the LCSH and finally the investigation into the patterns revealed in the research and how this might have relevance for further research in either information or picture book studies or a practical implication for cataloguers or designers of information retrieval systems.

### **5.2 Literature review and methodology**

A literature review was carried out with a special focus on folksonomies, booksonomies, categorising tags, mapping tags, user warrant, 'aboutness' and the position of the picture book within society. This informed the methodology on many levels- the choice of LCSH sampled from WorldCat compared with the LibraryThing tags was based on the many other studies that chose these two open access systems (Guyot, 2013; Heymann & Garcia-Molina, 2009; Lawson, 2009; Lu, Park & Hu, 2010; Adler, 2009; Iyer & Bungo, 2011; Bartley, 2009; Smith, 2008).

The sample retrieval method was influenced by Guyot's method using a clean-up process that imported the raw data via Microsoft Word into Excel maintaining the tag frequencies as they give a "more accurate view of the importance of each tag and this weights the addition of the tag to the category into which it is categorised" (2013, p.41).

#### **5.2.1 Long tail**

In this research the 'long tail' that was removed included only those tags applied once, this accounted for 81.2% of the total tags. To compare this rate with that of Guyot who used tags applied three times or more the average over the two trial titles in this research was 75.2 % compared to Guyot's 83 % (p.38). A larger sample would have to be used in

order to draw any conclusions from this but perhaps as her booksonomies were more populated with individual tags and a higher number of users had applied them her sample showed a higher rate of consensus within the tag distribution.

The Zipfian Power Law was observed within the distribution of tag frequencies at Mathes (2004) predicted would always be the case with folksonomy-based systems (Mathes, 2004, p.11; Guy & Tonkin, 2006; Golder & Huberman, 2006; Guyot, 2013). Additionally, just as Ke & Chen (2012) observed, the tag category distribution also obeyed the power law (see Figure 4.2).

The high number of tags categorised within the concept of ‘use’ that were used only once meant that due to the scope of this study they were unable to be investigated- further study into these descriptive ‘long tail’ tags would perhaps be an interesting one- especially if it was found that a certain group of taggers use this very personal ‘task based’ technique throughout their collections.

### **5.3 The categorisation process**

The category model was designed by combining various methods and approaches researched as part of the literature review. The structure, with the two main “Aboutness” and “Isness” group of tags was based on a combination of Golder & Huberman’s (2006) ‘extrinsic’ and ‘intrinsic’ sections (yet redefined), Thomas, Caudle & Schmitz’s (2009) adaptation of this, and an interpretation of Beghtol’s user warrant theory (1986) and Shatford’s “Aboutness” and “Isness” concepts (1986).

Contextual cataloguing based on the method employed by Iyer and Bungo (2011) was taken a step further when matching ‘unknown’ tags by using the hotlinks to look at the tag in question within the community context with which it was applied. This technique resulted in a low number of over all ‘Trashy’ tags of 5.9%. Acronyms, misspellings and foreign language tags were for the best part included after some investigation also contributing to this low level of “Trashy” tags compared to the data discussed by Thomas, Caudle & Schmitz in their study “Trashy tags: problematic tags in

LibraryThing” (2010). Most studies that they discuss included misspellings, foreign language terms and non-alphabetic tags in their “Trashy” tag category- in their research this was at 34% of the overall sample which they concluded was comparable to the levels observed in other studies of folksonomies (p. 229).

### **5.3.1 ”Aboutness” group of tags**

44% of the sample fell within the “Aboutness” group results- this included ‘*subject*’, ‘*literary/textual*’ related tags and ‘*use*’. Within the ‘*subject*’ category there was little deviation from the mean average across all ten booksonomies. “Pigs” had the lowest count at 22% but it had a much higher than the mean average percentage of ‘*literary*’ tags instead- possibly because as a ‘post-modern’ retelling of a fairy story it was perceived to be more “about” it’s own form and stories in general than about a wolf and some pigs! It is difficult to compare the category results of this survey with the category results of many other surveys because the categorisation model is fundamentally different in terms of the conceptual groupings but the closest comparable result in percentage of ‘subject’ tags is with Guyot’s 27% (p. 42) compared to 35% found in the picture book booksonomies. The slightly higher percentage is perhaps due to the contextual analysis of the tags as opposed to Guyot’s aggregated data set.

“The Very Hungry Caterpillar” had the highest percentage share of ‘*subject*’ tags- perhaps because Spanish, French and Dutch words were included in this list as the book has been translated into many different languages (see Figure 4.6).

### **5.3.2 ‘Literary’ category tags**

Guyot discussed what she called “miscellaneous tags” which she claimed were “useful’ and her examples included ‘multiple plot’, ‘book within a book’, and ‘unreliable narrator’ which would all have fallen within the ‘*literary*’ category within this study (Guyot, 2013, p. 40) No other evidence of ‘*literary*’ category tags as examined here were discovered in research but they may have been included within different tag categories in other category systems. The ‘*literary*’ category was the only one of the categories to have a wide range between booksonomies with a mean average of 4% over all of the sample.

The percentages ranged from “Brown Bear” with the 13.5% of the total tags relating directly to the text as opposed to “Corduroy’s” 0.3%.

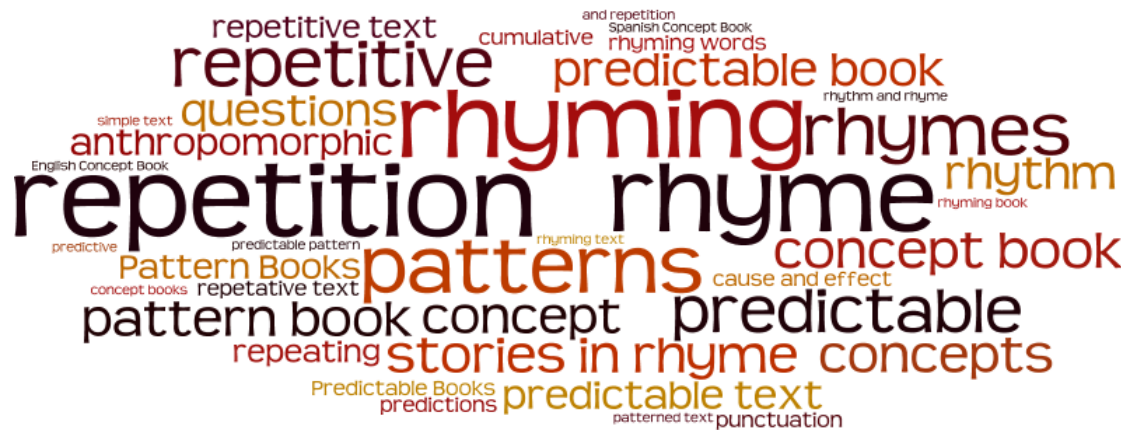


Figure 5.1 ‘Literary’ tags applied to “Brown Bear”, weighted according to frequency and visualised using Wordle. (Wordle.net, 2016)

If you compare the two sets of tags however it is clear that this disparity is due to the nature of the two books and also the literary techniques discussed with children of this age. “Brown Bear” as the tags suggest is a repetitive, rhyming book for the very young and in both the American Common Core curriculum and the British National Curriculum there is an emphasis on making even very young children aware of simple literary techniques such as these so we could surmise by this result that the ‘user warrant’ is tending towards wanting to find rhyming and repetitious books- possible to use as a teaching aid. “Corduroy” on the other hand is a traditional narrative with an alive and feeling teddy bear but no other major literary devices that might want to be used to teach the theory therefore there is a low count on this category for this title.

