



CACIC2013

XIX Congreso Argentino de Ciencias de la Computación

XVII Escuela Internacional de Informática

Ingeniería del software basada en la evidencia (IBSE) (PARTE II)

Marcela Genero Bocco

Grupo Alarcos

marcela.genero@uclm.es

Universidad de Castilla-La Mancha

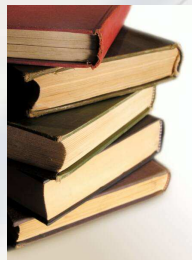
Escuela Superior de Informática

Departamento de Tecnologías y Sistemas de Información

Ciudad Real - España



How to Perform Systematic Reviews: Theory and Examples




Marcela Genero Bocco

ALARCOS Research group

Department of Technologies and Information Systems

University of Castilla-La Mancha


Marcela.Genero@uclm.es



Agenda


- Introduction
- Motivation
- Review process
- Lessons learned
- Conclusions→A SLR of SLRs
- Relevant literature

3



Introduction

What is a SLR?



- A means of **evaluating** and **interpreting all available research relevant** to a particular **research question** or phenomenon of interest.
- SLRs aim to **present a fair evaluation of a research topic** by using a trustworthy, rigorous, and auditable methodology.

4



Introduction

Evidence Based Medicine

Our vision is that healthcare decision-making throughout the world will be informed by high-quality, timely research evidence


<http://www.cochrane.org>



**THE COCHRANE
COLLABORATION®**

Working together to provide the best evidence for health care to [The Cochrane Library](#)

5



Introduction

SLRs are getting relevance in SE


Examples of SLRs of SLRs

- **Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey, J., Linkman, S. (2009).** Systematic literature reviews in software engineering – A systematic literature review. Information and Software Technology 51, 7–15.
- **B. Kitchenham et al. (2010).** Literature reviews in software engineering – a tertiary study, Information and Software Technology 52 (8) 792–805.
- **Fabio Q.B. da Silva, André L.M. Santos, Sérgio Soares, A. César C. França, Cleviton V.F. (2011).** Six Years of Systematic Literature Reviews in Software Engineering: An Updated Tertiary Study Information and Software Technology. 53 (9), 899-913.
- **Zhang, H., Ali Babar, M. (2013).** Systematic reviews in software engineering: An empirical investigation. Information and Software Technology, 55(7),1341–1354. → **142 (72 SLR y 72 SMS)**

6





Motivation



- Most **research starts**, or should start with a **literature review of some sort**.
- Unless a **literature review is thorough and fair**, it is of little scientific value.
- A systematic review **synthesizes existing work** in a manner that is fair and seem to be fair.



8



Motivation

- SLRs must be undertaken in accordance with a predefined **search strategy**, that must allow the completeness of the search to be assessed.
- Researchers performing a SLR must make every effort **to identify and report research that does not support their preferred research hypothesis** as well as identifying and reporting research that supports it.


9



Motivation

- To summarize the existing evidence concerning a treatment or technology (e.g. **to summarize the empirical evidence of the benefits and limitations of a specific agile method**).
- To **identify any gaps** in current research in order to suggest areas for further investigation.
- To provide a **framework/background** in order to appropriately position new research activities.
- To **examine** the extent to which **empirical evidence supports/contradicts theoretical hypotheses**, or even to assist the generation of new hypotheses.

10

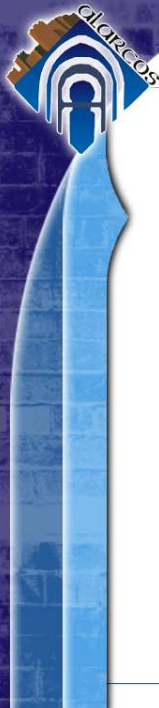


Motivation

Prons & Cons

- The well-defined methodology makes it **less likely** that the **results** of the literature **are biased**.
- They can **provide information** about the effects of **some phenomenon** across a **wide range** of **settings** and **empirical methods**.
- In the case of quantitative studies, it is **possible** to combine data using **meta-analytic techniques**.
- The major **disadvantage** of systematic literature reviews is that they require considerably **more effort** than traditional literature reviews.

11




Motivation

Other reviews

- Individual studies contributing to a systematic review are called **primary** studies
- A systematic review is a form of **secondary** study.
- A systematic review that analyzes the information in SLR is a **tertiary** study.

12




Motivation

Other reviews

Systematic Mapping Studies

- When it is discovered that **very little evidence is likely to exist or that the topic is very broad** then a systematic mapping study may be a more appropriate exercise than a systematic review.
- A systematic mapping study allows **the evidence in a domain to be plotted at a high level of granularity**.
- This allows for the **identification of evidence clusters and evidence deserts** to direct the focus of future systematic reviews and to identify areas for more primary studies to be conducted.

13




Introduction

Other reviews

Systematic Mapping Studies

Elementos	SMS	SLR
Objetivos	Clasificación y análisis temático de la literatura sobre un tema específico de la ingeniería de software	Identificar las mejores prácticas con respecto a procedimientos, tecnologías, métodos o herramientas específicas, mediante la agregación de información obtenida a partir de estudios empíricos.
Pregunta de investigación	Genérica, relacionada con tendencias de investigación, como por ejemplo: qué investigadores, cuantos estudios, qué tipo de estudios, etc.	Específica, relacionada con resultados de estudios empíricos, como por ejemplo: ¿Es mejor el método/tecnología A que la B?
Proceso de búsqueda	Definido por el área de estudio o de interés	Definido por la pregunta de investigación la cual identifica la tecnología específica que está siendo investigada
Alcance	Amplio – se incluyen todos los artículos sobre un área de interés, pero sólo se extrae de ellos datos para clasificarlos	Centrado– solo se incluyen artículos que contengan estudios empíricos relacionados con las preguntas de investigación y se extrae de ellos información detallada sobre los resultados obtenidos en cada uno de ellos

14




Motivation

Other reviews

Systematic Mapping Studies

Elementos	SMS	SLR
Requisitos de la estrategia de búsqueda	A menudo menos estricta si sólo se buscan tendencias, por ejemplo se puede buscar solo en un conjunto específico de publicaciones, limitándolas a artículos de revistas, o limitándolas a una o dos bibliotecas digitales	Extremadamente exigente– se deben encontrar todos los artículos relevantes. Generalmente además de buscar en las fuentes establecidas, puede ser necesario buscar en las referencias de los estudios primarios seleccionados o consultar a los expertos para incluir el mayor número de artículos posible
Evaluación de la calidad	No es esencial. Al incluir tanto estudios teóricos como empíricos de cualquier tipo, suele ser muy difícil definir un mecanismo de evaluación	Es importante asegurarse de que los resultados se basan en la evidencia de mejor calidad
Resultados	Un conjunto de artículos relacionados con un área de interés clasificados en una serie de dimensiones, especificando el número total de artículos en cada dimensión	Se agregan los resultados de los estudios empíricos para contestar a las preguntas de investigación

15





Motivation

Other reviews

Tertiary Reviews

- In a domain where a number of systematic reviews exist already it may be possible to conduct a tertiary review, which is **a systematic review of systematic reviews**, in order to answer wider research questions.
- A tertiary review **uses exactly the same methodology** as a standard systematic literature review.
- It is potentially **less resource intensive** than conducting a new systematic review of primary studies but is dependent on sufficient systematic reviews of a high quality being available.

