Chapter 8

Extracting Study Data

8.1 Overview of data extraction 93
8.2 Examples of extracted data and extraction procedures 95
Examples from quantitative systematic reviews 96
Examples from qualitative systematic reviews 97
Example from a mapping study 98

The objective at this stage of the review process, which is highlighted in Figure 8.1, is to extract, from the reports of the primary studies, the data needed to address the research questions. The strategy for data extraction, including the data extraction form (or forms) needs to be defined and justified. The use of data extraction forms can help to maintain consistency (across studies and between extractors), and, where these are held electronically, the extraction and data recording can be performed in one step. Although the forms will have been piloted during the planning phase, it is possible that they will have to be revised during the data extraction stage when a broader range of studies are processed. Some tools that provide support for data extraction and subsequent data management are indicated in Chapter 13.

The structure and balance for this chapter is somewhat different from the three previous chapters with a greater emphasis being placed on providing a set of examples to illustrate the strategies that have been followed across a range of quantitative systematic reviews, quantitative systematic reviews and mapping studies. The data to be extracted for a review of any kind is very closely related to the specific research questions for that review and also to the requirements of the synthesis/aggregation phase. In the examples, we show this connection as well as illustrating a range of procedures for extracting, recording and validating the data.

As noted in Chapter 7, although data extraction is quite distinct from quality assessment, these two stages can be performed sequentially or together.

8.1 Overview of data extraction

Different types of data are usually extracted for the different types of review although all usually include some ‘standard data’ that records, for example, publication details for each paper included in a review and information

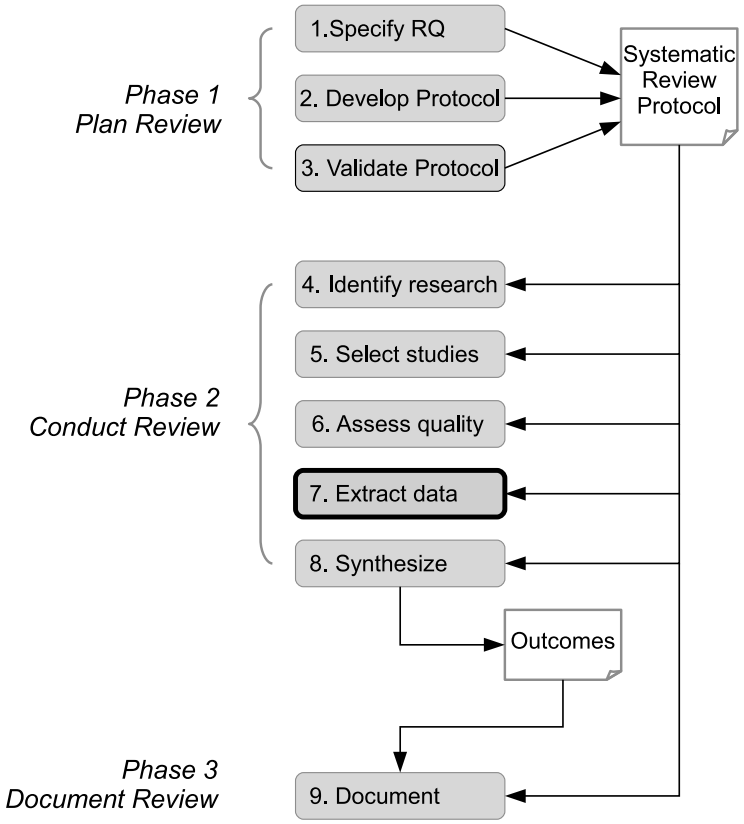

FIGURE 8.1: Data extraction stage of the systematic review process.

about the extractor and date of extraction. What other data is extracted depends very much on the research questions for a review. It should be noted also that sometimes the data needed will be spread across a number of papers.

For *quantitative systematic reviews*, data is most commonly in numerical form although it may include some qualitative data relating, for example, to the context of a primary study (such as the characteristics of participants in an experiment), or to opinions expressed by participants in a primary study, or to recommendations based on the findings of a primary study. This will particularly be the case where vote counting, as opposed to meta-analysis, is to be used as the method of synthesising the outcomes of the primary studies. Here qualitative data can be used to investigate possible explanations for differences in the outcomes of the primary studies (see Section 10.4.6). If the research questions relate, for example, to specific metrics such as defects found, time taken to perform a task or estimates of development costs, then this information is extracted for each study and recorded in a table, a spreadsheet or a special-purpose systematic review support tool.