Tag Name	Lit F
personification	6
anthropomorphic	3
cause and effect	3
anthropomorphic toys	2
anthropomorphism	2
dialogue	2
sentence structure	2
<b>Total</b>	<b>20</b>

Table 5.2 ‘Literary’ tags applied to “Corduroy

### 5.3.3 ‘Use’ category

It could be argued that the tags in the ‘literary’ category also belong in the ‘use’ category but it was felt to be too much of a leap to imply that all of the tags describing the language were for teaching or educational purposes. No other booksonomy researched observed the ‘use’ category of tags in this way- this may be because of the unique ‘function’ of the picture book in society and education or this might be because out of context these tags would be difficult to interpret and identify for the processes of categorisation.

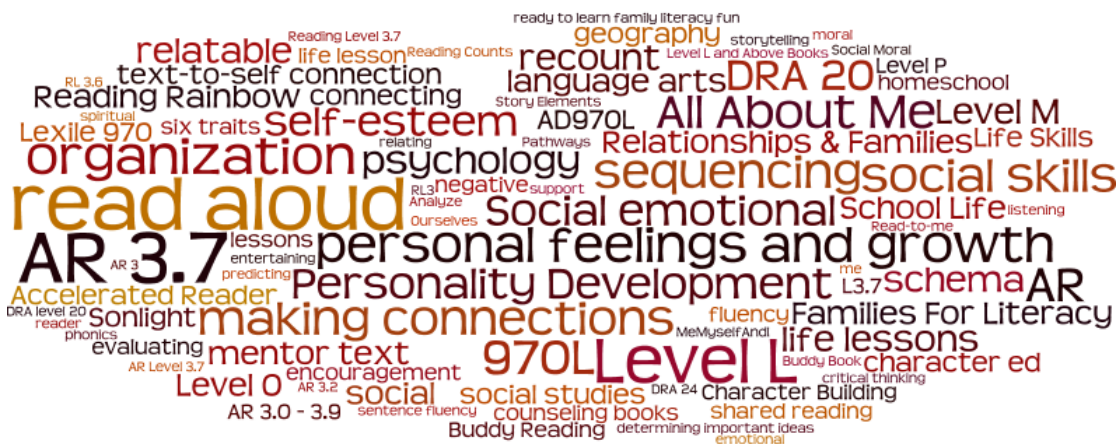


Figure 5.3 ‘Use:pedagogical’ tags applied to “Alexander”, weighted according to frequency applied (Wordle.net, 2016)

5% of the tags were categorised as implying ‘*use*’ of various kinds which were further analysed to subcategories of pedagogical, informal, social and affect then the pedagogical subset was broken down further to separate those referring to an actual named pedagogical product or programme or those that gave the book a level from a graded reading scheme. Guyot found 0.2% of the tags in her study referred to reading systems booksonomies on LibraryThing (2013, p. 42). She doesn’t break her statistics down into all of their different mediums but she does specify that within the ‘Young Adult’ group of books there were 0.2% ‘reading system’ related tags and within the non-fiction selection none at all (p.46). This observation suggests that ‘*use*’ tags (certainly the ‘*level*’ category within this) occurs in booksonomies for young people when it doesn’t appear in those for adults. Further studies would have to be undertaken to investigate this any deeper however.

#### **5.3.4 Implications of the categories discovered**

What’s not reflected in the traditional folksonomy is the “cultural warrant” reflected in the ‘*use*’ category and the definite sub categories that has been revealed by analysing the tag data. As a distinctly connotative category this might be seen as a rich source of information for picture book researchers to observe picture book booksonomies evolving over time and possibly within particular user groups to observe cultural warrant forming a consensus about certain tags or books in a manner similar to Adler’s study of people who tag books about transgender (Adler, 2009). Primary school teachers, parents and school librarians could all be identified as separate user groups within LibraryThing via the groups pages or forums to carry out a study of this kind in the future.

The results suggest that there would also be enough data here within the picture booksonomies to carry out a longitudinal study using either categories or language used to see whether as Colomer suggests there is a change in consensus about what picture books ‘should be about’ or perhaps how they are re-interpreted over time (Colomer, 2010). The retrospective application of the 1996 ‘conduct of life’ LCSH to “Where the Wild Things Are” (1963) is a relevant example. It is defined by the Library of Congress as a term that relates to “works on moral and ethical values in everyday life.” (Library of



Congress, n.d.). which appears to comply with what Colomer was arguing about the shift in values within the picture book publishing world between the 70s and the year 2000, away from “the power of imagination” as the focus in the 70s to that of social protection and emotional literary in 2000 (Colomer, 2010, p.42-46).

Colomer’s research was carried out looking at the themes of books published at those times but perhaps by tracing the application of different ‘descriptors’ applied to ‘classic’ picture book titles both in formal and informal taxonomies or the forming of new consensus over stable tags that imply the ‘interpretation’ and perception of a picture book could also be traced over time to observe this phenomena.

### **5.3.5 Implications of the findings for cataloguers and IRS**

When Beghtol wrote about ‘cultural warrant’ in reference to “personal and professional cultures of information seekers” demanding the establishment of “fields, terms, categories or classes in a [particular] knowledge representation and organization system” (Beghtol, 2005, p. 2) she could have been referring to the school and children’s librarian community. The market leading information retrieval system (IRS) for schools in the UK, Microlib does not supply fields within their ‘Junior Librarian’ or ‘Eclipse’ management systems for adding terms relating to ‘use’. A recent project by the researcher to catalogue 5,000 books for 2-8 year olds within a school library context highlighted the need for other entry points into this IRS to reflect the ‘use’ that the books will be put to and the need for them to be discoverable in this way. The only solution to this issue at present within Junior Librarian is to create controlled terms within their “Key Term” field and create a personal taxonomy for the specific library in that school but this could cause issues in federated catalogues across multiple sites.

The results of this study have indicated that there is a definite ‘user warrant’ for an IRS used in a school to have fields denoting other different categories e.g.

