**How and Why to Document Searches in the Sciences: A Hands-On Workshop**

Instructor: Kristine Alpi, Director, William Rand Kenan, Jr. Library of Veterinary Medicine, NCSU Libraries, Kristine_alpi@ncsu.edu, 919-513-6219. Questions: http://www.lib.ncsu.edu/askus

**Goals:**
- improved clarity of methods sections in proposals and in manuscripts
- enhanced communication with advisors, collaborators or reviewers about your exploration of relevant literature
- reproducibility by others where the search or analysis strategy is part or all of your experiment

**Strategies to be covered:**
- search logs (Sample on last page)
- saved searches in databases (including date last updated and parameters that change over time)
- capturing searches and results in the moment
- storing searches and results with citations to keep organized

**Cases of Documentation** from Required to Required/Recommended to Suggested to Implied
- Animal Studies – Institutional Animal Care & Use Committee requires documented search
- Systematic review/Meta-analysis - finding all relevant studies is part of method, PRISMA guidelines recommendations (software engineering and JAMA examples)
- *In Silico* bioinformatics - BLAST search and how described in Cell and Science articles
- Human Subjects Research – Institutional Review Board - Johns Hopkins case of not searching the older literature

**REQUIRED EXAMPLE:** Animal Use Studies

```markdown
NORTH CAROLINA STATE UNIVERSITY APPLICATION FOR VERTEBRATE ANIMAL USE (Revised by IACUC 07-15-2010)

SECTION G. Consideration of Alternatives

COMPLETION OF THIS SECTION IS REQUIRED FOR ALL APPLICATIONS INVOLVING THE USE OF ANIMALS AT THE LEVEL OF CATEGORIES C, D OR E. For descriptions of animal use categories, see [http://www.ncsu.edu/iacuc/forms.html](http://www.ncsu.edu/iacuc/forms.html), “Guidance on Pain and Distress Levels.” [NB: This is everything except “Animal use activities that involve only breeding, conditioning, or holding”]

There must be a written narrative description of the methods and sources which were consulted to determine the availability of alternatives (reduction, refinements, replacement).

```
IACUC form continued…

1. Please respond to items a-d regarding your literature search for alternatives:
   a. Which databases were searched?

   b. Indicate the range of dates searched within the database(s) (i.e. 1900-2009, or 1987-present):

   c. Keywords should include those likely to yield information on alternatives to the potentially painful or distressful procedures or conditions that are part of this protocol. **As such, one of the keywords included in your search terms should be either “alternative” or “alternatives.”**

   List keywords used:

   d. Provide the most recent date that a full search was performed:

**Elements of Documentation:**

1. **Date of Search(es)** (legal responsibility – was knowledge “published” and “available” at time of search; scientific responsibility – be aware of new info that comes out later that may cause you to want to adjust your protocol OR address new information in your discussion.)

2. **Databases Searched**
   a. Web of Science versus Web of Knowledge constituent databases for “All Databases” search – look at “Select a Database”
   b. Compendex or INSPEC versus Engineering Village -- constituent databases vary with subscription and selections
   c. Google Scholar – constantly changing content. Capture the pages of results that you actually perused as documentation.

3. **Time Periods Covered**
   a. MEDLINE/PubMed going back in time – how do you know where it was at the time you searched it. See OLDMEDLINE load data chart for example: [http://www.nlm.nih.gov/databases/databases_oldmedline.html](http://www.nlm.nih.gov/databases/databases_oldmedline.html)
   b. Web of Science and other database time frames vary by institution’s subscription

**DISCUSSION and HANDS-ON EXERCISE:** What databases do you normally search? Confirm now what years they cover. Share some of these examples as group.
4. Keywords used and how they were combined
Documenting nesting, truncation, Boolean operator use: (game* or gaming or Halo or Wii)

5. Citations (or unique ID # of citations) followed for “pearl-growing,” “snowball searching,” “cited by,” and “citing references”

*Hint: Most of these databases like Web of Science and Google Scholar have strategies at the search or citation level that you can use to keep up for the duration of your work on a project.*

6. Selection of Abstracts and Following through to Full Text
You are responsible for everything that came up in the search results even if you didn’t choose to read most of the abstracts, etc., unless you make it clear about your cut off points. For Google Scholar/web results – at what point did you cut off looking at results? Capture what you reviewed using Print to Adobe PDF or other capture strategies.

For systematic reviews and meta-analyses, the decision about what to include and exclude is part of your methodology as well.