For *qualitative systematic reviews* and *mapping studies*, data is often extracted in textual form or through the use of a set of classification schemes.

For *qualitative systematic reviews*, information about, for example, fac-

tors, barriers, motivators, recommendations or experiences is extracted and recorded. However, the extraction of this type of data can be more susceptible to bias and the data is less amenable to statistical analysis, making spreadsheets a less useful recording medium. For this type of study, data extraction and synthesis are very closely linked and are likely to be combined within a single process. See, for example, the meta-ethnography process described in Section 10.4.1 and thematic analysis, described in Section 10.4.4.

For *mapping studies*, data extraction and aggregation may be performed iteratively with the classification schemes (for example, relating to the technique used for a particular software engineering task, or the method used in an empirical study) being revised as more knowledge about the topic is gained through the extraction and aggregation process.

The commonly used procedures for data extraction and validation are mostly the same as those described for quality assessment (see Section 7.3), namely:

- Independent extraction by two (or more) reviewers followed by reconciliation through discussion or moderation
- For a lone researcher, taking a test-retest approach and comparing outcomes
- For a lone researcher such as a PhD student, engaging a member of the supervisor team to extract data for a sample of studies

Additionally, especially for qualitative data, extraction may be undertaken as a team, with two or more reviewers working together to reach agreement about the data to be extracted. Whatever approach is taken, if agreement between extractors (or extractions for the test-retest case) is poor then some review of data descriptions, and possibly of the research questions, may be needed.

8.2 Examples of extracted data and extraction procedures

We summarise the data extraction strategies for the following examples:

1. two quantitative systematic reviews which take rather different approaches to data validation,
2. two qualitative systematic reviews: a technology-oriented review and a research-oriented review,
3. a technology-oriented mapping study about the use of Open Source projects in teaching about software engineering.

Examples from quantitative systematic reviews

The first example in this category is a meta-analysis undertaken by Hannay et al. (2009) which aimed to determine the effectiveness of pair programming compared to solo/individual programming. The systematic review addressed the research question:

> How effective, in terms of quality, duration or effort is pair programming compared to solo programming?

Data was extracted about the type of treatment, the type of system, the type of tasks, duration of the study, number of groups, group assignment, type of subjects and their experience with pair programming, number of pairs, number of individuals, outcome variable, means, standard deviations, counts, percentages and p-values.

Data was extracted from all studies by three of the four reviewers and discrepancies were resolved through discussion amongst all four reviewers (that is, the four authors of the review paper). Reviewers performed separate meta-analyses for quality, duration and effort.

The second example of data extraction for a quantitative systematic review is taken from a study carried out by Hall et al. (2012) which reviewed the performance of fault prediction models. This systematic review addressed three research questions relating to context, independent variables included in the fault prediction models and the performance of specific modelling techniques. Three sets of data were extracted with a different procedure being followed for each set:

1. *Context data* - source of the data, maturity, size, application area and programming language of the systems studied. This data was collected by one of the reviewers.

2. *Qualitative data* - data relating to the research questions of the review that was reported in the findings and conclusions of the primary studies. Two reviewers extracted the data independently and discussed disagreements until these were resolved.

3. *Quantitative data* - predictive performance data for all models reported in a study. The form of the data depended on whether the study reported results via categorical or continuous dependent variables. For categorical studies, where possible, data about precision, recall and/or f-measure was recorded. For continuous studies, results in the primary studies were reported in terms of the number of faults predicted in a unit of code using measures based on errors (for example, Mean Standard Error) or differences between expected and observed results (such as Chi Square). Data was extracted and recorded using whatever measure was used in each study. For this data, two reviewers worked together to identify and extract the data from each study.

The reviewers intended to carry out meta-analyses of quantitative data; however, this subsequently proved problematic and so they chose to take a qualitative thematic approach.

Examples from qualitative systematic reviews

The first example in this category is taken from a review which aimed to 'plot the landscape' of reported knowledge about what motivates and demotivates software engineers (Beecham et al. 2008). The research questions for this review are shown in Section 4.3. Data was extracted about how each study answered each of the research questions, with the extractor recording information about:

- Software engineer characteristics,
- Software engineer motivators,
- Software engineering de-motivators,
- External signs or outcomes of motivated software engineers,
- External signs or outcomes of de-motivated software engineers,
- Software engineering as a motivator (e.g. what is motivating about the type of development used; task of coding, testing etc),
- Frameworks/models that reflect how software engineers are motivated.

