An Evidence-Based Approach to Scoping Reviews

Antonio Hidalgo Landa, Istvan Szabo, Liam Le Brun, Ian Owen, Graham Fletcher and Martin Hill
Cranfield University, Shrivenham, UK

a.hidalgo@cranfield.ac.uk
i.szabo@cranfiled.ac.uk
i.lebrun@cranfield.ac.uk
i.owens@cranfield.ac.uk
g.p.fletcher@cranfield.ac.uk
m.hill@cranfield.ac.uk

Abstract: Scoping reviews are used to assess the extent of a body of literature on a particular topic, usually in order to ensure that further research in that area is a beneficial addition to world knowledge. It is typically the first stage of a research project, informing the following work, and so can have a significant impact on both methods and results.

Without systematic methods, reviews are subject to personal and community biases, whether intended or not. Evidence-based science is a broad collection of techniques intended to improve research quality by methodically analysing documented relevant facts, rather than relying on expert opinion. Systematic reviews are a key element and are used in many fields from medicine to public policy to engineering.

We propose a repeatable, documented, evidence-based approach to search and manage the literature, in order to assure demonstrated objectivity of the resulting review. Software tools are prototyped, issues identified and some resolved.
It is not clear from the three test cases whether the effort spent in researching, designing and developing the guidelines directly outweighs expert-based reviews. Further applications of the guidelines and comparative tests are needed.

Keywords: Evidence-based protocols, systematic review, literature review, information management.

1. Introduction

This paper is in three parts. This first introduces the work, explains the benefits of this approach and some of the relevant terminology. The main part of the paper follows, describing the proposed protocol. Finally in the concluding section we describe our experience in developing and using the protocol.
The protocol suggested here describes how sources and papers are selected. A random sub-selection are manually reviewed and later verified against other sub-selections by independent researchers. All steps are recorded for repeatability and review.
We intend the protocol to be reusable for other reviews and in other disciplines, with changes as appropriate.

1.1 Scoping study

Good scoping reviews are systematic literature reviews that provide specialists with relevant and quantified results about the knowledge available on a particular topic, within and without particular research disciplines and communities.

They aim "to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available, and can be undertaken as stand-alone projects in their own right, especially where an area is complex or has not been reviewed comprehensively before" [Mays et al., 2001]

"A systematic review is a summary of available research on a given topic that compares studies based on design and methods. It summarizes the findings of each, and points out flaws or potentially confounding variables that may have been overlooked. A critical analysis of each study is done in an effort to rate the value of its stated conclusions. The research findings are then summarized, and a conclusion is provided." [McGill, 2006]

1.2 Evidence and Expert-based Reviews

The literature review is the foundation of a research project, and requires the analysis of vast amounts of information: 'the literature'. To control this analysis, we suggest an approach based on evidence rather than expertise to assure its objective quality; the evidence-based paradigm requires that review quality is measurable, visible, objective and verifiable [Higgins and Green, 2009].

Experts can introduce various unintentional personal and community biases to the review. These biases might be the researcher's own interests, a lack of training or broad experience, limited views of the total available research due to discipline or language, and many others. The review may then not properly scope the literature actually available, at best excluding some work that is then repeated by the researcher, at worst distorting the picture so that the following research is misguided.

Evidence-based systematic reviews have been widely adopted and developed in medicine (the Cochrane Institute is an early example) and are intermittently used in education, policing, teaching, management and engineering. It appears to have made little impact on information systems research. In this paper we hope to introduce some of the benefits of an evidence-based approach to conducting what is often the starting point of an enquiry: the literature scoping review.

1.3 Reducing Bias

In most research fields it is impossible to review the total literature available due to the sheer quantity of it. A selection has to be made, and skills – expertise – developed in making that selection representative and useful to the work. This expertise in literature review is not necessarily the same as expertise in the research topic of interest, and is subject to various biases.

The Oxford Dictionary defines bias as: "Inclination or prejudice in favour of a particular person, thing, or viewpoint" [O.E.D., 2010] Bias influences the quality of literature reviews and scoping studies by distorting the selection of material. Any bias introduced at the start of a research project is likely to skew the subsequent enquiry and so the results.

Therefore we should take measures to minimise the external and internal influences that can lead to bias.

