

ARE SYSTEMATIC REVIEWS BETTER, LESS BIASED AND OF HIGHER QUALITY?

Boell, Sebastian K., University of New South Wales, SISTM, Quadrangle Building L2, Sydney, 2052 NSW, Australia, sebastian.boell@unsw.edu.au

Cezec-Kecmanovic, Dubravka, University of New South Wales, SISTM, Quadrangle Building L2, Sydney, 2052 NSW, Australia, dubravka@unsw.edu.au

Abstract

Literature reviews are central to any academic research. Whether journal article, conference presentation or research thesis it is necessary for all of them to provide an overview of earlier research in order to contextualize one's own findings. More fundamentally, the beginning of any research is crucially dependent on an appropriate literature review. In recent years a supposedly new form of literature review has emerged, so called, systematic reviews. The aim of this paper is to question the key premises of systematic reviews and demonstrate that the claims they are less biased and more rigorous than so called narrative reviews do not hold. This paper briefly introduces the origin of systematic reviews and explains how they are undertaken. Based on this introduction the paper shows that key premises of systematic reviews cannot be fulfilled and that they by no means guaranty the creation of 'better' literature reviews. In contrast, to systematic reviews which put importance on the literature identification and selection process, it argues that reading is central to reviewing literature. Reading enables academics to improve their understanding of the subject area and therefore to further advance their searches. Better literature reviews can only be achieved through better understanding of the subject area. Proper understanding of search techniques will then allow researchers to identify further relevant literature. Reviewing literature is therefore better described as a hermeneutic process.

Keywords: Literature reviews, systematic review, database searches, hermeneutic circle, hermeneutic literature review.

1 Introduction

All academic achievements are based on earlier achievements by others. Predecessors provide the foundation for one's own research, for example, by illuminating areas prospective for future research or through the development of methods and theories. This basic characteristic of academic work was succinctly put in the well known quote by Isaac Newton as: "if I have seen a little further it is by standing on the shoulders of giants".

This of course requires that one is familiar with earlier work of relevance to the phenomenon studied. Commonly this is done by reviewing earlier literature and presenting the most important aspects of such reviews to the research domain. Literature reviews are therefore important for both the conduct of research and the presentation of research findings as they aid readers in understanding how the research presented extends current knowledge of a phenomenon.

In recent years an approach to literature reviews named 'systematic review' has emerged. The term systematic review was proposed in the mid 1990s in the context of medical research (Chalmers and Altman, 1995) and has since spread via health informatics (e.g. Shiffman et al., 1999) into software engineering (Kitchenham, 2004) and only recently into information systems (IS) (e.g. Williams et al., 2009). While in medical research systematic reviews were initially closely associated with meta-analysis of earlier studies (e.g. Clarke and Stewart, 1995; Eysenck, 1995), later guides developed in software engineering (e.g. Kitchenham, 2004) seem to reinterpret systematic reviews mainly as 'replicable database searches' (e.g. Beecham et al., 2006).

As systematic reviews were developed outside IS it is important to ask if they can indeed help to improve literature reviews in IS research. There are several reasons for a critical investigation of systematic reviews. Firstly, the original context was quite narrow, looking at meta-analysis of studies on medical treatments and not at literature reviews in general. This raises the question to what extent the concept of systematic reviews is transferable to literature reviews in general. Secondly, medical research can be more readily associated with the 'hard' and 'exact' sciences, while the IS field is more closely associated with the 'softer' social sciences. It is therefore important to ask if the concept of systematic reviews can be easily and fruitfully adopted across domains of social sciences. And lastly, one has to look if the translation process of systematic reviews from medicine to other domains has led to distortions in the understanding of their execution and their capabilities.

This warrants to have a closer look at systematic reviews and especially its relevance and value for IS research. This paper aims to contribute to better understanding of systematic reviews, especially their limitations, and to offer a reflective critical discussion of their adoption in IS. Specifically it will: (1) provide an overview of the origin and assumptions of systematic reviews; (2) elaborate on the process of undertaking systematic reviews; and based on this (3) identify key problems that can be associated with systematic reviews. Finally, it will argue that (4) reviewing literature is fundamentally a hermeneutic process that can, therefore, be better described by referring to the hermeneutic circle.

2 The origin of systematic reviews

The phrase "systematic review" first started to appear in the mid 1990s and since then has gained momentum in being adopted for literature reviews (Figure 1). Before that the phrase only appeared in a hand full of publications in the context of classification of species in biology (e.g. Mees, 1957) and zoology (e.g. Alm, 1916). More specifically the origin of the phrase can be attributed to a meeting organised by the *British Medical Journal* and the *Cochrane Centre* in London in 1993 (Chalmers and Altman, 1995).