### **-Named pedagogical products, programmes**

e.g. if the book is one of the books that teachers base a teaching strand on or that is part of a formal reading programme such as Accelerated Reader

### **-Levels**

The ability to identify ‘levelled’ books from reading schemes. In some cases (e.g. the Accelerated Reader scheme) ‘real’ picture books are included within a larger collection of picture books but need to be identified for this specific use.

### **-Formal pedagogical use**

The research has shown a strong ‘user warrant’ for a method of key word entry point based on the ‘*use*’ concept to be part of a formal taxonomy system with an IRS for schools. E.g. to be able to quickly identify books that would help teach key early literary concepts like ‘narrator’ or ‘repetition’

## **5.4 The mapping process**

The mapping technique used in this study was adapted from the ‘LCSH Tree’ methods used by Heymann & Garcia-Molina (2009), Yi (2009) and Yi & Chan (2009) in that it simulated the syndetic structure of formal taxonomies in order to match the two sets of descriptors. A three step ‘level’ approach was adopted much like the ‘exact’ and ‘almost exact’ matches and ‘semantic equivalents’ stages used by Heymann & Garcia-Molina (2009, p.2). The crucial difference between this research and the three examples of research mentioned above is that the previous research uses the LCSH as the ‘benchmark’ set of data with which to match the tags whereas here the tag set for each book has been matched to the LCSH – the process is inverted because the focus of this present study is in the investigation of the tag sets in comparison to the LCSH rather than the other way around.

Heymann & Garcia-Molina found that in their study 48 % of their LCSH tags had been matched by their second stage (either equivalent or almost equivalent) (2009, p. 3) and if

we look at the data in this research the level over the ten picture books using the LCSH as the benchmark is even higher at 90.7%. ‘Almost all’ of their LCSH had been matched by the end of the semantic matching process and likewise all of those used in the picture book booksonomies had been matched.

	1 <sup>st</sup> Level	2 <sup>nd</sup> Level	3 <sup>rd</sup> Level	No. of tags
Pigeon	10	10	13	13
Caterpillar	3	3	3	3
Moon	2	2	2	2
Bear	2	2	2	2
Pigs	1	2	2	2
Corduroy	3	3	3	3
Alexander	1	1	1	1
Ducklings	5	9	10	10
Wild Things	4	6	6	6
Mouse	1	1	1	1
	<b>74.4%</b>	<b>90.7%</b>	<b>100.0%</b>	<b>43</b>

Figure 5.4 % of LCSH matched at each stage using the LCSH as the ‘benchmark’

The higher number of matches in this research may be due to the subjective matching process of the researcher- for example the LCSH ‘animal welfare’ was matched at the second level to ‘animal rights’ when perhaps another person may have seen this as a related rather than an equivalent term. Misspellings and foreign language words for the same term were also used as ‘equivalents’ at the second level so in this way it is perhaps unfair to compare these results to those of Heymann and Garcia-Molina who had more rigid guidelines about the second level of matches. Further research would have to be undertaken to test whether there is a wider point here to be made about the LCSH applied to picture books as a group compared to other samples.

To return to the mapping process that is outlined in this research (using the tags as the benchmark set) there was a 67.7% matching rate by the 3<sup>rd</sup> Level of matches. Even by the 2<sup>nd</sup> Level of the mapping process 38.1% of the ‘*subject*’ tags had been matched- these

tags were semantically the same as the LCSH, e.g LCSH metamorphosis to metamorphisim (sic.) and would have been dealt with within a controlled vocabulary.

There was a degree of variation between the booksonomies and as there was also a wide range in the number of LSCH applied to each book so a test was done to see whether there was any correlation between these two factors. As you can see in Table 5.5 and Figure 5.6 there was no correlation between these factors that could be detected within this very small sample of booksonomies.

	No. of LCSHs	Tags not matched (Fr)
Pigeon	13	12.9%
Ducks	10	2.7%
Wild Things	6	29.3%
Caterpillar	3	52.5%
G'Night Moon	2	40.7%
Brown Bear	2	8.1%
Pigs	2	16.4%
Corduroy	2	49.3%
Alexander	2	56.8%
Mouse	1	54.1%

Table 5.5 LCSH tags for each book compared to the % of tags not matched

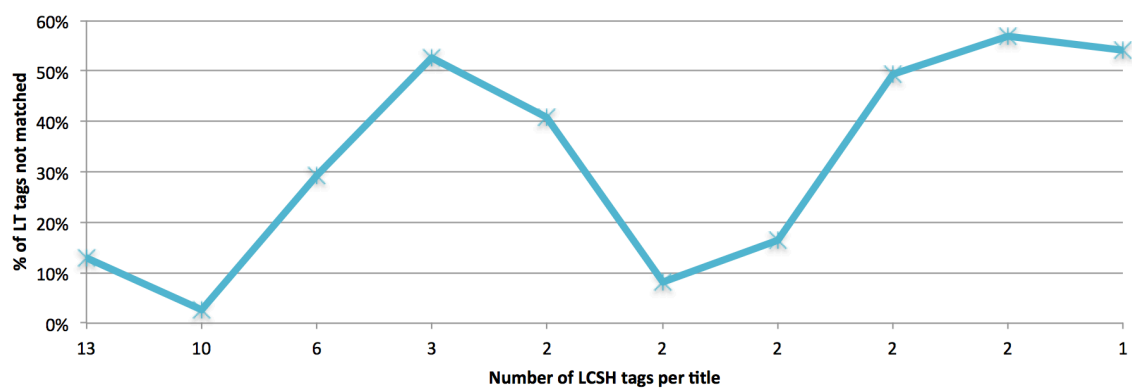


Figure 5.6 LCSH tags for each book compared to the % of tags not matched

### 5.4.1 Subject tags not matched and emerging consensus.

There were a significant number of tags not matched by the third level of the mapping process. In order to investigate whether there was any evidence of an emerging consensus within the non-matched tags each set was arranged into groups by concept and totals were created of those that seemed to have a high frequency when combined in a basic level variation with related terms.

The booksonomy with the most significant number of tags that formed distinct ‘*subjects*’ not represented by the LCHS was “The Very Hungry Caterpillar”. The LCSH of ‘*butterfly*’, ‘*caterpillar*’ and ‘*metamorphosis*’ do not reflect the very clear “user warrant” that the booksonomy reveals for subject terms that reflect the very distinct concepts in Table 5.7 below.

Subject area	Tags (Fr)	% of total tags
Hunger/Eating	703	7.33%
Counting	592	6.17%
Life Cycles	392	4.08%
Days of the week	386	4.02%

Table 5.7 subject headings ‘suggested’ by the non-matched subject tags within the “Caterpillar” data  $n=9597$

Added together these four new tags would amount to a further 2073 tags being matched, or 22% of the total subject tags (Fr). The breakdown of the individual tags within these concepts are in Appendix F, with similar sets arising from other titles in Appendices G-I.