"Cognitive bias" is the conscious or unconscious actions of the researchers that systematically distort work. A reviewer may examine literature about a limited set of topics or methods or technologies, or from only a limited set of authors or research communities, leaving aside other relevant literature through ignorance or reputation. Preferred topics or authors might be specified by funders. The researcher may unconsciously guide the results of a survey toward findings that are either expected or preferred. The researcher – especially new researchers such as PhD students – may be unfamiliar with the relevant literature, and so may miss large regions of it or misread the impact of some aspects. Even more experienced researchers may be familiar with certain fields but in large subject domains are rarely up to date with all related work.

Such biases can be reduced by separating the design and implementation of a study, either in time or by the personnel involved [Dorak]. We propose that the methodology of the review is developed and documented before searches are run. By agreeing and fixing the study protocols ahead of the search, they cannot be changed by the results, so the selection of material is not distorted by the early review. This also makes the review repeatable by other research teams so reducing any biases of a single researcher, and the method and actions are visible to any monitoring staff so encouraging rigour.

The analysis of a large number of papers in a structured manner should reduce the effect of bias from any one of the primary studies. [Vest and Gamm, 2009]

1.4 Repeatability

In many research fields results are not considered valid until they have been repeated and independently verified. In this case, repeatability allows quality assurance activities to be completed independently of the primary research team; checking the results by repeating the work or samples of the work elsewhere improves confidence in the results.

The protocol ensures that the results of the scoping study and the literature review can independent verified.

2. Method Guidelines

This section describes the proposed guidelines to carry out an objective, repeatable evidence-based scoping study. It is based on the "evidence based review protocols" and "systematic reviews" used in healthcare. [Higgins and Green, 2009].
The main steps are as follows:

1. Define & refine research search terms ("What are we going to look for?")

2. Identify databases and search engines ("Where are we going to look?") and query using the search terms.

3. Create and apply the inclusion & exclusion criteria filters

4. Verify the sub-selection is representative

These steps acquire and manage lists of candidate documents, and these lists need to be managed carefully. We propose bibliographic tools and databases to index the documents and bibliographic managers to run queries.

We found that we needed to be careful when specifying the scope of our research, to ensure that the number of papers retrieved can be feasibly analysed. A strong scoping statement is required to prevent irrelevant material from being included within the study and to ensure that results meet the needs of the work programme such as time and staffing. Requirements for papers to be included in the study are implemented by the exclusion and inclusion criteria.

Each step should be recorded and the results logged and tagged, in order to ensure repeatability.

2.1 Research terms definition

Defining the right research questions is important; the nature of the question influences the success of the review. Complicated questions make the review task difficult, and unanswerable questions make it more so.

Search terms are in this case lists of words, but we are interested in the concepts behind these words. Certain 'jargon' terms are commonly used in particular research communities, but work on similar concepts outside those communities may use other words for similar things; "sensor meshes", for example, share many similar concepts to "web services". In order to capture work by these other, possibly unknown and unaware communities, we broaden our queries to include other terms that may refer to the same concepts.

Formally, the relationships between terms would be represented using a semantic diagram called an ontology. These describe concepts, terms and the relationships between the concepts and terms.

Expanding research terms is not trivial. We suggest two mechanisms to perform this task: "Thesauri-based" and "reports-based".

2.1.1 Thesauri-based expansion

This uses synonyms as defined in a thesaurus to expand the search terms. This is best done manually so that the meaning of the terms remains consistent, but some automation appears feasible as long as contexts and tenses can be specified.

To keep track of the relationships, create a table where the terms are placed down a column and each thesaurus along a row. For each term for each thesaurus, enter the appropriate synonyms in the cross-indexed cell, and add the terms to the search term column (see Table 1). Carry this out for each of these new terms, and so on for an acceptable number of iterations or until new ones peter out.

Table 1: Thesauri-based expansion

	Search terms	Encarta	Roget's
Initial/Root terms	Mesh	lattice, netting, web, weave	net, network, skein, web
	Network	System, web, linkage	web
	Grid	framework	n/f
Iteration 1	Web	(nothing new)	(nothing new)
	System	Structure, coordination	organisation
	Framework	structure, frame, scaffold	scaffold, platform
Iteration 2	Structure	...	...
	Scaffold	...	...
	…	…	…

2.1.2 Report-based expansion

This uses the terms in the reports found by the original search terms to look for further reports. The outline process is as follows:

1. Select (and record) the search engines which will be used to expand the terms.

2. Query those engines with the initial search terms.

3. Select, say, the first ten documents found.

4. Extract all the analogies, definitions and different terms referred to by each research term concept in the selected documents.

5. Add those terms to the list of search terms.

6. Repeat for an acceptable number of iterations or until new terms peter out.

To keep track of the semantic relationships, use a table as above with the search engines and documents along the top.