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement provides an evidence-based minimum set of items that for reporting systematic reviews and meta-analyses, and is an update and expansion of the QUOROM Statement. Although it focuses on randomized trials, the PRISMA Statement can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions.

From the Checklist: (there is also a Flow Diagram)

**Information sources** 7: Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.

**Search** 8: Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.

**Study selection** 9 State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).

Example of Documenting Use of Above Guidelines and Clarifying Author Roles

Jennifer Keiser; Jürg Utzinger. **Efficacy of Current Drugs Against Soil-Transmitted Helminth Infections: Systematic Review and Meta-analysis.**
JAMA, April 23/30, 2008; 299: 1937-48. [http://jama.ama-assn.org/cgi/content/abstract/299/16/1937](http://jama.ama-assn.org/cgi/content/abstract/299/16/1937)

**Abstract: Data Sources** A systematic search of PubMed, ISI Web of Science, ScienceDirect, the World Health Organization library database, and the Cochrane Central Register of Controlled Trials (1960 to August 2007). **Study Selection** From 168 studies, 20 randomized controlled trials were included.
Actual Paper: METHODS We adhered to the Quality of Reporting of Meta-analyses (QUOROM) guidelines. We searched PubMed (http://www.ncbi.nlm.nih.gov) (1966 to August 2007), ISI Web of Science (http://www.isiknowledge.com) (1960 to August 2007), ScienceDirect (http://www.sciencedirect.com) (1960 to August 2007), the Cochrane Central Register of Controlled Trials (http://www.mrw.interscience.wiley.com/cochrane/cochrane_elcentral_articles_fs.html) (1960 to August 2007), and the World Health Organization library database (1960 to August 2007) to identify clinical trials, studies, and case reports pertaining to the use of albendazole, mebendazole, levamisole, and pyrantel pamoate for treating infections with *A lumbricoides*, hookworm, and *T trichiura*. No restrictions were set on year or language of publication. We used the terms *albendazole*, *mebendazole*, *levamisole*, and *pyrantel pamoate* in combination with *trial* or *study* or *case report* and *ascariasis*, *Ascaris lumbricoides*, *hookworm*, *Ancylostoma duodenale*, *Necator americanus*, *trichuriasis*, *Trichuris trichiura*, and *soil-transmitted helminths*. Bibliographies of identified articles were screened for additional relevant studies.

Exercise: Format the above search terms in a reproducible way using nesting and Boolean.

Author Contributions: Dr Keiser had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Keiser, Utzinger. Acquisition of data: Keiser. Analysis and interpretation of data: Keiser, Utzinger. Drafting of the manuscript: Keiser, Utzinger. Critical revision of the manuscript for important intellectual content: Keiser, Utzinger. Statistical analysis: Keiser. Obtained funding: Keiser, Utzinger. Role of the Sponsor: The Swiss National Science Foundation had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript. Additional Contributions: We thank the library team of the Swiss Tropical Institute for its help in obtaining relevant articles.

Example from a non-life sciences discipline:

Software engineering research for computer games: A systematic review

Abstract: Objective: This study aims to assess the state of the art on research concerning software engineering for computer games and discuss possible important areas for future research. Method: We employed a standard methodology for systematic literature reviews using four well known digital libraries.

Full Text: 2.2. Search process

A systematic review on a “new topic” should identify and highlight specific sources about the subject under study. However, no such sources exist in the domain of software engineering for computer games. Thus, related studies may be published in journals and conferences that are related either to “traditional software engineering” or to “entertainment computing”. In the area of entertainment computing some
dominant publishers such as Association for Computing Machinery (ACM), IEEE Computer Society (IEEE), Springer and Elsevier provide several journals and conferences, such as “Computers in Entertainment”, “Entertainment Computing”, “Conference on Advances in Computer Entertainment”, “Conference on Foundations of Digital Games”, “SandBox Workshop”, “Conference on Future Play” and “Conference on Entertainment Computing”.

The search procedure aimed at the identification of candidate primary studies that would be either included or excluded from the final set of the review studies. The search plan involved automated search into four well known digital libraries: ACM Digital Library, IEEE Digital Library, ScienceDirect, and SpringerLink. The search parameters included one keyword, i.e. game, which should be included in the article title and another keyword, i.e. software, which should be mentioned at least once in the article full text. In addition to that, before the manual observation of the studies, automated filtering was employed. In ScienceDirect the search was limited to subject “Computer Science” and in SpringerLink, the search was limited to subject “Software Engineering”.