The members of this meeting were not satisfied with the way in which reviews were undertaken on earlier research in clinical medicine. They were concerned that reviews on clinical treatments lack the use of "scientific principles" potentially causing severe consequences for advancing medical treatments: "because reviewers have not used scientific methods, advice on some lifesaving therapies has been delayed for more than a decade, while other treatments have been recommended long after controlled research has shown them to be harmful" (Oxman and Guyatt, 1988 cited by Chalmers and Altman, 1995).

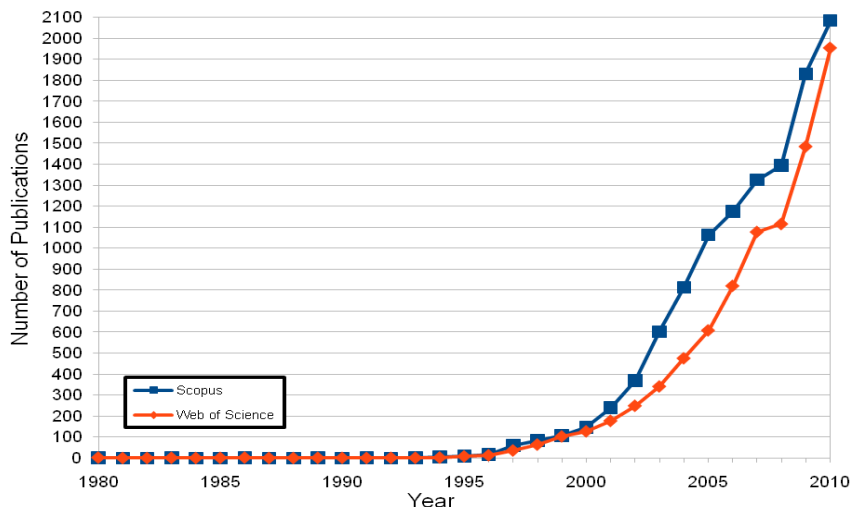


Figure 1. Number of publications in Web of Science and Scopus for the phrase “systematic literature review” OR “systematic review” in the title of publications; results are limited to publications classified as reviews.

The arguments provided for this dramatic judgement are sound and can be summed up as follows: (1) authors of literature reviews are selective in the literature they chose for their reviews; (2) studies on medical treatments differ in their research design and consequently in the quality of their findings; (3) studies can report different and inconclusive findings; and (4) clinical studies are expensive and generally use only a limited number of subjects (Mulrow, 1995).

Systematic reviews were therefore introduced as a way for overcoming these shortcomings. Consequently points elaborated by members of the 1993 meeting should be seen as central to systematic reviews. For example, in order to overcome bias resulting from selective coverage systematic reviews should aim to be as comprehensive as possible. Knipschild (1995) argued therefore that potential sources of bias arise if literature is selected on the basis of, for example, prestige of journals, authorship or language. Rather the main criterion has to be the quality of studies (Oxman, 1995). In the context of medical studies this means that randomized controlled trials are preferred (Clarke and Stewart, 1995). Inconsistencies in findings and limited number of subjects can then be mitigated by combining original results of all studies through meta-analyses (Eysenck, 1995; Thompson, 1995). Combining findings of individual studies increases the sample size and therefore the statistical power for estimating the effectiveness of treatments (Mulrow, 1995).

In conclusion the phrase “systematic review” originated as a label for a certain approach to undertaking literature reviews on earlier research on medical treatments closely associated with meta-analyses of earlier studies. The general aim is to uncover all relevant studies on a specific topic independently of source. After studies are identified only those meeting a minimum standard regarding the soundness of their research findings are included in further analysis. The effort necessary for undertaking systematic reviews is justified with potentially much higher costs and time requirement for additional clinical studies involving enough subjects to have sufficient statistical power.

3 The Process of undertaking systematic reviews

This section briefly introduces the systematic review process according to Kitchenham (2004), Kitchenham and Charters (2007) and Petticrew and Roberts (2006). While other guidelines for undertaking systematic reviews might slightly differ, the general approach for undertaking systematic reviews outside of medical research are similar.

As the label indicates systematic reviews follow a structured approach. The process of undertaking systematic reviews is therefore mostly linear with a sequence of steps following each other as described by Figure 2. Generally one step has to be finished before the review can move on. It is important to stress that proponents of systematic reviews believe that this introduces rigour into the literature review process helping to achieve “unbiased” reviews (Kitchenham and Charters, 2007) and “higher quality” reviews (Dybå and Dingsøy, 2008) that are “potentially reproducible” (Greenhalgh, 1997; Kitchenham and Charters, 2007).

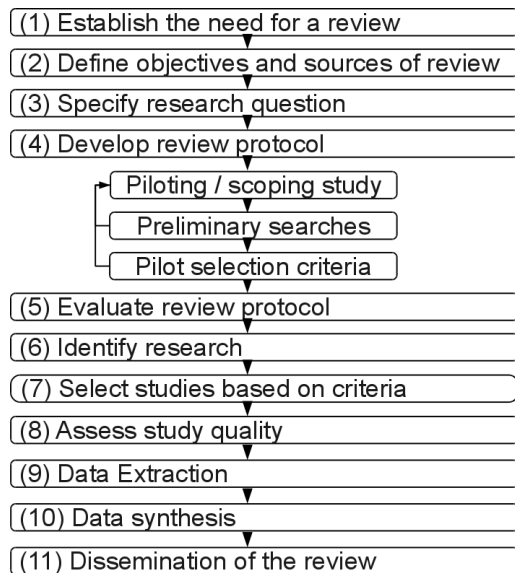


Figure 2. Overview of the systematic review process.

