

The wisdom of the cataloguers: LCSH, indexer inconsistencies and collective intelligence

Gary Steele

Glasgow Caledonian University

Abstract

The process of subject analysis and Library of Congress Subject Heading assignment is, despite the availability of rules and standards, a subjective one. Disagreements and inconsistencies between cataloguers regarding the correct Library of Congress Subject Headings for a given resource are widespread. This paper attempts to address the problem of these indexer inconsistencies by utilising the wisdom of the crowd. The various headings suggested by different cataloguers, for a particular resource from a large number of library catalogues, can be collated to create a coherent, valid, and consistent set of Library of Congress Subject Headings that represent the collective wisdom of the cataloguers.

This paper seeks to address the issue of the inconsistent selection and application of Library of Congress Subject Headings (LCSH). Most articles about indexer inconsistencies concentrate on how they can be eliminated. I contend that these inconsistencies can never be eliminated and are entirely legitimate expressions of an individual cataloguer's interpretation of the rules and policies for the application of LCSH. Although the Library of Congress has attempted to produce objective rules and policies for determining the assignment of headings, the headings chosen by any cataloguer are the result of individual subjective decisions. These decisions ultimately lead to the phenomenon most commentators label indexer inconsistency. Basing my arguments on the theory of the wisdom of the crowd I propose a solution that re-evaluates the concept of indexer inconsistencies. If such inconsistencies are viewed as legitimate expressions of individual cataloguer's different opinions of the correct headings, then these suggestions can be aggregated to create a coherent, valid, and consistent set of LCSH that represents the collective wisdom of the cataloguers.

LCSH admittedly do have many failings, not least because they can be ethnocentric, exclusionist, culturally insensitive, and rife with gender bias. These problems cannot be dismissed lightly but they are beyond the scope of this paper and instead I wish to concentrate on the selection and assignment of LCSH. Although initially conceived and developed exclusively for use in the Library of Congress, LCSH are now used in libraries all over the world. When confined to the Library of Congress, LCSH could maintain a unity of purpose and content which should be emulated in every library catalogue that uses LCSH. As a controlled indexing language and as a standard there are rules and principles for the use of LCSH to ensure the consistent application of headings. The principle of the user and usage is fundamental to the selection and application of LCSH. Charles Cutter wrote of the 'convenience of the public' (1904, p. 6) and David Haykin promoted the principle of the 'reader as a focus' (1951, p. 7). Both envisaged a coherent, definable 'public' that was the user of the library. This is clearly an untenable stance and writers such as Prevost (1946), Dunkin (1969), and Chan (2004) have highlighted the problem of even attempting to conceive of the users of a library as a discrete entity. As Lois Mai Chan (2004) has argued, if headings are not chosen to suit a pre-conceived audience or 'public', then headings should instead be chosen in a logical and consistent manner;

a manner the user can learn and understand. The same heading should be applied to the same book regardless of where it is catalogued and a strict interpretation of the LCSH rules implies that 'individual libraries shouldn't change the headings or interpret them to meet local needs' (Broughton, 2004, p. 104).

Clearly, libraries do change and interpret LCSH. Moreover, further inconsistencies arise in the process of subject heading indexing. The 1991 article 'Cataloging must change!' by Dorothy Gregor and Carol Mandel argued that indexer inconsistencies are inevitable and must be accepted, stating that 'only a reasonable degree of inter-indexer consistency can ever be achieved' (1991, p. 46). From their meta-analysis of indexer inconsistency studies Gregor and Mandel concluded that the degree of indexer consistency is as low as 'ten to twenty percent'. Thomas Mann published a strong rebuttal to Gregor and Mandel's article in *Cataloging and Classification Quarterly* in 1997 arguing that only a structured vocabulary control mechanism such as LCSH can overcome and resolve issues of inconsistency. Mann insisted that provided cataloguers have full knowledge and competency with LCSH there is no reason they should not always arrive at the correct heading every time. According to Mann, the rule of specific entry functions as a means of choosing between a set of potentially appropriate headings and creates a situation of 'objective right and wrong in subject heading assignment' (1997, p. 30). But do indexer inconsistencies reflect the extent to which cataloguers apply the 'wrong' headings or do they represent genuine disagreement about which 'right' heading is most appropriate? David Bade has pointed out that most studies of indexer inconsistencies are counts of 'indexer consistency and not of appropriateness' (as cited in Hjørland, 2005, p. 146) and are thus ultimately pointless because they are entirely self-referential. In order to assess the process they are examining, that is to say the assignment of subject headings, the researcher must first perform the same task and decide which heading is 'correct'. The researcher's judgement of which heading is the correct one is itself a subjective decision. To put it another way, the process of creating and selecting LCSH can be judged objectively but the headings actually assigned can only be judged subjectively.