The article set that has been returned from the aforementioned query consisted of 3463 articles. However, the majority of these articles were identified marginally related to software engineering. The exclusion of irrelevant articles was manually conducted according to the inclusion and exclusion criteria defined in Section 2.3.

2.3. Inclusion and exclusion criteria

The papers that are selected as primary studies in the review had to be relevant to software engineering, i.e. the article should be classified under a software engineering topic, as described in Section 2.5.1. In line with [5], there are four stages in filtering the article set in order to produce the primary study data set. These stages are:

(a) identify relevant studies – search digital libraries (on the completion of this stage the article set consisted of 3472 articles),

(b) exclude studies on the basis of titles (on the completion of this stage the article set consisted of 223 articles),

(c) exclude studies on the basis of abstracts (on the completion of this stage the article set was consisted of 130 articles),

(d) obtain studies and select those relevant to software engineering on the basis of full text (on the completion of this stage the final primary studies dataset consisted of 84 articles).

All papers that have been considered in the primary study selection phase, after the completion of filtering out papers according to their title, are presented in the paper’s website (http://sweng.csd.auth.gr/~apamp/survey_games.html). In line with [1] the title and abstract were examined by the first author whereas the full papers which were not rejected at the first three stages were examined by both authors. During the survey, we came up with a variety of interesting papers, but we preferred to include in the review only research that was closely related to software engineering issues. The main issues that are related to game development but are not related to software engineering are “game based learning”, “artificial intelligence”, “social impact of gaming”, “networking” and “graphics algorithms”. Finally, in the review only journal, full conference and workshop papers have been considered, conference short papers, smaller than four pages, and posters have been removed from the
survey. The final primary study dataset consisted of eighty-four (84) research articles that are presented in an Appendix by the end of the paper.

**DISCUSSION:** Writing it up -- does it have to be so complicated? Compare abstract info with full article. Compare the non-PRISMA example above with the search description below. Identify the key documentation elements. Indicate whether you would be able to reproduce the search and/or results.

**Complex Example: Study Identification**
A senior reference librarian with expertise in systematic reviews (P.J.E.) designed a strategy to search MEDLINE, Scopus, CINAHL, EMBASE, ERIC, TimeLit, Web of Science, Dissertation Abstracts, and the University of Toronto Research and Development Resource Base for relevant articles. Search terms included delivery concepts (such as Internet, Web, computer-assisted instruction, e-learning, online, virtual, and distance), study design concepts (such as comparative study, evaluative study, pretest, or program evaluation), and participant characteristics (such as education, professional; students, health occupations; internship and residency; and specialties, medical). eTable 1 describes the complete search strategy. We restricted our search to articles published in or after 1990 because the World Wide Web was first described in 1991. The last date of search was January 17, 2008. Additional articles were identified by hand-searching reference lists of all included articles, previous reviews, and authors' files.


**Storage Options and Issues**
- Save search histories in the database they were created (can create alerts from these also).
  - Date management – does your database track the “last searched” or “last updated” date?
- Print web search history and results pages to PDF and store PDFs as documents in your RefWorks account or other storage areas. Date printed on PDF output.
- Copy and paste search history into search log in Word and maintain digitally. Provide file information in your lab notebook or project tracking system OR print and include dated and numbered pages in your notebook.
- Combination of the above. 😊

**SAVING/STORING DISCUSSION:** How do you keep track now of what you have searched? What does your advisor say/ask about literature review progress?

**DEMONSTRATION:** PubMed’s MyNCBI, Saved Searches, Show What’s New, EDAT and Collections and Outside Tool for linking to NCSU Full-Text.

**HANDS-ON EXERCISE:** Your choice of BLAST or Google Scholar search. Save the search and save the results by printing to PDF. Did you lose anything you needed? What did you gain automatically that is part of the documentation?
In Silico Bioinformatics

Cell Instructions for Authors - A more detailed version of the procedures and details such as oligo sequences, strains, and specifics of how constructs were made can be included in Supplemental Information, but it is not appropriate to move the majority of the Experimental Procedures to Supplemental Information in order to shorten the text. Please see our complete Supplemental Information guidelines for more information. http://www.cell.com/supplemental_information_guide


DEMONSTRATION: Compare regular article section on Determining Lateral Gene Transfer with what is provided in the Extended Experimental Procedures. Identify documentation elements.