16






Introduction

Running examples


- (AGILE) Tore Dybå, Torgeir Dingsøy. **Empirical studies of agile software development: A systematic review**. Information & Software Technology 50(9-10): 833-859 (2008).
- (GSE) Darja Šmite, Claes Wohlin, Tony Gorschek, Robert Feldt. (2010). **Empirical evidence in global software engineering: a systematic review**. Empirical Software Engineering, 15, 91–118. (SLR)
- (UML) Marcela Genero, Ana M. Fernández, H. James Nelson, Geert Poels, Mario Piattini. (2011). **A Systematic Literature Review on the Quality of UML Models**. Journal of Database Management, 23(3), 46-70. (SMS)

17

Motivation

Running examples (AGILE)



ELSEVIER

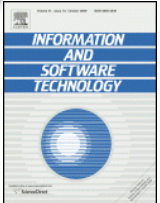
Available online at www.sciencedirect.com

ScienceDirect

Information and Software Technology 50 (2008) 833–859

INFORMATION AND SOFTWARE TECHNOLOGY

www.elsevier.com/locate/infsof



(más referenciado)

Empirical studies of agile software development: A systematic review

Tore Dybå*, Torgeir Dingsøy

SINTEF ICT, S.P. Andreassen 15B, NO-7465 Trondheim, Norway

Received 22 October 2007; received in revised form 22 January 2008; accepted 24 January 2008
Available online 2 February 2008

Abstract

Agile software development represents a major departure from traditional, plan-based approaches to software engineering. A systematic review of empirical studies of agile software development up to and including 2005 was conducted. The search strategy identified 1990 studies, of which 36 were identified as empirical studies. The studies were grouped into four themes: introduction and adoption, human and social factors, perceptions on agile methods, and comparative studies. The review investigates what is currently known about the benefits and limitations of, and the strength of evidence for, agile methods. Implications for research and practice are presented. The main implication for research is a need for more and better empirical studies of agile software development within a common research agenda. For the industrial readership, the review provides a map of findings, according to topic, that can be compared for relevance to their own settings and situations.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Empirical software engineering; Evidence-based software engineering; Systematic review; Research synthesis; Agile software development; XP; Extreme programming; Scrum

Contents

1. Introduction	834
2. Background – agile software development	834

18




EXAMPLE

Motivation

Running examples (GSE)

Empir Software Eng (2010) 15:91–118
DOI 10.1007/s10664-009-9123-y



Empirical evidence in global software engineering: a systematic review

Darja Šmite · Claes Wohlin · Tony Gorschek ·
Robert Feldt

Published online: 15 December 2009
© Springer Science+Business Media, LLC 2009
Editor: James Herbsleb

Abstract Recognized as one of the trends of the 21st century, globalization of the world economies brought significant changes to nearly all industries, and in particular it includes software development. Many companies started global software engineering (GSE) to benefit from cheaper, faster and better development of software systems, products and services. However, empirical studies indicate that achieving these benefits is not an easy task. Here, we report our findings from investigating empirical evidence in GSE-related research literature. By conducting a systematic review we observe that the GSE field is still immature. The amount of empirical studies is relatively small. The majority of the studies represent problem-oriented reports focusing on different aspects of GSE management rather than in-depth analysis of solutions for example in terms of useful practices or techniques. Companies are still driven by cost reduction strategies, and at the same time, the most frequently discussed recommendations indicate a necessity of investments in travelling and socialization. Thus, at the same time as development goes global there is an ambition to minimize geographical, temporal and cultural separation. These are normally integral parts of cross border collaboration. In summary, the systematic review results in several

19

EXAMPLE

Motivation

Running examples (UML)

46 Journal of Database Management, 22(3), 46-70, July-September 2011

A systematic mapping study

RESEARCH REVIEW

A Systematic Literature Review on the Quality of UML Models

Marcela Genero, University of Castilla-La Mancha, Spain
Ana M. Fernández-Saez, University of Castilla-La Mancha, Spain
H. James Nelson, Southern Illinois University, USA
Geert Poels, Faculty of Economics and Business Administration, Ghent University, Belgium
Mario Piattini, University of Castilla-La Mancha, Spain

ABSTRACT

The quality of conceptual models directly affects the quality of the understanding of the application domain and the quality of the final software products that are ultimately based on them. This paper describes a systematic literature review (SLR) of peer-reviewed conference and journal articles published from 1997 through 2009 on the quality of conceptual models written in UML, undertaken to understand the state-of-the-art, and then identify any gaps in current research. Six digital libraries were searched, and 266 papers dealing specifically with the quality of UML models were identified and classified into five dimensions: type of model quality, type of evidence, type of research result, type of diagram, and research goal. The results indicate that most research focuses on semantic quality, with relatively little on semantic completeness; as such, this research examines new modeling methods vs. quality frameworks and metrics, as well as quality assurance vs. understanding quality issues. The results also indicate that more empirical research is needed to develop a theoretical understanding of conceptual model quality. The classification scheme developed in this paper can serve as a guide for both researchers and practitioners.

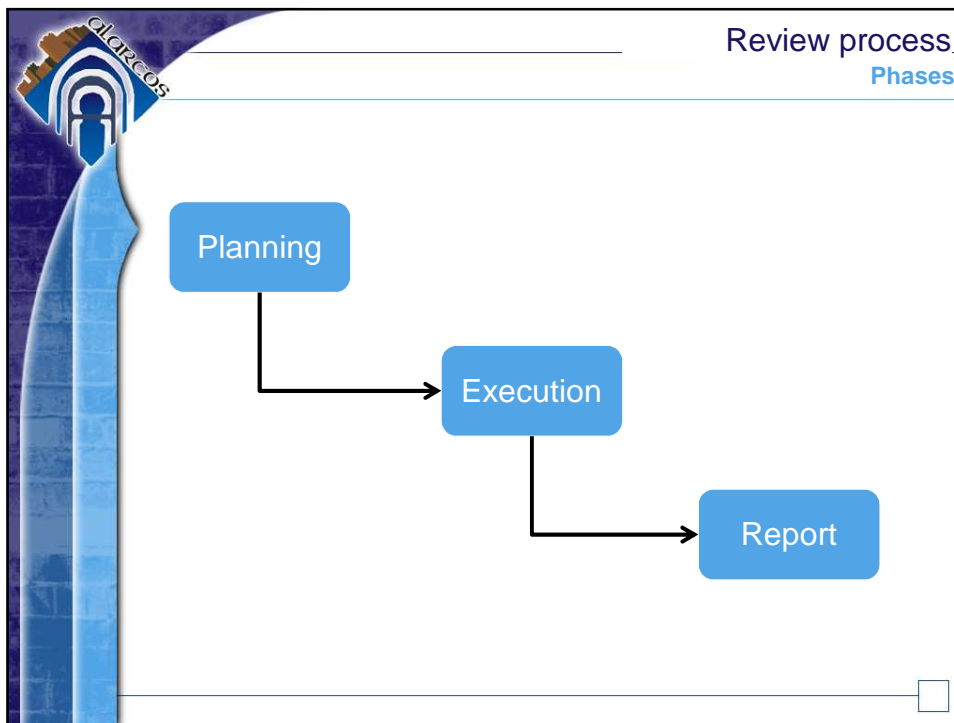
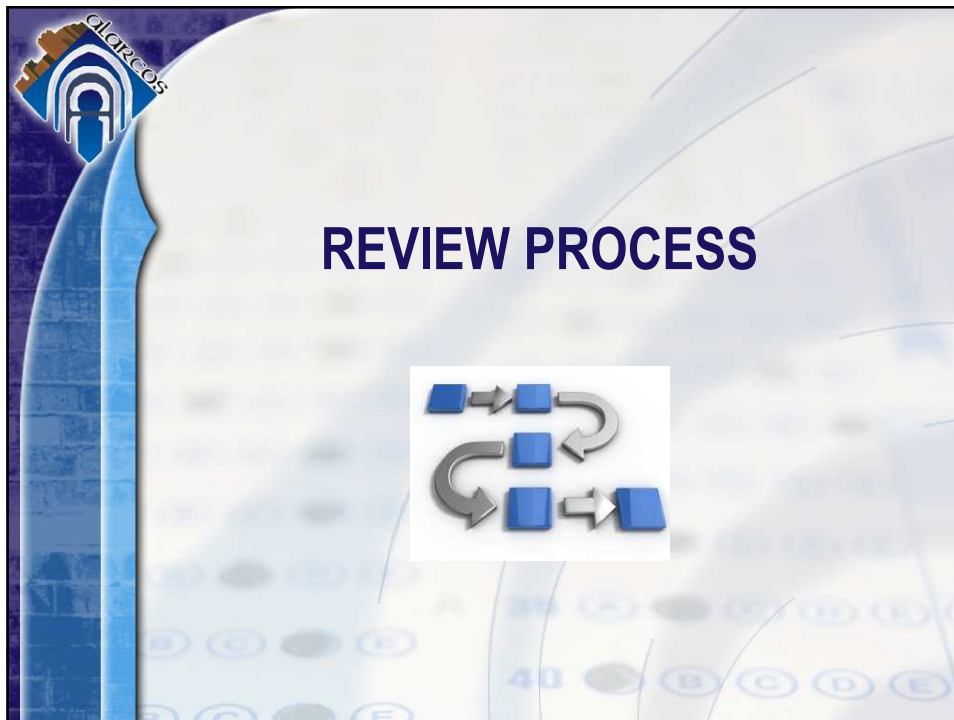
Keywords: Conceptual Model Quality, Conceptual Models, Software, Systematic Literature Review, Unified Modeling Language (UML)