LCSH are ultimately a means of indicating aboutness. The aboutness of a document is just that; what it is about, the subject, the topic. Aboutness is, as Birger Hjørland defines it, 'that "something" that subject analysis and retrieval are supposed to identify' (2011, para. 1). The concept of aboutness in relation to information science was first proposed by R. A. Fairthorne in the article 'Content analysis, specification and control' published in 1969. Fairthorne drew a distinction between the author's intended aboutness and the aboutness actually expressed in the document. Building on Fairthorne's concept, authors including Maron (1977), Hutchins (1978), Begthol (1986), and more recently Birger Hjørland (2001) have developed a further epistemological distinction between the objective and subjective view of aboutness. The objective or document-orientated approach states that indexing should summarise the aboutness of a document in itself while the subjective or request-orientated approach argues indexing should reflect the aboutness of the document in relation to the user and how they may request the document. LCSH are designed for the objective, document-orientated approach to determining aboutness. Library of Congress practice states that a cataloguer should 'assign headings based on an analysis of the contents of the work being cataloged' (2011, Library of Congress Policy and Standards Division [PSD], rule 2). However, this analysis can never be truly objective as it is mediated through a subjective agent, that is to say, the cataloguer. Hence any evaluation of appropriateness of subject headings can only ever be relative. The problem lies not in judging the truth or falsehood of a heading, but rather in judging if the heading is the 'right' one for the situation. The art of subject heading assignment is the art of finding the 'fittest' heading for a particular item. Fittest here is used in the sense of Darwin's (by way of Herbert Spencer) aphorism 'survival of the fittest', meaning the correct heading is the one that 'fits' the criteria for the assignment of LCSH. Thus it is more useful to discuss LCSH in degrees of appropriateness rather than the dichotomy true or false.

The general rule for assigning LCSH states that a cataloguer should assign 'one or more subject headings that best summarize the overall contents of the work and provide access to its most important topics' (2011, PSD, rule 1). Important topics are defined as those that 'comprise at least 20% of the work'. There are two specific principles, rule 4 'specificity' and rule 14 'objectivity', detailed in section H180 'Assigning and

Constructing Subject Headings' of Library of Congress's *SHM: Subject Headings Manual* that I believe may account for some of the indexer inconsistency when applying LCSH. Rule 4 'specificity' states that headings should be as 'specific as the topics they cover' but that 'specificity is not a property of a given subject heading; instead, it is a relative concept that reflects the relationship between a subject heading and the work to which it is applied' (2011, PSD). In other words, a heading such as Sociology, in its own terms, is a broad heading but is specific when applied to an introductory text on sociology. Rule 14 deals with the concept of 'objectivity' and sets out the Library of Congress's concept of objectivity in relation to subject analysis. To be objective, headings should not 'express personal value judgements regarding topics or materials' (2011, PSD). The rule acknowledges that a cataloguer's personal knowledge and judgement influence the process of subject analysis but clearly states that 'headings should not be assigned that reflect a cataloguer's opinion about the contents', rather cataloguers must 'consider the intent of the author or publisher and, if possible, assign headings for this orientation without being judgmental' (2011, PSD). Cataloguers are instructed to consider the authors' and publishers' intent for the aboutness of the item being analysed and assign appropriate, specific headings. Any heading chosen can only ever reflect the cataloguer's subjective opinion of its specificity in relation to the author's intent. Inconsistencies arise because there can be a number of equally valid and appropriate responses to the subjective evaluation of LCSH. Attempts to use LCSH as a means of addressing the subjective, request-orientated approach to determining aboutness contravene the Library of Congress's rules of specificity and objectivity. That said, even if the headings are chosen using the objective document-orientated approach, the absence of any objective means to select the 'right' LCSH from a group of equally valid headings still pervades.