A more detailed description of each step can be found in Kitchenham (2004), Kitchenham and Charters (2007), or Petticrew and Roberts (2006). While, for the purpose of this paper, it is not necessary to elaborate in depth on each of the steps involved, it is important to note:

(1) The rationale for undertaking systematic review is similar to those of other literature reviews. Kitchenham and Charters (2007) list three main reasons for undertaking systematic reviews: to summarize results of earlier research; to identify gaps in current research; and to provide a framework for new research.

(2) The first different step in systematic reviews is then to define the objective of the review. This assumes that a researcher already has an appropriate understanding of the research domain which will allow defining research objectives upfront. In contrast, neophyte researchers are usually trained to develop research objectives after a research gap is identified. A task that usually can only be achieved *after* a literature review is undertaken.

(3) In line with classical research training the research question is developed out of the research objectives. Guidelines stress the importance of this step for systematic reviews: “Specifying the research question is the most important part of any systematic review ... [it] drives the entire systematic review” (Kitchenham and Charters, 2007:9).

(4) Based on the research question a review protocol is developed. This step is crucial as it locks in the rest of the review process. It defines the terms used for searching and specifies the criteria according to which the literature will be assessed. Because of the importance of the step an initial scoping of a study may be undertaken during which some preliminary searches can be undertaken. This will help to identify search terms, their abbreviations and synonyms. Documents found during this round may also be used for developing and assessing selection criteria.

(5) The finalized review protocol should then be assessed by external experts or in the case of PhD students their supervisors.

(6) According to the final review protocol searches are then taken out. Usually this involves the development of an elaborated search string which may be adjusted for different databases. In most cases brackets and the Boolean operators AND and OR are heavily used in order to include synonyms, alternative spellings and abbreviations of search terms (for an example see Beecham et al., 2006). This step is crucial to systematic reviews and the importance of the way searches are taken out in systematic reviews is stressed in guidelines: “the rigour of the search process is one factor that distinguishes systematic reviews from traditional reviews” (Kitchenham, 2004:7).

(7) Search results are then screened to select relevant literature. Usually this is done in descending order by first looking at titles, then abstracts and finally the full text. Here authors of guidelines indicate that the selection process may also involve “practical issues”. For example, literature may be excluded on the basis of language, journal, authorship or date (c.f. Kitchenham and Charters, 2007:19).

(8) All relevant studies that could be identified are then screened in order to assess their quality. Studies not meeting quality standards specified in the review protocol will be excluded. Important criteria for quality are the absence of any form of bias in the reported research (e.g. selection or sampling bias) as well as internal and external validity.

(9) Data is then extracted from all selected studies and (10) synthesized before the final review is disseminated (11).

Systematic reviews have since 2000 spread beyond the medical fields. While not disputing the arguments for systematic reviews in medical fields (especially metastudies) this paper questions their adequacy and claimed benefits for social sciences and in particular IS.

4 Critical review of the systematic review process

As proponents of systematic reviews make strong claims about their advantages it is important to investigate if such a process can indeed lead to reviews of “higher quality” (Dybå and Dingsøyr, 2008). Moreover, Kitchenham (2004, p.2) stated that “systematic reviews require considerably more effort than traditional reviews”. Consequently it is worthwhile to investigate if additional effort is likely to lead to an improvement of the outcome in the final review. This section is therefore critically looking at systematic reviews undertaken according to the guidelines developed outside of medicine such as the approach introduced in the previous section. Two central issues for systematic reviews are stressed above. On the one hand the importance of the research question (step 3) and on the other hand the importance of the search process (step 6). Both of these points will be critically discussed.

4.1 The role of the research question

As the formulation of the research question takes an important place in systematic reviews it is worthwhile to further investigate if this can have consequences for the research. As indicated above research questions are usually developed out of research objectives which are commonly established after a literature review successfully identified a gap in earlier research. In a way guidelines for systematic reviews put this process on its head. Moreover, systematic reviews work with a single research question rather than a set of related questions. The reason for this may be attributed to the origin of systematic reviews in medicine.

Medical research for which systematic reviews are undertaken can usually be well defined at the onset of the research. Medical researchers generally would know the treatment they are interested in as well as the population for which they are doing research. Subsequently, a reasonably closed research question can be formulated at the onset of the research. For example, finding *all* studies “using randomized controlled trials investigating the effect of vitamin C on the common cold”.