INTRODUCTION

Software is becoming increasingly complex. So complex, in fact, that it is widely acknowl-

the customer, of deepening the understanding of how software works, and ultimately of reducing the complexity of software, is through the use of models (Thomas, 2004). Over the years, we

20



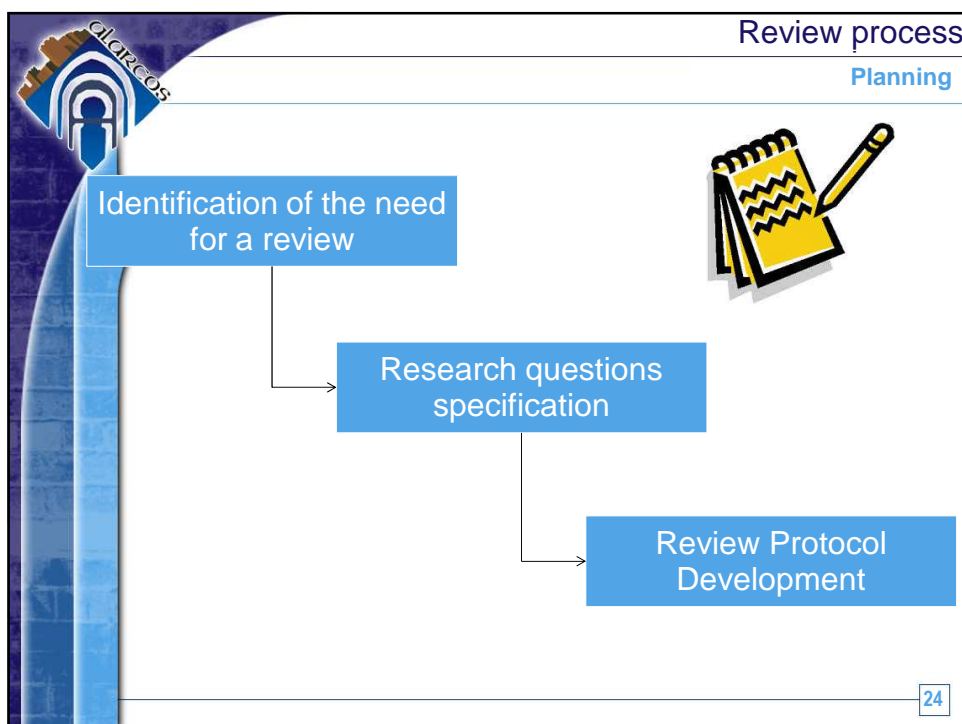



Review process

Phases


- The stages may appear to be sequential, but it is important to recognise that many of the **stages involve iteration**.
- **Many activities** are initiated during the protocol development stage, and **refined** when the review proper takes place. For example:
 - **The inclusion and exclusion criteria** are initially specified when the protocol is drafted but may be refined after quality criteria are defined.
 - **Data extraction forms** initially prepared during construction of the protocol will be amended when quality criteria are agreed.
 - **Data synthesis methods** defined in the protocol may be amended once data has been collected.

23





Review process


 Planning

Identification of the need for a review


Researchers should identify and review any existing systematic reviews of the phenomenon of interest against appropriate evaluation criteria:

- What are the review's objectives?
- What sources were searched to identify primary studies? Were there any restrictions?
- What were the inclusion/exclusion criteria and how were they applied?
- What criteria were used to assess the quality of primary studies and how were they applied?
- How were the data extracted from the primary studies?
- How were the data synthesised? How were differences between studies investigated? How were the data combined? Was it reasonable to combine the studies? Do the conclusions flow from the evidence?

25



Review process

 Planning


Research question(s)

Specifying the research questions is the most important part of any systematic review.

The review questions drive the entire systematic review methodology:


- The search process must identify primary studies that address the research questions.
- The data extraction process must extract the data items needed to answer the questions.
- The data analysis process must synthesise the data in such a way that the questions can be answered.

26



Review process

Research question(s)


 Planning

The critical issue in any systematic review is to ask the right question.

The right question is usually one that:


- Is **meaningful** and important to **practitioners** as well as **researchers**.
- Will lead either to **changes in current software engineering practice** or to increased confidence in the value of current practice.
- **Identify discrepancies** between commonly held beliefs and reality.

27





Review process

Question types


 Planning

- **Assessing** the effect of a software engineering **technology**.
- Assessing the **frequency** or rate of a **project development factor such** as the adoption of a technology, or the frequency or rate of project success or failure.
- Identifying **cost and risk factors** associated with a technology.
- Identifying **the impact of technologies** on reliability, performance and cost models.
- **Cost benefit analysis** of software technologies.

28

Review process

 Planning



RQ Example (AGILE)

RQ1. What is currently known about the benefits and limitations of agile software development?


RQ2. What is the strength of the evidence in support of these findings?

RQ3. What are the implications of these studies for the software industry and the research community?

29



Review process

 Planning


RQ Example (GSE)

- **Question 1:** What is the state-of-the-art in empirical studies of GSE?
 - Who is Involved in GSE?
 - Where are the Development Sites Located?
 - What is Studied in GSE?
 - How Successful are the Cases Reported in Literature?
 - Why are Companies Involved in GSE?
- **Question 2:** What is the strength of the empirical evidence reflected in the empirical GSE?

30



Review process

Planning

RQ Example (UML)

RQ1. Which type of UML model quality has been investigated by researchers?



RQ2. Which research methods are used in research on UML model quality?

RQ3. What is the nature of the research results on UML model quality?


RQ4. Which research goals are aimed at in research on UML model quality?

RQ5. Which type of UML diagrams is the focus of the research on UML model quality?

31




Review process


Planning

Review protocol

- A review protocol specifies the methods that will be used to undertake specific systematic review, **to reduce the possibility of researcher bias.**
- The components of a protocol include all the elements of the review plus some additional planning information.

32





Review process
 **Planning**

Review protocol development

- **Background.** The rationale for the survey.
- The **research questions** that the review is intended to answer
- The **search strategy** that will be used to search for primary studies including **search terms** and **resources to be searched**. Resources include digital libraries, specific journals, and conference proceedings.
- **Study selection criteria.** Study selection criteria are used to determine which studies are **included** in, or **excluded** from, a systematic review. It is usually helpful to pilot the selection criteria on a subset of primary studies.