Individually cataloguers may not be able to overcome the subjectivity of heading assignment but James Surowiecki's 2004 book *The Wisdom of Crowds* provides a potential means of choosing the most appropriate LCSH for any resource. Surowiecki's theory is based on a short article by Francis Galton - 'Vox Populi' - published in the journal *Nature* in 1907. While attending a 'fat stock and poultry exhibition' held in Plymouth, Galton observed the results of a weight-judging competition. Some 800 people entered the competition to guess the weight of a 'fat ox' and Galton calculated that the mean of the crowd's guesses was closer to the ox's actual weight than any one individual guess. Expanding on Galton's initial observations Surowiecki formulated the theory of the wisdom of the crowd arguing that when our 'imperfect judgement [is] aggregated in the right way, our collective intelligence is often excellent' (2004, p. XIV). This collective intelligence is the product of many individual choices, but the wisdom of the crowd is not aggregated guess work as all participants must have some knowledge of the matter at hand. Surowiecki identifies four additional factors that are required for crowd wisdom to work successfully and arrive at a wise decision. These are; diversity of opinion, independence of decision, decentralisation of expertise, and a method to summarise these individual decisions.

The process of subject heading determination already follows the first three factors identified by Surowiecki. Diversity of opinion evidently abounds with the assignment of LCSH. The indexing inconsistencies discussed prove there is diversity of opinion among cataloguers when assigning LCSH. The second factor, independence of decision, is more problematic as ideally participants in wise crowd decisions should not know their fellow participants choices. However, when assigning headings cataloguers' decisions, while guided, are not determined, by other cataloguers' opinions. Wise crowd decisions are also reliant on decentralisation which allows for individuals to draw on specialism and local knowledge. The power of decision must not reside in one central location and the use of LCSH by libraries around the world decentralises the decision making process. The first three criteria for the wisdom of the crowd to function successfully have been met by the use and application of LCSH in libraries throughout the world. The intellectual task of choosing the most appropriate subject headings has already been carried out by the cataloguers at these libraries. Taken together they represent years of knowledge and the broadest spectrum of cataloguers' opinions. The final factor required for a wise crowd decision is a means of summarising individual choices into a collective decision. By accessing the catalogues of these libraries and collecting the subject headings assigned it is possible to collate and choose the optimal group of LCSH for a particular

resource, the final step in the successful application of the wisdom of the crowd. For example, collating the headings from 20 random academic libraries from New Zealand, Australia, Canada, Ireland, the US and the UK for any edition of *The Order of Things: An Archaeology of the Human Sciences* by Michel Foucault returns 36 records with the following suggested headings shown in Table No. 1.

Headings	Number of occurrences
Learning and scholarship.	32
Civilization--History.	21
Postmodernism.	12
Knowledge, Theory of.	12
Learning and scholarship--Europe.	1

Table 1: Collated headings for *The Order of Things: An Archaeology of the Human Sciences* by Michel Foucault

All five of these heading are valid, authorised LCSH, but as Table No. 1 shows, only two of the headings – ‘Learning and scholarship’ and ‘Civilization--History’ occur in more than half of the records examined. Although not an objective method of assigning headings, by summarising individual cataloguers’ choices into a collective decision I argue that the number of occurrences for these two headings, which clearly demonstrate a consensus, can be deemed to represent the wisdom of the cataloguers.

The wisdom of crowds works most effectively when applied to defined, quantifiable values. The standard example cited as representing the wisdom of the crowd, guessing the weight of an ox, is predicated on known truths or absolutes. I have argued that there are no absolutes, rather degrees of appropriateness, when applying LCSH. Thus, utilising the wisdom of the crowd to select LCSH could potentially fall foul of the same self-referential problems David Bade has highlighted in relation to indexer inconsistencies studies. This could render the application of crowd wisdom to the selection of LCSH worthless but Thomas Kuhn's work on the philosophy of science addresses this very issue. In *The Structure of Scientific Revolutions* Kuhn argues that although objective criteria exist to enable scientists to choose between different theories, their choices are always in reality subjective individual choices. In the absence of any objective means to dictate the selection of the right LCSH from a group of equally valid headings a viable solution is to utilise the wisdom of the crowd for, to paraphrase Thomas Kuhn (1970), what better criterion could there be than the decision of the catalogue group?