#### **5.4.2 Implications of the tag mapping findings for traditional taxonomies**

Evidence of emerging consensus within the subject tags from the folksonomy indicate that there is a clear relevance in trying to incorporate this into the formal taxonomy systems that exist. When Cutter was first imagining a ‘syndetic’ system in 1875 it is clear that he actually favoured ‘user warrant’ over the authority of the cataloguer- “Useage (sic)...is the supreme arbiter- the useage, in the present case, not of the cataloguer but of the public in speaking of subjects” (Cutter, 1875, p.69):-

“the practice of reducing a name to the substantive form is often a good one; but should not be insisted upon as an invariable rule, as it might lead to the adoption of some very out-of-the-way names....As is often the case with language, useage will be found not to follow any uniform course” (p. 74)

Perhaps Cutter would have considered that “Conduct of life” to mean “behaviour” or “attitude” as applied to “Where the Wild Things Are” to be an ‘out-of-the-way’ name. It wasn’t matched by any of the tags from the folksonomy sample within the mapping process.

Cutter could hardly have dreamt of the advent of Library 2.0 and the possibility of informal folksonomies entirely driven by user warrant, reflecting language and cultural change and different interpretations of language and meaning. Studies have concluded that a system based entirely on a folksonomy however has limitations, as Ransom and Rafferty discuss further research into the nature of folksonomies has “led to the general consensus that user-tagging is likely to compliment rather than to replace formal classification systems” (2011, 1039)

#### **5.5 Summary of discussion**

This study shows that the investigation of booksonomies can reveal information relevant to researchers, cataloguers and designers of IRS as well as providing an alternative, more dynamic and varied system for describing and discovering books. The research points to some very specific practical applications that might be made to IRS used in schools as

suggested by the 'cultural warrant' reflected in the "use" tag categories and the sub categories within them.

## **6. Conclusion**

### **6.1 Introduction**

The aim of this study was to investigate picture book booksonomies sampled from LibraryThing with a particular focus on what these tags ‘revealed’ about ‘user warrant’ both in terms of ‘use’ indicated by the tag categories and their frequencies and in the concept of their ‘subject’ in particular. The objectives of the study are reviewed in this chapter as are the methodology, results and discussion and the limitations of this research. Possible areas of further study and the practical implications of the findings are also discussed.

### **6.2. A review of the aims and objectives**

#### **6.2.1 Aims**

To investigate picture book booksonomies on LibraryThing with a particular focus on the ‘user warrant’ that might be revealed in the tags or tag categories that are discovered.

#### **6.2.2 Objectives**

*1. To conduct a literature review into folksonomies and social tagging with an emphasis on categorisation, mapping with authorised terms such as LCSH and tagger motivation*

A review was undertaken and many relevant studies were discovered which in turn informed my methodology and approach. A gap was discovered in the research as no research has been previously carried out into either children’s booksonomies or into the relevance of bibliographic data either formal or non-formal to the field of picture book study.

*2. To categorise tags obtained from a sample of picture books on LibraryThing adapting existing models where relevant and to investigate the ways in which tags are applied to picture books by analysing the conceptual categories that they fall into.*



Tags from ten picture book booksonomies were analysed after discarding the long tail within the sample, the distribution of the tag frequencies were found to conform to the Zipfian Power Law so a stable community of tags was in evidence. **5,568** individual tags were applied **80,465** times within the sample and it was the frequency of the tag that has been used within this study in order to get a clearer view of tagging behaviour.

81.2% of the tags within the two trial booksonomies categorised were applied twice or more leaving a long tail of just 18.8%. This long tail did however contain a high percentage of tags that fell within the '*use*' category but it is beyond the scope of this survey to investigate all ten books completely. It would be advisable if further study into picture book booksonomies was carried out to include which the 'long tail' tags as there is evidence that it might contain tags relevant to the study of user behaviour and the '*use*' to which picture books are put to.

The largest single category across all of the booksonomies was the '*subject*' category which accounted for 34.3% of the tags (Fr). When the category groups were analysed 44.4% fell into the "Isness" or denotative group, 44% into the "Aboutness" group and only 5.9% into the "Trashy" tags group using the model developed. It was difficult to compare the results to those found by other researchers as my category model fundamentally differed in definitions of '*subject*' and in the existence of the '*use*' and related categories.

The evidence suggests that picture book booksonomies do indeed display the '*dulcie et virtute*' philosophy as the '*use*' category accounted for 5% of the sample. Further investigation revealed tags being used to denote different kind of '*use*'- formal pedagogical, informal pedagogical, social, affect and a reference to book '*reading levels*'. Further research could be undertaken to see whether using the category model and definitions outlined in this research any evidence of this '*use*' is to be found in other groups of booksonomies e.g. adult fiction, young adult fiction and longer fiction for children.

The study was limited by the size of the sample, the choice to remove the ‘long tail’ tags and by the subjective categorisation process

*3. To extract the ‘subject’ category tags and use these to ‘map’ against LCSH, investigating any conceptual groups of tags left for possible emerging consensus for additional subject headings.*

The mapping of LSCH to LibraryThing tags described in the methodology resulted in 67.7% match of tags across the systems. 38.1% of these occurred by the 2<sup>nd</sup> Level of the mapping process which included those tags not literally but semantically the same as the LCSH. The study was limited by the subjective nature of the matching process and by the small size of the sample.

Despite the high number of tags matched with the LCSH investigation into the semantic groups of tags left unmatched showed strong evidence of emerging subject terms. If a larger sample had been used it might have been possible to assess whether there were any trends in terms of the kinds of subject terms that were ‘suggested’ in more than one example the emerging term referred to the ‘family’ or ‘emotions’ where the LCSH had no mention of them. Picture book studies such as the kind carried out by Colomer into the changing nature of picture books over time relating to societal changes might be reflected in this disconnect between the stable terms both matched and unmatched within the booksonomies- certainly it might appear from this small sample that the ‘emotional’ and ‘social’ interpretations of picture book ‘value’ for adults are reflected in these terms.

4. To consider whether the booksonomy of a picture book might be a relevant source of data for further research into both social tagging and picture books.

Picture book booksonomies investigated within this study displayed a stable consensus of terms and definite trends in terms of tag category patterns- especially in terms of the ‘use’ concept. This suggests that they might be a rich source of data for the investigation of picture book users and taggers as well as the picture books themselves.

Further research could be carried out using a similar category model to better compare the picture book sample with other samples of fictional works- in this way it would be clearer to see whether the ‘*use*’ category is peculiar to picture books in particular or whether it is actually a specific and general fiction tag category that had previously been ‘lost’ within non-intrinsic conceptual categorisation processes or in other tag groups e.g. ‘personal task based’.