33




Review process
 **Planning**


Review protocol development

- **Study selection procedures.** The protocol should describe how the selection criteria will be applied e.g. how many assessors will evaluate each prospective primary study, and how disagreements among assessors will be resolved.
- **Study quality assessment checklists and procedures.** The researchers should develop quality checklists to assess the individual studies. The purpose of the quality assessment will guide the development of checklists.
- **Data extraction strategy.** This defines how the information required from each primary study will be obtained.

34




Review process

 Planning


Review protocol development

- **Synthesis of the extracted data.** This defines the synthesis strategy. This should clarify whether or not a formal meta-analysis is intended and if so what techniques will be used.
- **Dissemination strategy.** How and where the results will be published or disseminated.
- **Project timetable.** This should define the review schedule.

35



Review process


 Planning

Review protocol: Search strategy


Search strategies are usually iterative and benefit from:

- **Preliminary searches** aimed at both identifying existing systematic reviews and assessing the volume of potentially relevant studies.
- **Trial searchers** using various combinations of search terms derived from the research question
- **Reviews of research results**
- **Consultations with experts** in the field

36



Review process


 Planning

Review protocol: Search strategy


- Initial searches for primary studies can be undertaken initially using electronic databases but this is not sufficient.
- Other sources of evidence must also be searched (sometimes manually) including:
 - Reference lists from relevant primary studies and review articles
 - Journals (including company journals such as the IBM Journal of Research and Development), grey literature (i.e. technical reports, work in progress) and conference proceedings
 - Research registers
 - The Internet

To identify expert researchers

37



Review process



 Planning

Review protocol: Search string


Constructed using the following steps:

- Define the major terms
- Identify alternative spellings, synonyms, related terms for major terms.
- Check the keywords in any relevant papers we already had.
- Use the Boolean, to incorporate alternative spellings, synonyms, related terms.
- Use the Boolean AND to link the major terms

38

Review process



Planning

Review protocol: Search String

An example (AGILE)

(1) agile AND software

(2) extreme programming

(3) xp AND software

(4) scrum AND software

(5) crystal AND software AND (clear OR orange OR red OR blue)

(6) dsdm AND software



(7) fdd AND software

(8) feature AND driven AND development AND software


(9) lean AND software AND development

1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9

39

Review process



Planning

Review protocol: Search String

An example (GSE)

The final search strings were based on the experience from the pilot searches and consisted of a Boolean expression:

(A1 OR A2 OR A3 OR A4) AND (B1 OR B2 OR B3 OR B4), where

A1—global software development

A2—global software engineering

A3—distributed software development

A4—distributed software engineering


B1—empirical


B2—industrial


B3—experiment

B4—case study

40





Review process
 **Planning**


Review protocol: Search String


An example (UML)


Major terms	Alternative terms
Quality	quality OR consistency OR maintainability OR understandability OR completeness OR comprehension OR comprehensibility OR testability OR defect OR effectiveness OR complexity OR readability OR metric OR measure OR efficiency OR validation OR verification OR layout
UML	UML OR Unified Modeling Language
Representation	Representation OR diagram OR model

Therefore, we defined the following **search string**:
 (UML OR UNIFIED MODELING LANGUAGE) **AND** (REPRESENTATION OR DIAGRAM OR MODEL) **AND** (QUALITY OR CONSISTENCY OR MAINTAINABILITY OR UNDERSTANDABILITY OR COMPLETENESS OR COMPREHENSION OR COMPREHENSIBILITY OR TESTABILITY OR DEFECT OR EFFECTIVENESS OR COMPLEXITY OR READABILITY OR EFFICIENCY OR VALIDATION OR VERIFICATION OR LAYOUT)

41





Review process
 **Planning**

Review protocol: Search Sources



An example (AGILE)

- ACM Digital Library
- Compendex
- IEEE Xplore
- ISI Web of Science
- Kluwer Online
- ScienceDirect – Elsevier
- SpringerLink
- Wiley Inter Science Journal Finder


In addition, we hand-searched all volumes of the following conference proceedings for research papers:

- XP
- XP/Agile Universe
- Agile Development Conference

42



Review process



Planning

Review protocol: Search Sources


An example (GSE)

- Compendex,
- IEEE Xplore,
- Springer Link,
- ISI Web of Knowledge,
- ScienceDirect,
- Wiley Inter Science Journal Finder,
- ACM Digital Library

43



Review process


Planning

Review protocol: Search Sources

An example (UML)

- SCOPUS database,
- Science@Direct with the subject Computer Science,
- Wiley InterScience with the subject of Computer Science, I
- IEEE Digital Library,
- ACM Digital Library,
- SPRINGER database.

44



Review process


Planning

Review protocol

Bibliography management and document retrieval

- Bibliographic packages such as Reference Manager, Endnote are very useful to manage the large number of references that can be obtained from a thorough literature research.
- Once reference lists have been finalized the full articles of potentially useful studies will need to be obtained.

45



Review process


Planning

Primary studies selection


Study selection criteria

- **Selection criteria** should be decided during the **protocol definition**.
- **Inclusion and exclusion criteria** should be based on the **research question**.
- They should be piloted to ensure that they can be reliably interpreted and that they classify studies correctly.

46



Review process

 Planning

Primary studies selection


Study selection process

- Initially, selection criteria should be interpreted liberally, so that unless studies can be clearly excluded based on titles and abstracts, full copies should be obtained.
- Final inclusion/exclusion decisions should be made after the full texts have been retrieved.
- Maintain a list of excluded studies identifying the reason for exclusion.


Reliability of inclusion decisions

- When two or more researchers assess each paper, agreement between researchers must be reached


47



Review process

 Planning

Primary studies selection





Inclusion criteria: Example (AGILE)


Inclusion criteria:

- Present empirical data on agile software development and passed the minimum quality threshold.
- Studies of both students and professional software developers were included.
- XP, Scrum, Crystal, DSDM, FDD, and Lean.
- qualitative and quantitative research studies, published up to and including 2005.
- Studies written in English were included.

48



Review process

Planning



Primary studies selection

Inclusion criteria: Example (AGILE)


Exclusion criteria:

- Editorials, prefaces, article summaries, interviews, news, reviews, correspondence, discussions, comments, reader's letters and summaries of tutorials, workshops, panels, and poster sessions.
- If their focus, or main focus, was not agile software development or if they did not present empirical data.
- "Lessons learned" papers (papers without a research question and research design) and papers merely based on expert opinion

49



Review process

Planning


Primary studies selection


Inclusion criteria: Example (UML)

Inclusion criteria:

- Papers which dealt with UML and the tangible results of the modelling process (the UML diagram),
- were written in English,
- and were published between 1997 and 2009.

50






Primary studies selection

Inclusion criteria: Example (UML)


Review process

 Planning

Exclusion criteria:

- pure discussion and opinion papers, studies available only in the form of abstracts or PowerPoint presentations,
- duplicates (for example, the same paper included in more than one database or in more than one journal),
- research focusing issues other than UML model quality (for example, functional size measurement), or where quality is mentioned only as a general introductory term in the paper's abstract and an approach
- or other type of proposal related to quality is not amongst the paper's contributions.
- Papers were also excluded if they dealt with the quality and complexity of UML as a language (for example, how to make UML the language simpler) rather than on the quality and complexity of the models produced by UML, and finally if the paper was a summary of a workshop.