While the formulation of such a closed research question is possible for problems associated with medical treatments it is not possible for most problems faced in IS. Generally in social sciences research questions are much wider with unclear boundaries. Moreover, usually not all relevant theoretical lenses are known at the onset of a research. In such cases the benefit of using a systematic reviews approach is questionable. For example, finding *all* studies discussing “knowledge management in software development” seems a considerably more difficult task than the one described above.

It is important to stress that the aim of systematic reviews in medicine is unearthing *all* studies fitting a research question. Usually this involves painstaking work involving among others the consultation of multiple databases, using citation tracking techniques such as snowballing, as well as consulting other colleagues (Greenhalgh and Peacock, 2005). Some of the relevant studies may even be identified through sheer luck and therefore the way researchers came across the whole body of relevant literature is inherently not reproducible by others.

In addition to the type of questions that can be addressed by systematic reviews it is also important to note that for most research in IS it is simply not possible to lock in a research question before the research is well under way. Only when the problem space is reasonably well understood can researchers gain sufficient understanding of potentially relevant research questions. What seemed a worthwhile research question at the onset may turn out to be trivial at a later point. Sticking with a question initially formulated may stop researchers from exploring a problem from alternative angles

discovered at a later point. As MacLure (2005:399) puts it “diversions into unanticipated areas are not encouraged ... learning from adjacent areas is not recommended either”.

In brief, the role of research questions in systematic reviews seems problematic. First of all, systematic reviews can only be undertaken for a highly specific and reasonably narrow research questions. While the formulation of such questions is possible for problems in the natural sciences and medicine, it severely limits the scope of research that can be undertaken in the social sciences. In addition, systematic reviews require the formulation of a research question early on. In most cases however, reasonable research questions can only be formulated after a sufficient understanding of relevant literature is achieved.

4.2 General shortcomings of database searches

A second aspect central to systematic reviews is “the rigour of the search process” (Kitchenham, 2004:7). As the role of research questions in systematic reviews is questionable, it is important to ask if systematic reviews can make up ground by using rigorous searches. There are several aspects of database searches that are relevant in this context: first, the role of search terms; second the way search strategies are used in systematic reviews; and finally the coverage of databases.

Review protocols used for systematic reviews require the specification of search terms as well as their extension by alternative spellings, synonyms and abbreviations. While this requirement seems to make sense the problem arises when it is assumed that such a process is rigorous resulting in a complete coverage of relevant literature. Formalization of the search process is mistaken for rigour. Furthermore, locking in a specific set of search terms at the beginning of a search process is problematic and not advisable. If academics are searching for literature they are supposedly not entirely familiar with the research domain they are interested in. However, at this point the selection of search terms is based on assuming what could be relevant terms rather than knowing what are actual central terms used in the literature. In addition, terms used to describe a topic are potentially non-finite and cannot be known a priori (Bodoff, 2009). Any list of search terms included in a review protocol will therefore be inherently incomplete, even wrong. Moreover, even the consultation of experts may not be sufficient as the following case shows:

“The lawyers and paralegals using the retrieval system were actively working on the lawsuit and were intimately familiar with the case, having worked on it for over a year prior to the test. They were also quite specific in their insistence that they needed to retrieve a *minimum* of 75% of the desired, or relevant, documents to conduct the defence of the lawsuit successfully. After the lawyers and paralegals used 51 individual queries to retrieve what they considered to be over 75% of the desired documents, the experimenters then tried to find whether they had missed any relevant documents. The results were striking: while the lawyers and paralegals were convinced that they were retrieving over 75% of the desired documents, they were, in actuality, retrieving only 20%! The principal reason for such low success rates in retrieval can be directly attributed to the indeterminacy of meaning in natural language. Stated succinctly, it is impossibly difficult for inquirers using a large full-text retrieval system to predict, by means of their search queries, the exact words, word combinations, and phrases that are used to express the content of the documents they desire but are *not used* to express the content of the documents they do not desire” (Blair, 2006:302).

This example not only shows that search terms cannot be sufficiently complete a priori but also highlight the importance of the search procedure. Usually systematic reviews use long search strings including the whole set of vocabulary identified in the review protocol. However, more complex search strategies do not guarantee superior searches. Long search strategies compiled in accordance to systematic review guidelines do not allow the assessment of what was contributed by individual search terms. Often masses of entirely unanticipated documents are retrieved. This problem is well known in information retrieval research as the recall-precision trade off (Buckland and Gey, 1994). Strategies developed to mitigate this problem like 'successive fractions' are not mentioned in systematic review guidelines possibly because they would compromise the 'rigour' of the search process believed to be leading to superior literature reviews. In the context of systematic reviews this means that a lot of time has to be spent sorting out results and judging their relevance. For example, Beecham et al. (2006) reported that only 16% of 1,445 documents retrieved were even considered for the next step. Eventually relevance could only be confirmed for 30 documents while being unable to establish relevance for a further 80 documents due to the lack of access to the full text of documents. This

means even if they were all relevant, less than 8% of all retrieved documents would turned out to be relevant.