*5. To consider the practical implications of the research for cataloguers and designers of information retrieval systems that contain picture books*

Spiteri said that “The catalogue is a critical bridge between a library and members of its community” (2012, p.211) This study has revealed a number of disconnects between what is revealed about the ‘user warrant’ and habits of the picture book ‘using’ community and the formal taxonomies and entry points into the majority of library management systems. A number of practical suggestions for fields suggested by this research is provided in the discussion chapter including ‘reading level’, ‘named product or programme’ and a field for key words denoting ‘*use*’ rather than ‘*subject*’.

The evidence of emerging subject headings within the non-matched ‘*subject*’ tag categories is further evidence in support of the development of augmented formal and non-formal taxonomies within information retrieval systems. Dilger & Thompson (2008) write in their book about radical cataloguing that the catalogue could become a ‘discursive space’ in the future- perhaps this could be a space in which marginalised picture books are finally valued for the rich and complex ‘work’ that they do and the ‘places’ (they) go’.

## References

- Adler, M. (2009). Transcending library catalogs: a comparative study of controlled terms in Library of Congress Subject Headings and user-generated tags in LibraryThing for transgender books. *Journal of Web Librarianship*, 3(4), pp.309-331.
- Ames, M. and Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. *CHI*, April-May.
- Angeletou, S., Sabou, M., Specia, L. and Motta, E. (2007). *Between Folksonomies and the Semantic Web: An Experience Report*. Milton Keynes: The Open University.
- Bartley, P. (2009). Book tagging on LibraryThing: how, why, and what are in the tags? *Proceedings of the American Society of Information Science and Technology*, 46, pp.1–22. doi: 10.1002/meet.2009.1450460228 [Accessed 10 Mar. 2016].
- Bates, J. and Rowley, J. (2011). Social reproduction and exclusion in subject indexing. *Journal of Documentation*, 67(3), pp.431-448.
- Beghtol, C. (1986). Semantic validity: concepts of warrant in bibliographic classification systems. *Library Resources and Technical Services*, 30(2), pp.109-125.
- Beghtol, C. (2005). Ethical decision-making for knowledge representation and organization systems for global use. *Journal of the American Society for Information Science and Technology*, 56(9), pp.903-912.
- Buckland, M. (2012). Obsolescence in subject description. *Journal of Documentation*, 68(2), pp.154-161.
- Cantador, I., Konstas, I. and Jose, J. (2011). Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1), pp.1-15.
- Colomer, T., Kummerling-Meibauer, B. and Silva-Diaz, C. (2010). (Eds.) *New Directions in Picturebook Research*, 1st ed. Abingdon: Routledge
- Colomer, T. (2010). Picture books and the changing values at the turn of the century. In: T. Colomer, B. Kummerling-Meibauer and C. Silva-Diaz, (Eds), *New Directions in Picturebook Research*, 1st ed. Abingdon: Routledge, pp.41-54.

- Colomer, T., Kummerling-Meibauer, B. and Silva-Diaz, C. (2010). Introduction: Current trends in picturebook research. In: T. Colomer, B. Kummerling-Meibauer and C. Silva-Diaz, (Eds), *New Directions in Picturebook Research*, 1st ed. Abingdon: Routledge, pp.1-10.
- Conduit, N. and Rafferty, P. (2007). Constructing an image indexing template for The Children's Society. *Journal of Documentation*, 63(6), pp.898-919.
- Cutter, C. (1904). *Rules for a Dictionary Catalog*. 4th ed. [ebook] Washington: Government Printing Office. Available at: <https://archive.org/details/rulesforadictio06cuttgoog> [Accessed 26 Mar. 2016].
- Denscombe, M. (2003). *Good Research Guide for Small-scale Social Research Projects*. 2. ed.. Maidenhead, Phil.: Open University Press.
- Font, F., Serra , J. and Serra, X. (2013). Folksonomy-based tag recommendation for collaborative tagging systems. *International Journal on Semantic Web and Information Systems*, 9(2), pp.1-30.
- Furner, J. (2007). User tagging of library resources: Toward a framework for system evaluation. In: *World library and information congress: 73<sup>rd</sup> IFLA General Conference and Council*. [online] Durban: WLIC. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.109.5202&rep=rep1&type=pdf> [Accessed 8 Aug. 2015].
- Golder, S. and Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), pp.198-208.
- Guy, M. and Tonkin, E. (2006). Folksonomies. *D-Lib Magazine*, 12(1).
- Guyot, A. (2013). *Understanding Booksonomies: how and why are book taggers tagging?* MSc. Aberystwyth. [online] Available at: <http://hdl.handle.net/2160/12099> [Accessed 22 Sept. 2015].
- Halpin, H., Robu, V. and Shepherd, H. (2007). The complex dynamics of collaborative tagging. In: *Proceedings of the 16th international conference on World Wide Web*. [online] New York: USA Available at: DOI: 10.1145/1242572.1242602 [Accessed 20 Jan 2015].

- Heckner, M., Muhlbacher, S. and Wolf, C. (2008). Tagging tagging. Analysing user keywords in scientific bibliography management systems. *Journal of Digital Information*. 9(2)
- Heymann, P. and Garcia-Molina, H. (2009). Contrasting controlled vocabulary and tagging do experts choose the right names to label the wrong things?. *Second ACM International Conference on Web Search and Data Mining (WSDM 2009)*, February 9-13, 2009. Barcelona: Spain. Available at: <http://ilpubs.stanford.edu:8090/955/1/cvuv-lbrp.pdf> [Accessed 12 Jan 2015].
- Hunt, P. (1999). *Understanding children's literature*. London: Routledge.
- Iyer, H. and Bungo, L. (2011). *An examination of semantic relationships between professionally assigned metadata and user-generated tags for popular literature in complementary and alternative medicine*. [online] Informationr.net. Available at: <http://www.informationr.net/ir/16-3/paper482.html> [Accessed 14 Apr. 2015].
- Ke, H. and Chen, Y. (2012). Structure and pattern of social tags for keyword selection behaviors. *Scientometrics*, 92, pp.43-62.
- Kipp, M. (2007). *@toread and cool: tagging for time, task and emotion*. Paper presented at Information Architecture Summit 2007, Las Vegas: USA. Available at: <http://eprints.rclis.org/9868/1/mkipp-iasummit2007.pdf> [Accessed 12 Apr 2015]
- Kipp, M. (2010). User, author and professional indexing in context: an exploration of tagging practices on CiteULike. *Canadian Journal of Information and Library Science*. 35(1)
- Laspa, C. (2013). The normal distribution in LibraryThing: investigation on the combined literature of fifteen books. *Collection Building*, 32(4), pp.145-153.
- Lawson, K. (2009). Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers. *The Journal of Academic Librarianship*, 35(6), pp.574-582.
- Library of Congress, (n.d.). *Library of Congress Children's Subject Headings - LC Linked Data Service (Library of Congress)*. [online] Id.loc.gov. Available at: <http://id.loc.gov/authorities/childrensSubjects.html> [Accessed 11 Aug. 2015].