2.2 Databases & Search Engine Analysis

Most academic literature is available online, and full-text databases can be queried using the defined search strings to obtain candidate studies for review. In our studies we used academic search engines such as SCOPUS [SCO, 2010], IEEE Xplorer [IEE, 2010], Google Scholar [Goo, 2010] and Citeseerx [Cit, 2010]. Other web service search engines such as Google and Bing can be used for other online resources.

Note that this excludes the considerably large areas of recorded knowledge such as books, those journals not subscribed to by the searching organisation, and publications in other languages. This exclusion of certain types of literature automatically introduces forms of community bias: commercial literature on management practice, for example, is typically published as books.

2.2.1 Automating Queries

Manually querying different search engines is laborious and prone to human error, so we investigated ways of automating this process.

Our search client software reads through a text file for search terms, constructs a suitable URL from the web page address and the search terms, submits this as an http request over TCP/IP, and stores the returned page as a file for later analysis.

Frequent and immediate re-querying are identified by some search engines as a "denial of service" attack and results are not returned, so delays between queries are required.

Most search engines have similar interfaces that provide a Boolean-logic query pattern to be submitted from the search terms, and which could be developed further from the ontology developed above.

Search engines and academic paper databases tend to return results in different HTML formats that are not structured for extracting common fields. Some academic databases and some search engines

(such as Google and Bing) provide XML interfaces for automatic querying, but this needs further investigation.

Some search client applications and libraries (such as JavRef and KBibTex) can query most of the commonly used databases and search engines. These applications retrieve the references in bibtext format and help to manage all the references including filtering and merging from different databases. They can also export the reference databases in several formats.

2.3 Post-Query Filters: Exclusion & Inclusion Criteria

The lists of papers returned from the queries are passed through a set of filters in order to reduce prior study bias, provide a number that can be feasibly reviewed, remove some irrelevant documents, and eliminate others of insufficient quality.

We use three different kinds of filters: global and background filters, semantic filters and evidence grade filters. Not all can be automatically applied. If there are too many papers to manually review, we then run a sampling filter to select a suitably randomised subset.

Our intent is to set out strict criteria so that all papers are treated equally. Personal preferences given to famous authors, companies, institutions or journals are reduced, although the effects are not removed altogether (nor do we necessarily want to). When the inclusion criteria specifically add weighting to one of these factors, this is an explicit documented decision.

2.3.1 Global Filters

The global and background filters use data about the document (metadata) such as author, year of publication and institutions to limit the results. For example:

- Include only one document by the same author in the same year.
- Exclude documents previous to 1975.

2.3.2 Semantic Filters

Semantic filters exclude those results that have been returned due to sharing terms with different meanings. For example, the search term "SOA" will return documents about "Service Oriented Architecture", but also documents about "Semiconductor Optical Amplifiers".

2.3.3 Evidence Grade Filters

Evidence grade filters are used to exclude studies of insufficient quality, and are designed to remove badly conducted science, conjecture and poorly designed studies from the results. Common filters might be:

- Is the supporting evidence shown?
- Is the supporting evidence complete?
- Are there (unacceptable) conflicts of or vested interests?
- Is it repeatable? Repeated?

2.3.4 Sampling the Result Set

When the filtered list is still too large enough to manually review, we need to select a suitable subset.
A simple randomised selection across the result set may be sufficient, if we assume that the number of papers retrieved on each subject indicates the relative relevance of that subject.
However this means that a popular topic in one discipline can swamp work in the area in other disciplines. For example, consider a case where we discover papers on our topic in five research areas: in one area we have found 10,000 papers on the topic, in another we have found 600, another 100, another 30 and in one area just 5 as given in Table 2.

Table 2: Proportion of papers found on a topic by area

	Amount of papers	Proportion
Area 1	10,000	93.6%
Area 2	600	5.6%
Area 3	100	<1%
Area 4	30	<1%
Area 5	5	<<1%

If we were to randomly sample 100 papers, we would expect to return mostly those from Area 1, a few from Area 2, and maybe one from one of the other three areas. Thus we would lose the perspectives on the topic from the other areas.

We can use 'softening metrics' to weight the proportions to the minor areas. This uses a function (such as the square root) of the number of papers to define the proportion, rather than the number of papers directly.