Faced with an almost overwhelming number of retrieved documents which all have to be checked in order to adhere to the systematic review guidelines authors have to use pragmatic approaches. Researchers may either decide to include further restrictions on their searches or if numbers are not too daunting to make a first quick judgment based on titles and abstracts. While the title of documents often gives a good indication of the content it is important to stress that titles can be deceptive. Moreover, not even abstracts may be successful in sufficiently conveying the research content of a publication (Hartley and Betts, 2009). Subsequently, relevant literature may be lost as it was falsely excluded on the basis of titles or abstract. Take the case of Williams et al. (2009), even after restricting their search only to document titles the authors obtained more than 10.000 results from *Web of Science*. They subsequently restricted their search only to subject categories deemed 'relevant' for them. Still retrieving more than 4.000 documents the authors then made the decision to further restrict their search to 19 particular journals.

The selection of 'relevant' journals and papers indicates that the supposed 'rigour' and 'reproducibility' are compromised as researchers generally differ in what they consider relevant research for a particular purpose. Relevance is inherently individual; something that is new, relevant and important to one person may not be to someone else. This is demonstrated by Staples and Niazi (2007) who report low agreement between two independent raters in their systematic review process. Even after discussing selection criteria again after a first round of coding, a subsequent second round of coding still lead to low agreement between raters.

As systematic review approaches emphasize the importance of database searches for the review process it is worthwhile to have a look at the coverage of databases. Fundamentally only what is covered by a database can be found when using it. A study by Hood and Wilson (2001) is enlightening regarding this aspect. The authors used the *Dialog.com* search system to compare coverage of thirteen different topics across databases. *Dialog* was chosen as it is a major player in providing access to electronic databases. It started offering access to online databases long before the WWW was developed and covers several hundred major databases. Hood and Wilson's findings are sobering. The median contribution across all topics for the most productive database was only 27%. To achieve 93% coverage it was necessary to search 20 databases. In order to obtain all references available through *Dialog* it was necessary to search more than 50 databases for some topics. It is also important to note that the number of unique documents that could only be found in one database ranged from 38% to 80% for the different topics.

In addition to the spread of relevant literature across many databases the question remains open as to how much of the relevant literature is covered by databases in the first place? Some literature is notoriously badly covered in databases. For example, books or book chapters, but also journals with a predominantly national focus and material published in languages other than English. To investigate database coverage Wilson (1999) compared her extensive collection of publications on Bradford's Law (Bradford, 1934) to articles covered by databases. She reports that after checking for literature in 130 databases she could only find 26% of all publications in her collection through database searches. Even after narrowing her collection to include only central journal articles she could only boost the number of articles found to 61%. Thus, her research clearly shows that a substantial amount of the relevant literature on a topic may not be found when searching databases. This is also underlined by systematic reviews in medicine. Knipschild (1995) reported that only 36% of all relevant literature was found through database searches and Greenhalgh and Peacock (2005) report that only 25% of the literature was found through searches in fifteen different databases. In both studies the bulk of the literature was identified through other means.

4.3 Summary of criticism

In sum, key aspects of systematic review guidelines seem problematic. First of all, they seem to invert the literature review process to some extent by requiring the formulation of research questions before a literature review is undertaken that could identify gaps in current research. Moreover, even if such research questions are only seen as questions driving the literature review rather than the whole research, the type of research questions possible to be addressed by literature reviews are severely limited. This is especially problematic in the social sciences as the formulation of research questions

adhering to systematic reviews is only possible for a limited number of problems. In addition, they deny the fruitful dance between enhanced understanding of a problem and understanding of what questions are most relevant for addressing a problem.

Furthermore, an essential aspect of systematic reviews in medicine is that reviewers attempt to obtain *all* relevant literature. Insisting on the execution of a specific retrieval strategy is unlikely to help in this regard. The number of terms used to describe a problem is potentially infinite. The indeterminacy of meanings in natural languages makes it impossible to come up with a list of search terms that can capture all of the relevant literature. Systematic reviews assuming that the execution of a rigorous search strategy will lead to identification of relevant literature seem to confuse search terms with concepts (Fugmann, 2007). What researchers are interested in are certain ideas and concepts of relevance to their research problem, not particular wording used to describe them. Moreover, not only may important search terms be unknown, the meaning of words can also differ between specific areas or streams of research as well as among different authors. This can lead to the retrieval of many undesired documents and a failure to identify a potentially large number of relevant documents. The problem for systematic reviews is then: the more inclusive their search strategy, the more irrelevant documents will be retrieved; the more precise and specific the search terms the more relevant documents will be missed.

Lastly the coverage of databases is limited. Even the most productive databases for a particular topic only cover a fraction of all of the relevant literature in comparison to what can be found when numerous databases are searched. Besides, there is no way for researcher to assess if they used the most productive databases for their topic as coverage for different topics can vary from one database to another. One database being highly productive for one research problem may be relatively useless for the next. Finally, even searching multiple databases cannot retrieve all of the relevant literature as substantial amounts of literature cannot be found through database searches.