- Library of Congress (n.d). *Conduct of life - LC Linked Data Service (Library of Congress)*. [online] Available at:  
<http://id.loc.gov/authorities/childrensSubjects/sj96005146.html> [Accessed 25 Mar. 2016].
- Librarything.com. (2016). *LibraryThing for Libraries*. [online] Available at:  
<https://www.librarything.com/forlibraries> [Accessed 1 Apr. 2016].
- Lu, C., Park, J. and Hu, X. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of Information Science*, 36(6), pp.763-779.
- Marlow, C., Naaman, M. and Davis, M. (2006). HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead. In: *Proceedings of the seventeenth conference on Hypertext and hypermedia*. , [online] pp. 31-40 Odense: Denmark. Available at:  
<http://pagesperso.lina.univ-nantes.fr/~prie-y/ens/06-07/M2R/articles/p31-marlow.pdf> [Accessed 8 Aug 2015].
- Mathes, A. (2004). *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. [online] Adammathes.com. Available at:  
<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html> [Accessed 9 Aug. 2015].
- Morrison, P. (2007). Why Are They Tagging, and Why Do We Want Them To?. *Bulletin of the American Society for Information Science and Technology*, October/November. , [online] 31(1) Available at: <http://asis.org/Bulletin/Oct-07/morrison.html> [Accessed 12 Jan. 2016]
- National Information Standards Organization, (2003). *Guidelines for the construction, format, and management of monolingual thesauri*. [online] Maryland: NISO Press. Available at: [http://www.niso.org/apps/group\\_public/download.php/12591/z39-19-2005r2010.pdf](http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf) [Accessed 10 Nov. 2015]
- Overell, S., Sigurbjörnsson, B. and van Zwol, R. (2009). Classifying Tags using Open

- [Content Resources.\[online\]](#) *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. Barcelona: Spain. Available at: <http://www.wsdm2009.org/papers/p64-overell.pdf> [Accessed 13 Feb 2016]
- Pejtersen, A. (1979). The meaning of ‘about’ in fiction indexing and retrieval. *Aslib Proceedings*. [online] 31(5), pp.251-257. Available at: <http://dx.doi.org/10.1108/eb050682>[Accessed 10 Jan 2016]
- Peters, I. (2009). *Folksonomies*. Berlin: De Gruyter/Saur.
- Peters, I. and Weller, K. (2008). Tag gardening for folksonomy enrichment and maintenance. *Webology*, [online] 5(3), p. 58. Available at: <http://www.webology.org/2008/v5n3/a58.html> [Accessed 14 Apr. 2015].
- Peterson, E. (2008). Parallel Systems: The Coexistence of Subject Cataloging and Folksonomy. *Library Philosophy and Practice*, [online] Paper 179. Available at: <http://digitalcommons.unl.edu/libphilprac/179> [Accessed 17 Apr. 2015].
- Pirmann, C. (2012). Tags in the Catalogue: Insights From a Usability Study of LibraryThing for Libraries. *Library Trends*, 61(1), pp.234-247.
- Ransley, A. (1987). Towards a fiction index: part one. *The Australian Library Journal*, 36(1), pp.44- 54.
- Ransom, N. and Rafferty, P. (2011). Facets of user-assigned tags and their effectiveness in image retrieval. *Journal of Documentation*,[online] 67(6), pp.1038-1066. Available at: <http://dx.doi.org/10.1108/00220411111183582> [Accessed 15 Apr. 2015]
- Ranta, J. (1992). The New Literary Scholarship and a Basis for Increased Subject Catalog Access to Imaginative Literature. *Cataloging & Classification Quarterly*, 14(1), pp.3-26.
- Santos-Neto, E., Ripeanu, M. and Iamnitchi, A. (2007). *Tracking Usage in Collaborative Tagging Communities*. In: *Proceedings of International ACMIEEE Workshop on Contextualized Attention Metadata: personalized access to digital resources*. Vancouver: Canada. Available at: [http://www.csee.usf.edu/~anda/papers/CAMA07\\_ready\\_v2.pdf](http://www.csee.usf.edu/~anda/papers/CAMA07_ready_v2.pdf) [Accessed 12 Apr.



2015]

- Sauperl, A. (2012). Pinning down a novel: characteristics of literary works as perceived by readers. *Library Review*, 61(4), pp.286-303.
- Shatford, S. (1986). Analyzing the subject of a picture; a theoretical approach. *Cataloguing and Classification Quarterly*, 6(3), pp.39-62.
- Shirky, C. (2005). *Shirky: Ontology is Overrated -- Categories, Links, and Tags*. [online] Shirky.com. Available at: [http://www.shirky.com/writings/ontology\\_overnated.html?goback=.gde\\_1838701\\_member\\_179729766](http://www.shirky.com/writings/ontology_overnated.html?goback=.gde_1838701_member_179729766) [Accessed 8 Aug. 2015].
- Sipe, L. (1998). How Picture Books Work: A Semiotically Framed Theory of Text-Picture Relationships. *Children's Literature in Education*, 29(2), pp.97-108.
- Smith, G. (2008). Tagging: Emerging Trends. *Bulletin of the American Society for Information Science and Technology*, August/September, pp.14-17.
- Smith, T. (2007). Cataloging and you: Measuring the efficacy of a folksonomy for subject analysis. In: *Proceedings 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*. [online] Milwaukee: American Society for Information Science and Technology. Available at: <http://dlist.sir.arizona.edu/2061/> [Accessed 17 Apr. 2015].
- Thomas, M., Caudle, D. and Schmitz, C. (2009). To tag or not to tag?. *Library Hi Tech*, 27(3), pp.411-434.
- Thomas, M., Caudle, D. and Schmitz, C. (2010). Trashy tags: problematic tags in LibraryThing. *New Library World*, 111(5/6), pp.223-235.
- Tourné, N. and Godoy, D. (2012). Evaluating tag filtering techniques for web resource classification in folksonomies. *Expert Systems with Applications*, 39(10), pp.9723-9729.
- Uddin, M., Duong, T., Nguyen, N., Qi, X. and Jo, G. (2013). Semantic similarity measures for enhancing information retrieval in folksonomies. *Expert Systems with Applications*, 40(5), pp.1645-1653.