Endnote[1] was used to record publication details for each paper and a results form was used to record how each study answered each of the research questions. An example of a populated form, which also shows the captured publication details and a potentially relevant study identified through backwards snowballing is described in Appendix B of the review protocol (Beecham, Baddoo, Hall, Robinson & Sharp 2006).

Data was validated by an independent expert on motivation in software engineering who recorded how each paper addressed each research question. Disagreements were discussed and for the small number of cases where they could not be resolved, a third independent researcher arbitrated. The approach taken to synthesising the data was to establish the frequency with which a characteristic or motivator was identified by the primary studies (most of which were surveys).

The second example in this category is a research-oriented qualitative systematic review focusing on the systematic review process (Kitchenham & Brereton 2013). The overall aim of the review was to identify, evaluate and

[1] http://endnote.com/

synthesise research about performing systematic reviews and mapping studies.The specific research questions address are listed in Section 4.3. Extracted data included:

1. Publication details
2. Review-specific data relating to:
    - Type of paper (problem identification and/or problem solution (PI) or lessons learned/opinion survey/discussion paper (E)),
    - Scope of study (mapping study, systematic review/both/update study/other)
    - Summary of aims
    - Main topics covered (multiple selections allowed from a list)
    - Method proposed - name or description
    - Validation performed? - yes or no
    - Actual validation method (as judged by the data extractors) - selected from a list or other (specified),
    - Claimed validation method
    - Summary of main results
    - Any process recommendations (determined by the data extractors)

Publication details were collected and recorded in a spreadsheet by the first author. For review-specific data, some discussion papers, lessons learned papers and opinions surveys (that is, E-type papers) were treated differently from other studies. If a paper covered a very specific topic and had a limited number of results then the data was collected, as for other studies, by both reviewers, and recorded in a spreadsheet. Disagreements were discussed until agreement was reached. If, however, the scope of an E-type paper was very broad (that is, if it covered many aspects of a review and/or included comments from a large variety of subjects), the spreadsheet was only partially completed and an additional data extraction form was used (see Table 8.1). For this third type of (textual) data, the first author extracted the data and the second author checked it.

Example from a mapping study

The example summarised here is a mapping study focusing on the use of Open Source projects in software engineering education (Nascimento, Cox, Almeida, Sampaio, Almeida Bittencourt, Souza & Chavez 2013). The study addressed three research questions:

1. "How are Open Source projects used in software engineering education?"

TABLE 8.1: Form for Recording Extra Textual Data (adapted from Kitchenham and Brereton (2013))

Issue Id	
Issue text	For each issue/problem raised/solution proposed specify this using the same text as the paper authors
Type	Advice, problem/challenge or value (benefit)
Suggestion for guidelines?	Yes or no
Novice issues?	Yes or no
Education issues?	Yes or no
Location in paper	Page number or table number/id
Stage in review process addressed	Research question/protocol/search/selection/data extraction/ quality assessment/data aggregation/synthesis/reporting
Importance	A ratio indicating number of votes out of the maximum possible number or a textual indication of relative importance
Related issue	Reference to any related issue
Comments	

2. "Are there any initiatives that combine open source projects with active learning in software engineering courses?"

3. "How is student learning assessed in such initiatives?"

In addition to publication details, data was extracted according to the following classification facets:

- *Software engineering topic* - based on the SWEBOK knowledge areas[2],
- *Research type* - using a set of approaches to research based on those of Petersen, K. Petersen, Feldt, Mujtaba & Mattsson (2008),
- *Learning approach* - using categories: active learning (general), case-based learning, game-based learning, peer/group/team learning, problem-/project-/inquiry-based learning, studio-based learning and other,
- *Assessment perspective* - where there is assessment, this can, for example, be from the perspective of the student (through peer or self assessment) or from the perspective of teaching staff,
- *Assessment type* - covering methods of assessing students (such as by

[2]www.swebok.org

examination, through developed software artifacts, interviews, exercises or surveys).

Nascimento et al. indicate that due to lack of time, data was extracted (that is the primary studies were classified) by only one reviewer. The authors recognise this as a limitation of their mapping study.