While conducting a systematic review is a time consuming exercise the results of such efforts are not likely to lead to “better reviews” as their proponents claim. Also the inclusion of selection criteria, such as concentrating on certain journals, contradicts the claim that systematic reviews are less biased. It is therefore worthwhile to see if there are alternative guides that can assist researchers when conducting a literature review.

5 Understanding literature reviews as hermeneutic process

All understanding is based on prior understanding. Humans make sense by relating their experiences to prior experiences. Therefore, even if an entirely new area is entered understanding of that area is based on earlier understanding of problems elsewhere. This aspect of human understanding is central to hermeneutics. While initially being developed in the context of the interpretation of biblical texts hermeneutics was later extended to the interpretation of all textual material and subsequently to understanding in general (Ramberg and Gjesdal, 2009). In the IS literature hermeneutics has been applied, for example, in the context of IS use and design (e.g. Boland et al. 2010).

One important aspect of hermeneutics is the hermeneutic circle. Simply put the hermeneutic circle symbolizes that human understanding of a whole is influenced by understanding of its parts, while at the same time the whole influences understanding of its parts. This process can be applied to understanding the literature and conducting literature reviews (Boell and Cecez-Kecmanovic, 2010). Understanding of a problem domain is gained by reading individual research publications. Through reading researchers gain better understanding of earlier research and therefore better understand of what aspects are important and have been studied within a domain. This in turn will enable researcher to have a better understanding of what further literature they need and how they can look for it.

This process is displayed as the hermeneutic circle of literature reviews in figure 3. As becomes apparent searching for literature is only one aspect of undertaking literature reviews. Only looking at the search process does neglect reading as at least equally important part of the literature review process. Reading helps to gain further understanding of the research domain. Researchers identify important concepts and the different terms used to describe them. After reading some publications researchers can be much more confident about what aspects they are looking for and what terms will be most appropriate when looking for them. Also while better understanding of earlier research in a field is gained researchers will become much more fluent in assessing other research, its quality and

how different research results are or could be related. This enables the identification of interesting contrasts or controversies in previous research as well as revelation of unexplored and unexplained phenomena, typically phrased as ‘gaps’ in current research. Such in depth investigations may then lead to the formulation of meaningful research questions.

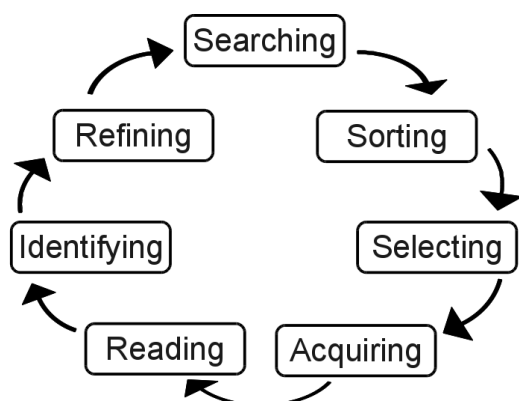


Figure 3. *The hermeneutic circle for undertaking literature reviews.*

While this process is theoretically open ended, it will enable researchers to reach a point where new publications are no longer leading to substantial shifts in the understanding of the research domain or problem. Moreover, going through subsequent iterations will ensure that important central publications are identified even if they were missed initially. Most importantly however, following a hermeneutic approach will allow the use of powerful searching and sorting techniques systematic reviews prohibit. It does therefore ensure that whatever time is available for a literature review can be used for critically engaging with relevant literature rather than sorting through long lists of search results.

6 Discussion and Conclusion

This paper started out by providing a brief overview of the origin of systematic reviews. It showed that systematic reviews originated in medicine as a reaction to sloppy reviews of earlier studies (especially metastudies) on medical treatments. One central point of systematic reviews is that the selection of literature on the basis of journals, the prestige of authors or particular databases severely compromises the coverage and quality of literature review and hence is not acceptable. However, this crucial point of systematic reviews seems to have been lost along the way as it is no longer reflected in guidelines for systematic reviews in the social sciences and the IS (which seems to be influenced by its neighbour discipline of software engineering). One wonders how systematic reviews conducted on this foundation should be less biased? More seriously, it seems that authors of systematic reviews conceal their biases rather than eliminate them. A worrisome development for academia.

It is worth reminding ourselves what it means to be less biased in literature reviews? Critically reading and approaching problems from different and unexplored angles are central aspects of all academic research. However, the types of questions that systematic reviews allow us to ask severely limit the intellectual horizon of research in the social sciences: "Systematic reviewers often set out to map fairly small fields with secure fences, and do not expect to look over the hedge" (MacLure, 2005:400). The question is therefore do systematic reviews while claiming to decrease bias in fact introduce a whole new dimension of bias to academic research? They only allow asking certain questions and discourage curiosity about alternative perspectives that might not be captured by their ‘review protocol’.