51



Quality assessment


It is generally considered important to assess the “quality” of primary studies

Review process


 Planning

- To provide still more detailed inclusion/exclusion criteria.
- To investigate whether quality differences provide an explanation for differences in study results.
- As a means of weighting the importance of individual studies when results are being synthesised.
- To guide the interpretation of findings and determine the strength of inferences.
- To guide recommendations for further research.

52



Review process

 Planning

Development of quality instruments

It is advisable to :


- **build checklists**
- **assign numerical scales** → numerical assessments of quality can be obtained.

Checklists are also developed by considering bias and validity problems that can occur at the different stages in an empirical study: Design, Conduct, Analysis, and Conclusions.


Kitchenham et al (2007) in the technical report provide:


- A quality checklist for quantitative studies
- A quality checklist for qualitative studies

53



Review process


 Planning




Quality assessment: example (AGILE)

1. Is the paper based on research (or is it merely a “lessons learned” report based on expert opinion)?
2. Is there a clear statement of the aims of the research?
3. Is there an adequate description of the context in which the research was carried out?
4. Was the research design appropriate to address the aims of the research?
5. Was the recruitment strategy appropriate to the aims of the research?
6. Was there a control group with which to compare treatments?
7. Was the data collected in a way that addressed the research issue?
8. Was the data analysis sufficiently rigorous?
9. Has the relationship between researcher and participants been considered to an adequate degree?
10. Is there a clear statement of findings?
11. Is the study of value for research or practice?

54



Review process

 Planning

Data extraction


- Design of data extraction forms

Contents:


- Name of Review
- Date of Data extraction
- Title, authors, journal, publication details
- Space for additional notes


- Data extraction procedures
- Multiple publications of the same data

55



Review process



 Planning




Data extraction form: Example (GSE)

- **General information:** metadata of the papers
- **Relevance:** Technical and methodological flaws of the study
- **Empirical Background, GSE Background:** Information about the sample, population or participants
- **Study:** Central focus of the study and the problem addressed
- **Qualitative evaluation:** Review of the key results

56

Review process



 Planning

Data extraction form: Example (AGILE)


Study description

1. Study identifier Unique id for the study
2. Date of data extraction
3. Bibliographic reference (Author, year, title, source)
4. Type of article (Journal article, conference paper, workshop paper, book section)
5. Study: aims What were the aims of the study?
6. Objectives: What were the objectives?
7. Design of study: Qualitative, quantitative (experiment, survey, case study, action research)
8. Research hypothesis: Statement of hypotheses, if any
9. Definition of agile software development given in study: Verbatim from the study

57

Review process


 Planning


Data extraction form: Example (AGILE)

Study description

10. Sample description Size, students, professionals (age, education, experience)
11. Setting of study Industry, in-house/supplier, products and processes used
12. Control group Yes, no (number of groups, sample size)
13. Data collection How was the data obtained? (questionnaires, interviews, forms)
14. Data analysis How was the data analyzed? (qualitative, quantitative)


58





Data extraction form: Example (AGILE)


Review process

 **Planning**

Study findings


1. Findings and conclusions: What were the findings and conclusions? (verbatim from the study)
2. Validity: Limitations, threats to validity
3. Relevance: Research, practice

59



Data synthesis


Review process

 **Planning**


- **Descriptive synthesis**
 - Extracted information should be tabulated
- **Quantitative synthesis**
 - Descriptive statistics
 - Meta-analysis
- **Qualitative synthesis**
 - Narrative synthesis
 - Thematic synthesis
 - Grounded theory
 - Case survey
 - Content analysis
 - Meta-ethnography

} (most used)

60




Review process

 Planning


Protocol review

- The protocol is a critical element of any systematic review.
- Researchers must agree a procedure for reviewing the protocol.
- If appropriate funding is available, a group of independent experts should be asked to review the protocol.
- The same experts can later be asked to review the final report.
- PhD or master students should present their protocol to their supervisors for review and criticism.

61



Review process

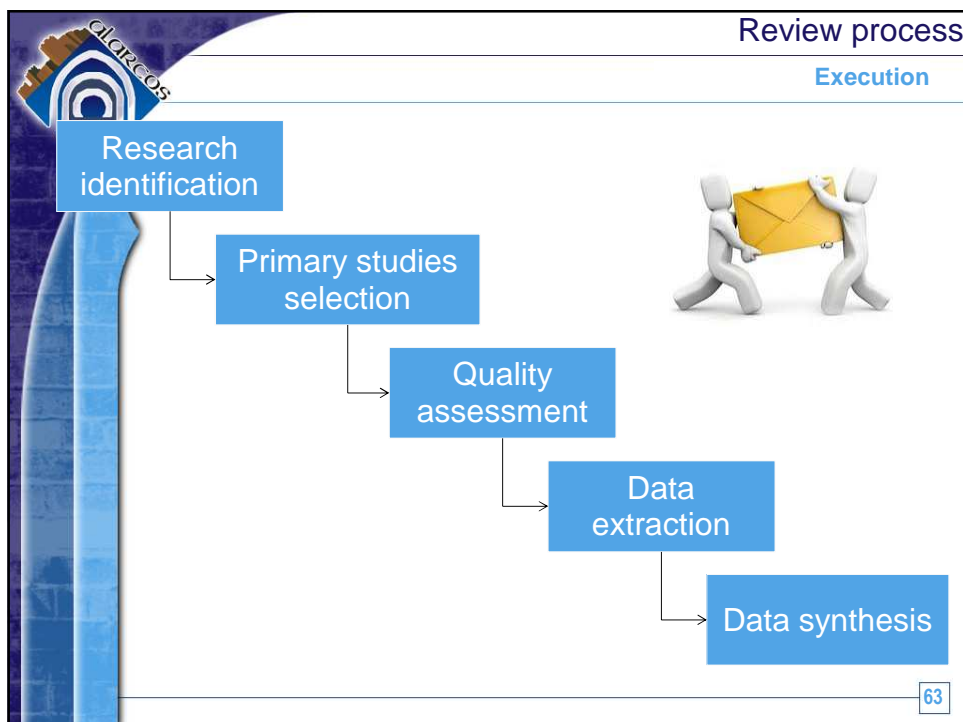
 Planning

Lesson learned from protocol construction

Brereton et al. (2007) identify a number of issues that researchers should anticipate during protocol construction:

- A **pre-review mapping study** may help in scoping research questions.
- Expect to **revise questions during protocol development**, as understanding of the problem increases.
- **All** the systematic review team **members** need to take an **active** part in developing the review protocol, so they understand how to perform the data extraction process.
- **Piloting the research protocol** is essential.
 - It will find mistakes in the data collection and aggregation procedures.
 - It may also indicate the need to change the methodology intended to address the research questions including amending the data extraction forms and synthesis methods.

62




Review process

Execution

Research identification

- Search the primary studies following the search strategy
- It could be necessary:
 - To refine the search string
 - To add search sources
 - To change the search eperiod
- Save the searches, the meta-data, teh abstract, in Bibliography managemment systems (EndNote, BibTex, etc.)
- Detect duplications (found in several search sources)

64



Review process
 Execution

Research identification

Documenting the search

The **process** of performing a systematic review must be **transparent** and **replicable**:

- The review **must be documented** in sufficient detail for readers to be able to assess the thoroughness of the search.
- The search should be documented as it occurs and **changes noted and justified**.

65




Review process
 Execution

Research identification

Documenting the search

Data Source	Documentation
Electronic database	Name of database Search strategy for each database Date of search Years covered by search
Journal Hand Searches	Name of journal Years searched Any issues not searched
Conference proceedings	Title of proceedings Name of conference (if different) Title translation (if necessary) Journal name (if published as part of a journal)
Efforts to identify unpublished studies	Research groups and researchers contacted (Names and contact details) Research web sites searched (Date and URL)
Other sources	Date Searched/Contacted URL Any specific conditions pertaining to the search

66




Review process
 Execution

Primary studies selection

- The selection of primary studies must **take into account** the **inclusion/exclusion criteria** and the procedure defined in the protocol.
- If any primary study it is not available to be downloaded, you can contact with other researchers or the authors.
- In this task the list of the selected primary studies is obtained.
- The primary studies must be managed by a bibliographic management tool (EndNote, BibTex, etc.).
- It is advisable to keep the list of the excluded studies and the motivation for this exclusion.