- Vander Wal, T. (2007). *Folksonomy :vanderwal.net*. [online] Vanderwal.net. Available at: <http://www.vanderwal.net/folksonomy.html> [Accessed 8 Aug. 2015].
- Voorbij, H. (2012). The value of LibraryThing tags for academic libraries. *Online Information Review*, 36(2), pp.196-217.
- Wallace, D. (2007). *Knowledge Management: Historical and Cross-disciplinary Themes (Libraries Unlimited knowledge management series)*. Santa Barbara: Libraries Unlimited Incorporated.
- Wordle.net. (n.d.). *Wordle - Beautiful Word Clouds*. [online] Available at: <http://www.wordle.net/> [Accessed 25 Mar. 2016].
- Yi, K. (2010). A semantic similarity approach to predicting Library of Congress subject headings for social tags. *Journal of the American Society for Information Science and Technology*, p.n/a-n/a.
- Yi, K. (2010). A semantic similarity approach to predicting Library of Congress subject headings for social tags. *Journal of the American Society for Information Science and Technology*, 61(8), pp.1658-1672.
- Yi, K. and Mai Chan, L. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation*, 65(6), pp.872-900.

## Bibliography

- Bell, J. (1987). *Doing your research project*. Milton Keynes: Open University Press.
- Brown, A. and Dowling, P. (1998). *Doing research/reading research*. London: Falmer Press.
- Brown, M. and Hurd, C. (1947). *Goodnight moon*. New York: Harper.
- Carle, E. (1987). *The very hungry caterpillar*. New York: Philomel Books.
- Denscombe, M. (2003). *Good research guide for small-scale social research projects*. 2. ed. Maidenhead, Phil.: Open University Press.
- Freeman, D. (1990). *Corduroy*. New York: Puffin Books.
- Kakali, C. and Papatheodorou, C. (2010). Exploitation of folksonomies in subject analysis. *Library & Information Science Research*, 32(3), pp.192-202.
- Martin, B. and Carle, E. (1992). *Brown bear, brown bear, what do you see?* New York: Henry Holt.
- McCloskey, R. (1941). *Make way for ducklings*. New York: Viking Press.
- Noruzi, A. (2007). Editorial: Folksonomies: Why do we need controlled vocabulary?. *Webology*, [online] 33(4). Available at: <http://www.webology.org/2007/v4n2/editorial12.html> [Accessed 17 Apr. 2015].
- Numeroff, L. and Bond, F. (1985). *If you give a mouse a cookie*. New York: Harper & Row.
- Proquest.com. (n.d). *LISA: Library and Information Science Abstracts - LISA: Library and Information Science Abstracts*. [online] Available at: <http://www.proquest.com/products-services/lisa-set-c.html> [Accessed 1 Apr. 2016].
- Rafferty, P. and Murphy, H. (2015). "Is there nothing outside the tags?: Towards a poststructuralist analysis of social tagging". *Journal of Documentation*, 71(3).
- Richards, A. and Sen, B. (2013). An investigation into the viability of LibraryThing for promotional and user engagement purposes in libraries. *Library Hi Tech*, 31(3),

pp.493-519.

Scieszka, J. and Smith, L. (1989). *The true story of the [three] pigs*. New York: Viking Kestrel.

Sendak, M. and Sendak, M. (1963). *Where the wild things are*. New York: Harper & Row.

Seuss, (1990). *Oh, the places you'll go!* New York: Random House.

Statistics.laerd.com. (n.d). *Mean, Mode and Median - Measures of Central Tendency - When to use with Different Types of Variable and Skewed Distributions | Laerd Statistics*. [online] Available at: <https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php> [Accessed 24 Mar. 2016].

Viorst, J. and Cruz, R. (1972). *Alexander and the terrible, horrible, no good, very bad day*. New York: Atheneum.

Willems, M. (2003). *Don't let the pigeon drive the bus*. New York: Hyperion

## **Appendix**

### **Appendix A: Final ‘Aboutness’/connotative group categories used during data analysis**

#### **Tag category: Subject**

Definition: a tag that refers to an actual character either within the words or pictures, objects, any themes that are explicitly mentioned within the book

Examples: *butterfly, eating, transformations*

#### **Tag category: Literary**

Definition: arising from the lexical or literary analysis of the text or narrative

Example: *metaphors, phonemes, circular story*

#### **Tag category: Use: Pedagogical: Formal**

Definition: a tag that implies use in formal education or to teach something formally

Example: *sequencing, inferring, predicting*

#### **Tag category: Use : Pedagogical: Formal: Named product**

Definition: a tag that refers to a named educational product or programme

Example: *FIAR, Sonlight, Common Core*

#### **Tag category: Use : Pedagogical: Formal: Levels**

Definition: a tag referring to graded reading books used to teach children to read

Example: *AR 3.7, Level K, DRA 40*

#### **Tag category: Use: Pedagogical: Informal**

Definition: a tag that implies informal ‘use’ of the book with others

Example: *read aloud, storytime, shared reading*

### **Tag category: Use: Pedagogical: Socialisation**

Definition: tags that imply ‘social teaching’ and that are ‘used’ as part of the socialisation process but are not explicit within the text.

Example: *social emotional, emotional literacy, morals*

## **Appendix B: Examples of ‘difficult’ tags and how they were interpreted**

### **Semantic clues**

There was a strict consideration of the semantic clues given by the form of the words applied so that the context (therefore category) of each tag could be understood.

For example:-

‘*predictable*’ = literary tag

‘*prediction*’ and ‘*prediciting*’ = use tag

In “Pigeon” which features a pigeon attempting to persuade the reader to “let him drive the bus” the tags

‘*persuade*’ and ‘*persuading*’ = subject tag

‘*persuasive text*’ and ‘*persuasive writing*’ = literary category because the pigeon was ‘speaking’ not writing so these tags refer to the text of the book and not the subject of the book’s narrative.

### **Contextual analysis**

One example of a group of tags that were difficult to position within the structure were those applied to “Pigs” describing it as a “twisted fairy-tale”. There were a number of these tags with a high level of frequency overall so a decision had to be made about whether these tags were describing a specific genre (in which case they should go in the ‘Isness’ group of categories) or whether this arose from the text specifically and was more of an interpretation whereby they should go in the ‘*literary*’ category in the ‘Aboutness’ group. After some thought these tags were placed within the ‘*genre*’

category as it was felt that logically they represented more of a factual statement about the book than an interpretative one.

## **Appendix C: LCSH/LibraryThing three stage mapping process details**

### **1<sup>st</sup> Level**

A straight and exact match

e.g. '*metamorphosis*' to '*metamorphosis*', what Heymann & Garcia-Molina called an 'exact' match (2009, p. 1)

### **2<sup>nd</sup> Level**

Semantically and almost syntactically the same, misspellings, plurals, direct variations, foreign language translations.