This paper questions that claims that systematic reviews based on a ‘rigorous’ and ‘reproducible’ process as defined by the review protocol will lead to ‘reliable’ results and ‘higher quality’ reviews. We demonstrated that while formal searches are reproducible, they by no means lead to comprehensive and reliable results. It is more likely that such systematic reviews will produce a very high number of documents majority of which are irrelevant while majority of relevant documents would remain unknown. More importantly, systematic reviews insist on a search rigour, which in most cases is irrelevant, while lacking critical engagement with literature in the research domain. The paper provides arguments that a literature review is a creative process and not a technically correct and ‘objective’ literature search. Literature reviews aim to identify relevant sources, results and ideas and

assesses them critically so as to identify problems in understanding, contradictions, mutual influences and fruitful or problematic research approaches and directions. As a creative process literature reviews needs to be critically reflective which is an intellectual endeavour going beyond the formal literature search process.

Based on the above argument, we claim that 'better' reviews can only be achieved through better and deeper understanding of the relevant literature. Following a rigorous search procedure as defined by protocols for systematic reviews is not likely to be helpful in this regard. It is actually more likely that such a procedure will cripple the literature review process by forcing a narrow focus while at the same time leading to a very large list of documents. Promoting systematic reviews as primary means for training PhD candidates in undertaking literature reviews is therefore highly inappropriate. Rather than giving neophyte researchers cookbook-like recipes for conducting literature reviews it is important to teach them search techniques. The hermeneutic circle of literature reviews was introduced as a framework that can be used for teaching relevant techniques and skills related to literature reviews. Moreover it can provide orientation for understanding of how these different processes hang together and how database searches are part of a larger literature review process.

To be clear this paper does not aim to outright reject systematic reviews as a method for undertaking literature reviews in IS but to show that they do not provide an approach that can be used for undertaking literature reviews in general. Our aim was to highlight some of the assumptions behind systematic reviews, and to expose shortcomings in current guidelines of systematic reviews. For instance, that search strings can sufficiently capture concepts represented in relevant literature. In many cases relying on database searches for identifying literature will not suffice. However, if systematic reviews are not understood as repeatable searches they may be applicable to address certain research questions. For instance they may be used in cases where meta-analysis of data is possible and findings from different studies can be aggregated.

Finally, literature reviews are an important part of all academic endeavours. Moreover, they also form an important type of literature in its own right. As Garfield (1987a:7) put it succinctly: "reviews provide scientists and scholars with informed notification of the published literature, help them maintain current awareness of related fields, and provide a backup to other methods of literature searching." For these reasons in all fields of academia literature reviews are of importance (Garfield, 1987b; Bensman, 2007). IS is no exception here. In fact, MISQ Review was established because researchers have found the lack of high quality literature reviews in IS (Watson, 2001). With Webster and Watson (2002) even arguing that this lack of reviews impedes progress in IS. This paper set out to advance the standing of literature reviews in IS by raising awareness on the shortcomings of systematic reviews.

Acknowledgements

We like to thank the reviewers of this paper and the track chairs for their helpful comments as well as the Australian Postgraduate Award (APA) for its financial support.

References

- Alm, G. (1916). Monography of Swedish fresh water ostracoda, along with the systematic review of Tribus Podocopa. Zoologiska Bidrag Från Uppsala. 4. 1-248.
- Beecham, S.; Baddoo, N.; Hall, T.; Robinson, H. and Sharp, H. (2006). Protocol for a Systematic Literature Review of Motivation in Software Engineering. Hatfield : University of Hertfordshire, Technical Report No 453. [online] <http://hdl.handle.net/2299/992>
- Bensman, Stephen J. (2007). Garfield and the Impact Factor. Annual Review of Information Science and Technology. 41. 93-155.
- Blair, David (2006). Wittgenstein, Language and Information. Back to the Rough Ground!. Dordrecht : Springer.
- Bodoff, D. (2009). Emergence of terminological conventions as a searcher-indexer coordination game. Journal of the American Society for Information Science and Technology. 60(12). 2509-2529.
- Boell, S.K. and Cecez-Kecmanovic D. (2010). Literature reviews and the hermeneutic circle. Australian Academic and Research Libraries. 41(2). 129-144.
- Boland, R.J.; Newman, M. and Pentland, B.T. (2010). Hermeneutical exegesis in information systems design and use. Information and Organization. 20(1). 1-20.