67



Review process
 Execution

Quality assessment

- Carry out the quality assessment according to the instrument defined in the protocol
- It could be necessary to exclude the primary study that does not reach the required level of quality

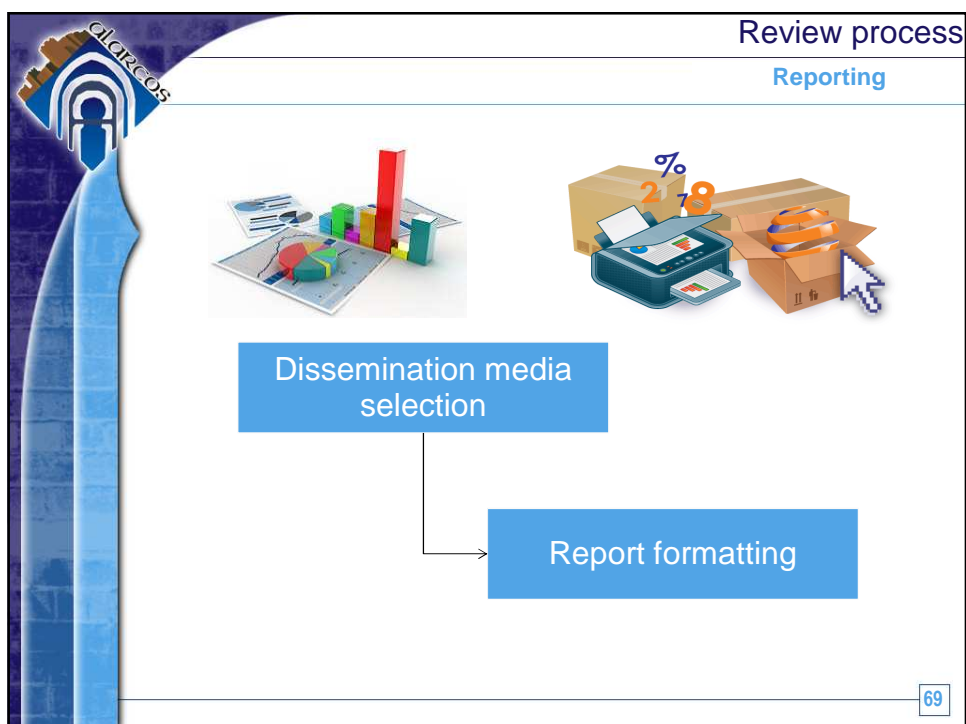
Data extraction

- Fill the data extraction form defined in the protocol
- The data extraction could be checked by other researcher
- Discrepancies must be solved
- If duplicates are found the most complete and recent study must be considered

Data synthesis

- The extracted data is synthesised using the techniques established in the protocol for answering the formulated research questions

68




Review process

Reporting

Dissemination strategy

- It is important to **communicate the results** of a systematic review effectively.
- Most guidelines recommend **planning the dissemination** strategy when preparing the systematic review protocol.

70



Review process

Reporting

Dissemination venues


Journals

- Information and Software Technology
- IEEE Transactions on Software Engineering
- Empirical Software Engineering
- IEEE Software – Voice of Evidence column
- ...

Conferences

- ESEM (Empirical software engineering and measurement)
- EASE (Evaluation and assessment in software engineering)
- ...

71



Review process


Reporting

Dissemination strategy

If the results of a systematic review are intended to influence practitioners, other forms of dissemination are necessary:

- Practitioner journals and magazines,
- Press releases to popular and specialized press,
- Short summary leaflets,
- Posters,
- Web pages,
- Direct communication to affected bodies.

72



Review process

Reporting

Report formatting

- Usually systematic reviews will be reported in at least two formats:

In a **technical report** or in a section of a **PhD thesis**.

In a **journal or conference paper**.
- A journal or conference paper will normally **have a size restriction**.
- In order to ensure that readers are able to properly evaluate the rigour and validity of a systematic review, **journal papers should reference a technical report or thesis that contains all the details**.

73

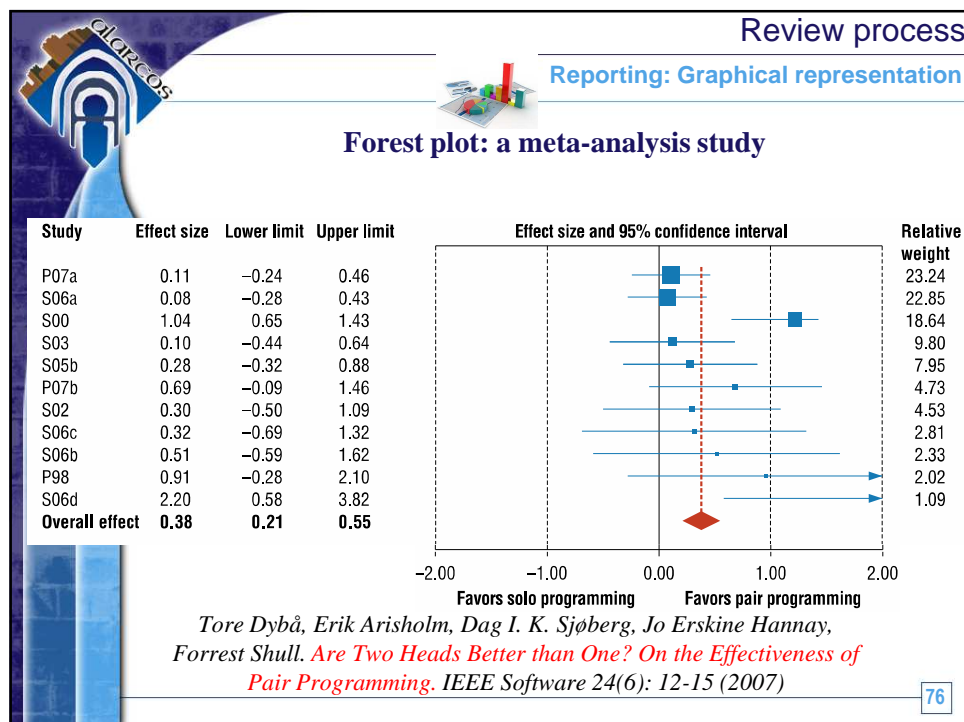
Table 9 Structure and contents of reports of systematic reviews

Section	Subsection	Scope
Title*		
Authorship*		
Executive summary or Structured Abstract*	Context	The importance of the research questions addressed by the review
	Objectives	The questions addressed by the systematic review
	Methods	Data Sources, Study selection, Quality Assessment and Data extraction
	Results	Main finding including any meta-analysis results and sensitivity analyses.
	Conclusions	Implications for practice and future research
Background		Justification of the need for the review. Summary of previous reviews
Review questions		Each review question should be specified
Review Methods	Data sources and search strategy	
	Study selection	
	Study quality assessment	
	Data extraction	
	Data synthesis	
Included and excluded studies		Inclusion and exclusion criteria List of excluded studies with rationale for exclusion