The fundamental purpose of LCSH has remained unchanged since their inception and they are still a means, as Charles Cutter put it in the introduction to *Rules for a Dictionary Catalogue*, ‘to enable a person to find a book... to show what the library has... [and] to assist in the choice of book’ (1904, p. 12). By utilising the wisdom of the cataloguers it may be possible to select a set of LCSH that is superior to any set of LCSH suggested by any individual catalogue. The potential shown by harnessing the wisdom of the cataloguers demonstrates a means of selecting the ‘right’ LCSH for each resource in a library catalogue by exploiting the very indexing inconsistencies others have endeavoured to eradicate. Only by ensuring the accurate and consistent application of subject headings can LCSH hope to survive in the online environment where all their faults and errors are readily apparent. LCSH are a means of conveying to the reader the expert opinion of the catalogue essentially saying here are other resources I as a Librarian recommend on this topic. And, the truth is, for all their failings, LCSH are still the most comprehensive and authoritative controlled subject access language in existence, and are themselves a product of decades of cumulative collective wisdom.

Acknowledgements

The author wishes to acknowledge the advice and criticism of Dr Richard Espley, The Caird Library, National Maritime Museum, Greenwich, London, and the assistance of Patrick Comerford in the preparation of this paper.

References

- Beghtol, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42(2), 84-113.
- Broughton, V. (2004). *Essential classification*. London: Facet.
- Chan, L. M. (2005). *Library of Congress subject headings: Principles and application*. 4th edition. Westport, Conn.: Libraries Unlimited.
- Cutter, C. A. (1904). *Rules for a printed dictionary catalog*. 4th edition. Washington, D.C.: Government printing office.
- Fairthorne, R. A. (1969). Content analysis, specification and control. *Annual Review of Information Science and Technology*, 4, 73-109.
- Gregor, D & Mandel, C. (1991). Cataloging must change! *Library Journal*, 116 (6), 42-47.
- Haykin, D. J. (1951). *Subject headings: a practical guide*. Washington, D.C.: Library of Congress.
- Hjørland, B. (1998). Theory and metatheory of information science: a new interpretation. *Journal of Documentation*, 54(5), 606-621.
- Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. *Journal of the American Society for Information Science and Technology*, 52, 774-778.
- Hjørland, B. (2005). Empiricism, rationalism and positivism in library and information science. *Journal of Documentation*, 61(1), 130-155.
- Hjørland, B. (2011, May 5). *Aboutness*. Retrieved from <http://www.iva.dk/bh/core%20concepts%20in%20lis/articles%20a-z/aboutness.htm>.
- Hutchings, W. J. (1978). The concept of 'aboutness' in subject indexing. *Aslib Proceedings*, 30 (5), 172-181.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. 2nd edition. Chicago, IL: University of Chicago Press.
- Kuhn, T. S. (1972). *The Essential Tension*. Chicago, IL: University of Chicago Press.
- Library of Congress Policy and Standards Division. (2011, May 5). *SHM: Subject Headings Manual - General Topics - H180 Assigning and Constructing Subject Headings*. Retrieved from <http://desktop.loc.gov>.
- Mann, T. (1997). 'Cataloging must change'" and indexer consistency studies: misreading the evidence at our peril. *Cataloging and Classification Quarterly*, 23(3/4), 3-45.

Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science and Technology*, 28, 38-43.

Q & A with James Surowiecki. (2011, May 5). Retrieved from <http://www.randomhouse.com/features/wisdomofcrowds/Q&A.html>.

Shubert, S. B. (1992). Critical views of LCSH - ten years later: a bibliographic essay. *Cataloging and Classification Quarterly*, 15(2), 37-97.

Soll, J. B., Mannes, A. E., & Larrick, R. P. (2011, May 5). *The "Wisdom of Crowds" Effect*. Retrieved from <http://faculty.fuqua.duke.edu/~jsoll/Soll,%20Mannes,%20Larrick%202011.pdf>

Stone, A. T. (2000). The LCSH century: a brief history of the Library of Congress Subject Headings, and introduction to the centennial essays. *Cataloging and Classification Quarterly*, 29(1/2), 1-15.

Surowiecki, J. (2004). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. London: Little Brown.

Corresponding author:

Gary Steele can be contacted at gary.steele@gcu.ac.uk