- Bradford, Samuel C. (1934). Sources of information on specific subjects. *Engineering*. 137(3550). 85-86. reprint in: *Journal of Information Science*. 1985. 10(4). pp. 173-180
- Buckland, M. and Gey, F. (1994). The relationship between Recall and Precision. *Journal of the American Society for Information Science*. 45(1). 12-19.
- Chalmers, I. and Altman, D.G. (1995). *Systematic reviews*. London : BMJ.
- Clarke, M.J. and Stewart, L.A. (1995). Obtaining data from randomized controlled trials: how much do we need for reliable and informative meta-analyses?. In: Chalmers, I. and Altman, D.G. (Eds.). *Systematic reviews*. London : BMJ. 37-47.
- Dybå, T. and Dingsøyr, T. (2008). Strength of Evidence in Systematic Reviews in Software Engineering. In *ESEM Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement* (pp. 178-187).
- Eysenck, H.J. (1995). Problems with meta-analysis. In: Chalmers, I. and Altman, D.G. (Eds.). *Systematic reviews*. London : BMJ. 64-74.
- Fugman, R. (2007). Informationstheorie: Der Jahrhundertbluff. Eine zeitkritische Betrachtung (Teil 1). *Information Wissenschaft und Praxis*. 58(8). 449-458.
- Garfield, E. (1987a). Reviewing review literature. Part 1, Definitions and uses of reviews. *Current Contents*. 18(May 4). 5-8. [online] <http://www.garfield.library.upenn.edu/essays/v10p113y1987.pdf>
- Garfield, E. (1987b). Reviewing review literature. Part 2, The Place of Reviews in the Scientific Literature. *Current Contents*. 19(May 11). 3-8. [online] <http://www.garfield.library.upenn.edu/essays/v10p117y1987.pdf>
- Greenhalgh, T. (1997). How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *British Medical Journal*. 315(7109). 243-246.
- Greenhalgh, T. and Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *British Medical Journal*. 331(7524). 1064-1065.
- Hartley, J. and Betts, L. (2009). Common weaknesses in traditional abstracts in the social sciences. *Journal of the American Society for Information Science and Technology*. 60(10). 2010-2018.
- Hood, W.W. and Wilson, C.S. (2001). The Scatter of Documents Over Databases in Different Subject Domains: How Many Databases Are Needed?. *Journal of the American Society for Information Science and Technology*. 52(14). 1242-1254.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, Eversleigh : Keele University and NICTA, Technical Report. [online] http://www.idi.ntnu.no/emner/empse/papers/kitchenham_2004.pdf
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. [online] <http://www.dur.ac.uk/ebse/resources/guidelines/Systematic-reviews-5-8.pdf>
- Knipschild, P. (1995). Some examples of systematic reviews. In: Chalmers, Iain; & Altman, Douglas G. (Eds.). *Systematic reviews*. London : BMJ. 9-16.
- MacLure, M. (2005). 'Clarity bordering on stupidity': Where's the quality in systematic review?. *Journal of Education Policy*. 20(4). 393-416.
- Mees, G. F. (1957). *A systematic review of the Indo-Australian zosteropidae*. Leiden : Brill.
- Mulrow, C.D. (1995). Rationale for systematic reviews. In: Chalmers, Iain; & Altman, Douglas G. (Eds.). *Systematic reviews*. London : BMJ. 1-8.
- Oxman, A.D. (1995). Checklists for review articles. In: Chalmers, Iain; & Altman, Douglas G. (Eds.). *Systematic reviews*. London : BMJ. 75-85.
- Oxman, A.D. and Guyatt, G.H. (1988). Guidelines for reading literature reviews. *CMAJ : Canadian Medical Association journal*. 138(8). 697-703.
- Petticrew, M. and Roberts, H. (2006). *Systematic reviews in the social sciences. A practical guide*. Malden : Blackwell.
- Ramberg, B. and Gjesdal, K. (2009). Hermeneutics. In: Edward N.Z. (Ed.). *The Stanford Encyclopedia of Philosophy*. Stanford : The Metaphysics Research Lab. [online] <http://plato.stanford.edu/archives/sum2009/entries/hermeneutics/>
- Shiffman, R.N.; Liaw, Y.; Brandt, C.A. and Corb, G.J. (1999). Computer-based guideline implementation systems: A systematic review of functionality and effectiveness. *Journal of the American Medical Informatics Association*. 6(2). 104-114.
- Staples, M. and Niazi, M. (2007). Experiences using systematic review guidelines. *Journal of Systems and Software*. 80(9). 1425-1437.
- Thompson, S.G. (1995). Why sources of heterogeneity in meta-analysis should be investigated. In: Chalmers, Iain; & Altman, Douglas G. (Eds.). *Systematic reviews*. London : BMJ. 48-63.
- Watson, R.T. (2001). Introducing MISQ Review - A new department in MIS Quarterly. *MIS Quarterly*. 25(1). 103-106.
- Webster, Jane; & Watson, Richard T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*. 26(2). xiii-xxiii.
- Williams, M.D.; Dwivedi, Y.K.; Lal, B. and Schwarz, A. (2009). Contemporary trends and issues in IT adoption and diffusion research. *Journal of Information Technology*. 24(1). 1-10.
- Wilson, C.S. (1999). Using Online Databases to Form Subject Collections for Informetric Analyses. *Scientometrics*. 46(3). 647-667.