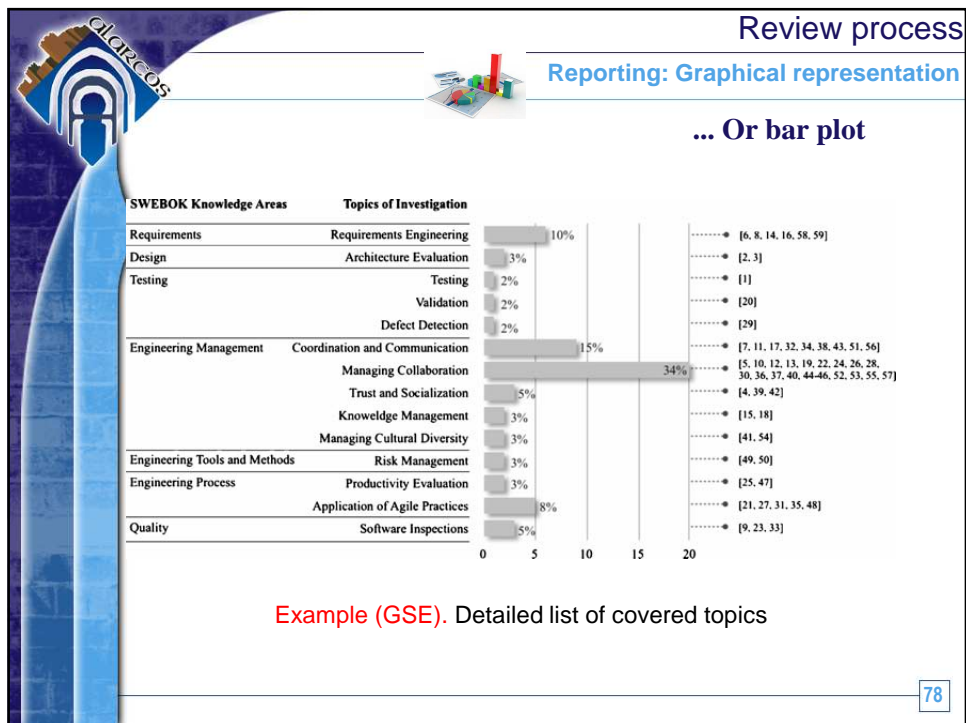
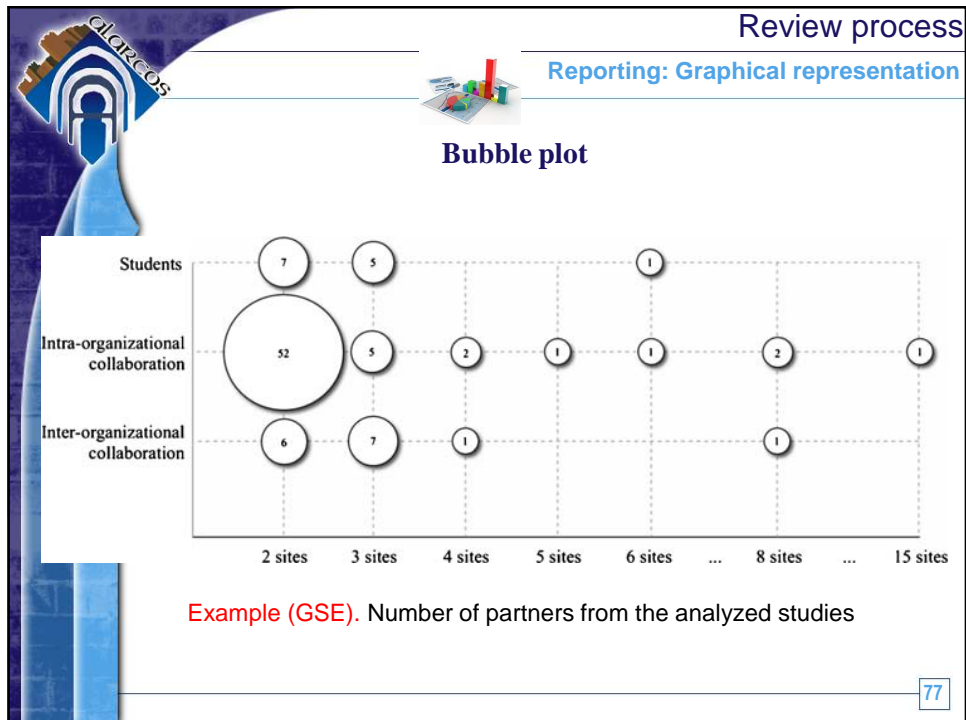
74


Results	Findings	Description of primary studies Results of any quantitative summaries Details of any meta-analysis
	Sensitivity analysis	
Discussion	Principal findings	
	Strengths and Weaknesses	Strength and weaknesses of the evidence included in the review Relation to other reviews, particularly considering any differences in quality and results.
	Meaning of findings	Direction and magnitude of effect observed in summarised studies Applicability (generalisability) of the findings
Conclusions	Recommendations	Practical implications for software development
		Unanswered questions and implications for future research
Acknowledgements*		All persons who contributed to the research but did fulfil authorship criteria
Conflict of Interest		
References and Appendices		

75




76





Review process

Reporting: Graphical representation




.. or just a table?

Type of quality	Number	Percent
Syntactic	15	5.64%
Semantic	135	50.75%
Pragmatic	103	38.72%
Syntactic + Semantic	6	2.26%
Syntactic + Pragmatic	0	0.00%
Semantic + Pragmatic	6	2.26%
Syntactic + Semantic + Pragmatic	1	0.38%
Total	266	100.00%


Example (UML). Percentage of Papers Addressing Different QualityTypes

79



Review process

Reporting




Evaluating systematic review reports

Technical reports are not usually subjected to any independent evaluation.

- If systematic reviews are made available on the Web so that results are made available quickly to researchers and practitioners, it is worth **organising a peer review**.
- If an **expert panel** were **assembled to review the study protocol**, the same panel would be appropriate to undertake peer review of the systematic review report, otherwise several researchers with expertise in the topic area and/or systematic review methodology should be approached to review the report.
- The evaluation process can use **quality checklists**.