Heymann & Garcia-Molina (2009) used a more strict definition of this level- what they called 'almost exact' – "if the LCSH is modified to remove parenthetical remarks, swap the ordering of the words around a comma, stem or add or remove an 's' (2009, p. 2). In this study the second level included foreign language translations of the same LCSH, misspellings and some terms that you could argue were closer than just a 'broader' or 'narrower' term. e.g. '*metamorphosis*' to '*metamorphism*', '*metamorphois*' and '*metamorphosis-fiction*'.

### **3<sup>rd</sup> Level**

Basic level variation and related terms.

As Yi and Chan explain "polysemy can be a barrier to finding semantics, collecting more terms closely related to subject headings might help tackle the task" (2009, p. 1663) As folksonomies are free and democratic and have no 'authority terms' a more precise way of assessing whether or not the LCSH reflect the 'subject' perceived by the picture book taggers is to compare the wider range of semantically relevant, broader and narrower terms as well. E.g. '*metamorphosis*' to '*transformation*', '*changing*', '*growing*'

**Appendix D: number of matches (Fr)**

	Total at 1st	Total by 2nd	Total by 3rd	Not matched
<b>Pigeon</b>	300	802	1666	247
<b>Caterpillar</b>	1504	2147	4559	5038
<b>G'Night Moon</b>	811	913	1508	1035
<b>Brown Bear</b>	978	1807	2420	214
<b>Pigs</b>	103	868	1146	225
<b>Corduroy</b>	540	593	1438	1400
<b>Alexander</b>	83	91	1260	1656
<b>Ducks</b>	738	1753	2721	74
<b>Wild Things</b>	1618	1747	3968	1648
<b>Mouse</b>	285	689	1004	1181

**Appendix E: number of matches % (Fr)**

Book/Sample	Total at 1st	Total by 2nd	Total by 3rd	Not matched
Alexander <i>n</i> =2916	2.9%	3.1%	43.2%	56.8%
Mouse <i>n</i> =2185	13.0%	31.5%	46.0%	54.1%
Caterpillar <i>n</i> =9597	15.7%	22.4%	47.5%	52.5%
Corduroy <i>n</i> =2838	19.0%	20.9%	50.7%	49.3%
G'Night Moon <i>n</i> =2543	31.9%	35.9%	59.3%	40.7%
Wild Things <i>n</i> =5616	28.8%	31.1%	70.7%	29.3%
Pigs <i>n</i> =1371	7.5%	63.3%	83.6%	16.4%
Pigeon <i>n</i> =1913	15.7%	41.9%	87.1%	12.9%
Brown Bear <i>n</i> =2634	37.1%	68.6%	91.9%	8.1%
Ducklings <i>n</i> =2795	26.4%	62.7%	97.4%	2.7%
<b>Mean Average</b>	<b>19.8%</b>	<b>38.1%</b>	<b>67.7%</b>	<b>32.3%</b>
<b>Median Average</b>	<b>17.4%</b>	<b>33.7%</b>	<b>65.0%</b>	<b>35.0%</b>



## Appendix F: Hungry Caterpillar new subject heading suggestions

<b>'Lifecycle' Tags</b>	<b>No. of tags (Fr)</b>
life cycles	85
butterfly life cycle	16
life cycle of a caterpillar	7
butterfly cycle	3
cycle of life	6
cycles	3
butterfly's life stages	3
insect life cycle	2
process of life	2
life cycle	265
<b>Total</b>	<b>392</b>

<b>'Food/Eating' Tags</b>	<b>No. of tags (Fr)</b>
food	435
eating	92
hungry	92
hunger	31
foods	21
eat	12
overeating	4
veel eten ('eat a lot' in Dutch)	4
Essen ('food' in German)	2
diet	2
food: babies	2

different foods	2
full at last	2
fat	2
<b>Total</b>	<b>703</b>

<b>Counting Tags</b>	<b>No. of tags (Fr)</b>
counting	480
numbers	87
counting book	12
tellen ('count' in Dutch)	8
count	3
Numbers & Counting	2
<b>Total</b>	<b>592</b>

<b>Days of the week Tags</b>	<b>No. of tags (Fr)</b>
days of the weeks	3
Counting Days of Week	4
days of the week	340
days	15
weekdays	9
dagen	4
dagen van de week	4
week	3
week days	4
<b>Total</b>	<b>386</b>

### Appendix G: Corduroy new subject heading suggestions

Subject area	Tags (Fr)	% of total tags
Friendship/Love	613	21.60%

Friendship Tags	No. of tags (Fr)
friendship	348
love	97
friends	86
acceptance	30
friend	22
belonging	22
human-animal relationships	3
Relationships & Families	3
companionship	2
care	2
<b>Total</b>	<b>615</b>

### Appendix H: Alexander new subject heading suggestions

Subject area	Tags (Fr)	% of total tags
Bad days/Hard times	575	19.72%
Families	492	16.87%

Bad Day Tags	No. of tags (Fr)
bad day	341
bad days	145
bad	21

bad luck	20
day	17
trouble	5
hard times	4
accidents	3
terrible day	3
having a bad day	3
Murphy's Law	3
accident	3
adversity	3
not every day is a good day	2
when things go bad	2
<b>Total</b>	<b>575</b>

<b>Families Tags</b>	<b>No. of tags (Fr)</b>
family	230
children	169
siblings	26
family life	17
brothers	8
brothers and sisters	8
friends and family	8
relationships	5
parents	5
Family Members	5
sibling rivalry	4
brothers and sister	3
Mom	2
sisters	2
<b>Total</b>	<b>492</b>

### Appendix I: Where the Wild Things Are new subject heading suggestions

Subject area	Tags (Fr)	% of total tags
Travel/Adventure/Journey	526	9.37%
Family relationships	257	4.58%
Emotions	187	3.33%

Travel Tags	No. of tags (Fr)
adventure	430
journey	28
travel	14
independence	9
adventures	8
running away	6
journeys	5
escape	5
exploration	5
adventurous	3
run away	3
discovery	2
exploring	2
leaving home	2
traveling	2
quest	2
<b>Total</b>	<b>524</b>

Family Tags	No. of tags (Fr)
family	108
Max	45
friendship	29
friends	16
parents	14
mother	10
relationships	9
Mom	5

mothers	4
mothers	4
family relationships	3
Mama	2
people	2
son	2
friends and family	2
mother's love	2
<b>Total</b>	<b>257</b>

<b>Emotions Tags</b>	<b>No. of tags (Fr)</b>
feelings	37
love	35
anger	32
emotions	27
courage	13
respect	9
fear	7
angry	6
loneliness	6
fears	6
lonely	5
frustration	2
comfort	2
<b>Total</b>	<b>187</b>