For example, Bradford [Bradford, 1985] suggests that scientific literature is distributed or scattered logarithmically as the relevance of the search terms decreases linearly. We can use the function `ln(num_papers+1)` to adjust the sampling weighting as given in Table 3:

Table 3: Weighted Bradford adjustment (to nearest %)

	Amount of papers	Proportion	ln(np+1)	Proportion ln(np+1)
Area 1	10,000	94%	9.2	37%
Area 2	600	6%	6.4	26%
Area 3	100	<1%	3.9	16%
Area 4	30	<1%	3.4	14%
Area 5	5	<<1%	1.8	7%

2.4 Verification mechanisms

The result of the above filters is a subset of papers from the total papers available. In order to verify that we have a representative set, we re-run the randomised sub selection filters and compare the abstracts of the new set with the original set.

If we find no significantly new material, or subjects in apparently significantly different categories, we can be confident that the original set is representative and so suitable to use.

3. Conclusions

3.1 Example Case Study

This protocol has been applied to reviews by our research group, and indeed it has been used in three Ph.D. research scoping reviews. In this section we examine one of these cases.
This protocol helped us to meet time limits and cope with the volume of information available. The following issues were identified and only partially resolved:

- Search terms can be ambiguous and difficult to expand sensibly without leading to very large and irrelevant result lists. This was partly dealt with by the expansion methods given above (using thesauri and dictionaries, and using previous reports).

- The search engines and literature databases do not share standard query languages or result formats. Some bibliographic managers (eg kbibtex and jabref) can help as they have interfaces to retrieve bibliographic references from most of the main databases.

- Managing the large numbers of papers required a systematic way of storing them and referring to them. Our final review included around 30 documents, extracted from 300 which were in turn extracted from several thousands. To manage lists of these documents we used the bibtext format, which is a common bibliographic format.

Our filters helped us to meet certain information quality requirements, and let us reduce the volume of information to meet deadlines. Also the protocol assured us that the scope of the review was complete and fairly distributed across subject areas.

3.2 Comparisons

No formal statistical comparisons have yet been done, but some impressions have been given by two of the contributors:

1) Manually reading abstracts to sub-select from large document sets to feasibly reviewable smaller sets appears to give similar results to the automatic filters above. However the two activities are not exclusive.

2) The time taken to design, develop and implement these methods has not been trivial, and it is not clear whether it would have been better to 'just get on with' an ordinary expert-based review.

3) Anecdotally, a citation-spider approach used by one contributor after he used these protocols yielded a wider variety of solutions than the automatic search terms had. Again, the two approaches are not exclusive.

3.3 Concluding Remarks

Reviewing is an art, but should also include proper, documented, scientific process. Researchers have their own favoured ways to do reviews but behind them should be a science: the science of reviewing. Good scientific practice should inform all aspects of scientific research.
We are pleased to provide a view to the information systems research discipline of the scientific review practices developed in other fields, and propose an approach to manage information during the early scoping stage of a research project.

While we have defined a protocol to deal with this kind of review, we recognise it may not be the ideal protocol for all subjects, and that researchers should omit or adapt it as required. Similarly, it is not a complete review method; it should be used in conjunction and comparison with, for example, citation spider searches.

We propose it as a reference for others to develop their own evidence based literature reviews, and encourage them to design and document a process to follow before the start of a literature review, and to share that process with others so their results can be replicated and so undergo independent verification.

References

Citeseerx. http://www.citeseerx.ist.psu.edu/, 2010.
Google scholar. http://scholar.google.co.uk/, 2010.
Ieee xplorer. http://ieeexplore.ieee.org/search/advsearch.jsp, 2010.
J.P.T. Higgins and S. Green. Cochrane handbook for systematic reviews of interventions. Wiley, 2009. O.E.D. Oxford English Dictionary. Oxford University Press, 2010.
J.R. Vest and L.D. Gamm. A critical review of the research literature on Six Sigma, Lean and StuderGroup's Hardwiring Excellence in the United States: the need to demonstrate and communicate the effectiveness of transformation strategies in healthcare. Implementation Science, 4 (1):35, 2009.
M.T. Dorak. Bias & confounding. Eri §.
McGill. Mcgill university 2006) glossary of terms. http://www.repar.veille.qc.ca/burnengine/?p=glossary, 2006.
N. Mays, E. Roberts, and J. Popay. Synthesising research evidence. Studying the organisation and delivery of health services: Research methods, pages 188–220, 2001.
S. C. Bradford. "sources of information on specific subjects". J. Inf. Sci., 10(4):173–180, 1985. ISSN 0165-5515. doi: http://dx.doi.org/10.1177/016555158501000406.
Scopus. http://www.scopus.com, 2010.