80

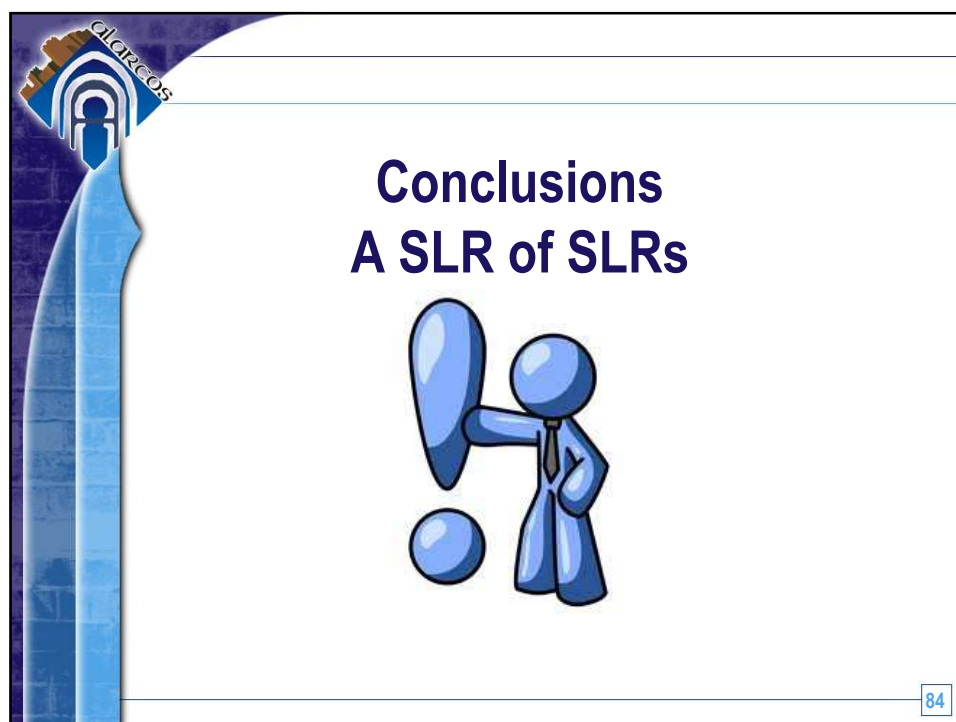
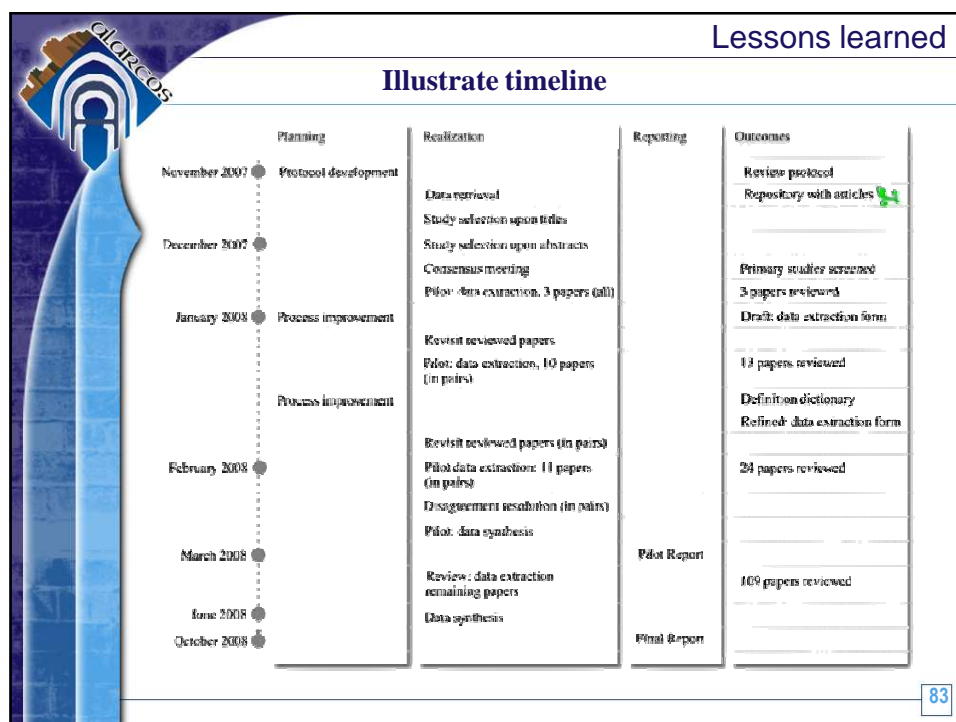





Lessons learned

- The **poor quality of search engines** available (precision, available fields)
- **Researchers** should **familiarize** themselves with how each **search engine** handles search terms.
- To avoid redundant searches, researchers should first plan which terms will be applied to which search engines and once completed, the results and timestamp are recorded.
- Due to the apparent fragility of some search engines a patient and opportunistic approach must be adopted.
- The **variable quality** of the **abstracts** available for Software Engineering papers.
- Use **structured abstracts** (context, objective, method, results, conclusion).
- More lessons learned in **Brereton et al. (2007)** and **Staples and Niazi (2007)**

82






A SLR of SLRs

- The software engineering research community is starting to adopt SLRs consistently as a research method.
 - number of SLRs is increasing.
 - number of researchers and organizations performing them is increasing.
- The integration of the results of the primary studies was poorly conducted by many SLRs.

Source (da Silva et al., 2011)

85




A SLR of SLRs

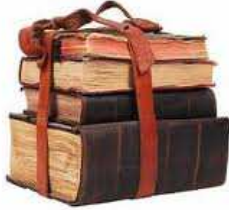
- There was very little consistency in the way the SLRs are organized.
- Many SLRs omitted essential data, including important parts of the review protocol.
- The majority of the SLRs:
 - not evaluate the quality of primary studies.
 - not provide guidelines for practitioners, thus decreasing their potential impact on software engineering practice.

Source (da Silva et al., 2011)


86



Relevant Literature




87



Relevant literature: Guidelines and lessons learned

- Kitchenham, B. and Charters, S. (2007). **Guidelines for performing Systematic Literature Reviews in Software Engineering**. Version 2.3 EBSE-2007-01.
- Brereton et al., (2007). **Lessons from applying the systematic literature review process within the software engineering domain**. Journal of Systems and Software, 80, 571-583.
- Staples, M., Niazi, M. (2007). **Experiences using systematic review guidelines**. The Journal of Systems and Software, 80(9), 1425–1437


88



Relevant Literature: Tertiary studies

- Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey, J., Linkman, S. (2009). **Systematic literature reviews in software engineering – A systematic literature review**. Information and Software Technology 51 7–15.
- Kitchenham, B. e tal.(2010). **Literature reviews in software engineering – a tertiary study**. Information and Software Technology 52 (8) 792–805.
- Fabio Q.B. da Silva, André L.M. Santos, Sérgio Soares, A. César C. França, Cleviton V.F. (2011). **Six Years of Systematic Literature Reviews in Software Engineering: An Updated Tertiary Study**. Information and Software Technology 53(9) 899-913.
- Zhang, H., Ali Babar, M. (2013). **Systematic reviews in software engineering: An empirical investigation**. Information and Software Technology, 55(7),1341–1354


89



Relevant literature: Synthesis

- Daniela Cruzes, Tore Dybå. **Research synthesis in software engineering: A tertiary study**. Information & Software Technology 53(5): 440-455 (2011)
- Daniela S. Cruzes, Tore Dybå. **Recommended Steps for Thematic Synthesis in Software Engineering**. ESEM 2011: 275-284
- Daniela S. Cruzes, Tore Dybå, Per Runeson, Martin Höst: **Case Studies Synthesis: Brief Experience and Challenges for the Future**. ESEM 2011: 343-346

90




Relevant literature

Example: Systematic literatures reviews

- Tore Dybå, Torgeir Dingsøy. **Empirical studies of agile software development: A systematic review.** Information & Software Technology 50(9-10): 833-859 (2008).
- Sarah Beecham, Nathan Baddoo, Tracy Hall, Hugh Robinson, Helen Sharp. **Motivation in Software Engineering: A systematic literature review.** Information & Software Technology 50(9-10): 860-878 (2008).
- Tracy Hall, Sarah Beecham, David Bowes, David Gray, Steve Counsell. **A Systematic Literature Review on Fault Prediction Performance in Software Engineering.** IEEE Trans. Software Eng. 38(6): 1276-1304 (2012)

91



Relevant literature

Example: Systematic mapping studies

- Adrian Fernandez, Emilio Insfrán, Silvia Abrahão. **Usability evaluation methods for the web: A systematic mapping study.** Information & Software Technology 53(8): 789-817 (2011).
- Fabio Q. B. da Silva, Marcos Suassuna, A. César C. França, Alicia M. Grubb & Tatiana B. Gouveia & Cleviton V. F. Monteiro, Igor Ebrahim dos Santos. **Replication of empirical studies in software engineering research: a systematic mapping study.** Empirical Software Engineering. (2013).
- Ana M. Fernández-Sáez, Marcela Genero, Michel R. V. Chaudron. **Empirical studies concerning the maintenance of UML diagrams and their use in the maintenance of code: A systematic mapping study.** Information & Software Technology 55(7): 1119-1142 (2013)

92



How to Perform Systematic Reviews: Theory and Examples



Marcela Genero Bocco
ALARCOS Research group

Department of Technologies and Information Systems
University of Castilla-La Mancha
Marcela.Genero@uclm.es