In this study, we investigated the distribution of the E. coli–type oxidative stress response by performing a BLAST search for SoxR and SoxS in the bacterial domain (15). SoxR was found in sequences from 176 strains in the phyla Proteobacteria and Actinobacteria (Fig. 1A), 123 of which come from completed genomes. The occurrence of SoxS was restricted to the family Enterobacteriaceae. To identify alternative SoxR targets in non-enterics, we searched all available complete bacterial genomes (616) for the presence of soxRboxes (i.e., SoxR-binding sites in the promoter regions of target genes) using a position weight matrix (PWM) derived from the soxRbox sequences of 12 diverse SoxR-containing bacteria (fig. S1B). This PWM permits statistically robust predictions of SoxR binding to a soxRbox. Of the 123 soxR containing genomes, 121 contain soxRboxes. SoxRboxes were also found in 27 genomes (19 were Firmicutes) that do not contain a soxR homolog. The results of our analysis (table S1 and http://soxRbox.mit.edu) were consistent with gene expression studies made in the Gram-negative bacteria E. coli, S. enterica (10), P. aeruginosa (8, 13, 14), and Agrobacterium tumefaciens (16), which validates our search algorithm.

View Supporting Online Material (Table S1): www.sciencemag.org/cgi/content/full/321/5893/1203/DC1

Human Subjects Research – It’s the Investigator’s Responsibility

Ellen Roche, a healthy, 24-year-old volunteer in an asthma study at Johns Hopkins University, died in June 2001 because a chemical she inhaled led to the progressive failure of her lungs and kidneys. The Baltimore Sun concluded that while the supervising physician, Dr. Alkis Togias, made "a good-faith effort" to research the drug's adverse effects, his search apparently focused on a limited number of resources, including PubMed, which at that time was searchable only back to 1966. Previous articles published in the 1950s, however, with citations in subsequent publications, warned of lung damage associated with hexamethonium. (Source: http://newsbreaks.infotoday.com/nbreader.asp?ArticleID=17534)
The principal investigator subsequently stated to the investigation committee that he had performed a standard PubMed search for potential hexamethonium toxicity and consulted standard, current edition textbooks of pharmacology and pulmonary medicine before submitting the application to the IRB. None of these sources mentioned hexamethonium-related pulmonary toxicity. Although the PI made a good-faith effort to uncover previous reports of hexamethonium-related toxicity, the vagaries of performing such a search are illustrated vividly by the following performed as part of this investigation: (1) None of the last four editions of Goodman and Gilman, or the current edition of Fishman’s Pulmonary Medicine mentioned this toxicity, but an older edition of Fishman’s text did; (2) During PubMed searches, "hexamethonium inhalation lung injury" gave 0 hits, "hexamethonium inhalation" gave 42 hits (but none referring to pulmonary toxicity), "hexamethonium lung" yielded 3 useful articles, "hexamethonium lung toxicity" gave 4 hits, but 0 useful articles, "hexamethonium lung hypersensitivity" gave 16 hits with 3 useful articles, and "hexamethonium lung fibrosis" gave 3 hits and 2 useful articles; (3) the Google search engine and Yahoo uncovered the pneumotox web site (see below), but LookSmart and GoTo.com did not; (4) the Micromedex data base had lung toxicity as the first adverse effect of hexamethonium. The pulmonary toxicity associated with oral, intramuscular, and/or subcutaneous hexamethonium administration for hypertension was first reported in 1953 (1) and between then and 1960 several articles including individual case reports and small series of autopsied cases were published (1-12). In 1970 a review article (13) listed six references from the 1950’s.

Institutional Review Board for the Use of Human Subjects in Research
http://www.ncsu.edu/sparcs/irb/forms.php  SUBMISSION FOR NEW STUDIES

POTENTIAL RISKS
1. State the potential risks (physical, psychological, financial, social, legal or other) connected with the proposed procedures and explain the steps taken to minimize these risks.

CONTINUING REVIEW
Is there any new information since the last IRB review that might impact the risks vs. benefits of the research? If so, please submit a summary of any recent literature, findings, or other relevant information, especially information about risks associated with the research.

WHAT CONSTITUTES SUBSTANTIVE AND MEANINGFUL CONTINUING REVIEW?
a summary of any recent literature that may be relevant to the research and any amendments or modifications to the research since the last IRB review;

The regulations do not state or even suggest that the IRB is required to perform or validate a review of the literature. Reviewing the literature is a scientific activity and as such is the responsibility of the investigator – the IRB receives the results of the review.
http://www.hhs.gov/ohrp/humansubjects/guidance/contrev0107.htm