RESEARCH METHODOLOGY IN SE
SYSTEMATIC LITERATURE REVIEW

AIUB. FALL 2019-2020

Dr. Mahbubul Syeed
Associate Professor and Head, Department. of CS, AIUB
mahbubul.syeed@aiub.edu
www.msyeed.weebly.com

HOW WE DO THE SYSTEMATIC LITERATURE REVIEW

AN EXAMPLE STUDY

THE REVIEW PROCESS

Three main phases:

- ✓ Planning the Review
- ✓ Conducting the Review
- ✓ Reporting the Review.

THE REVIEW PROCESS

Three main phases:

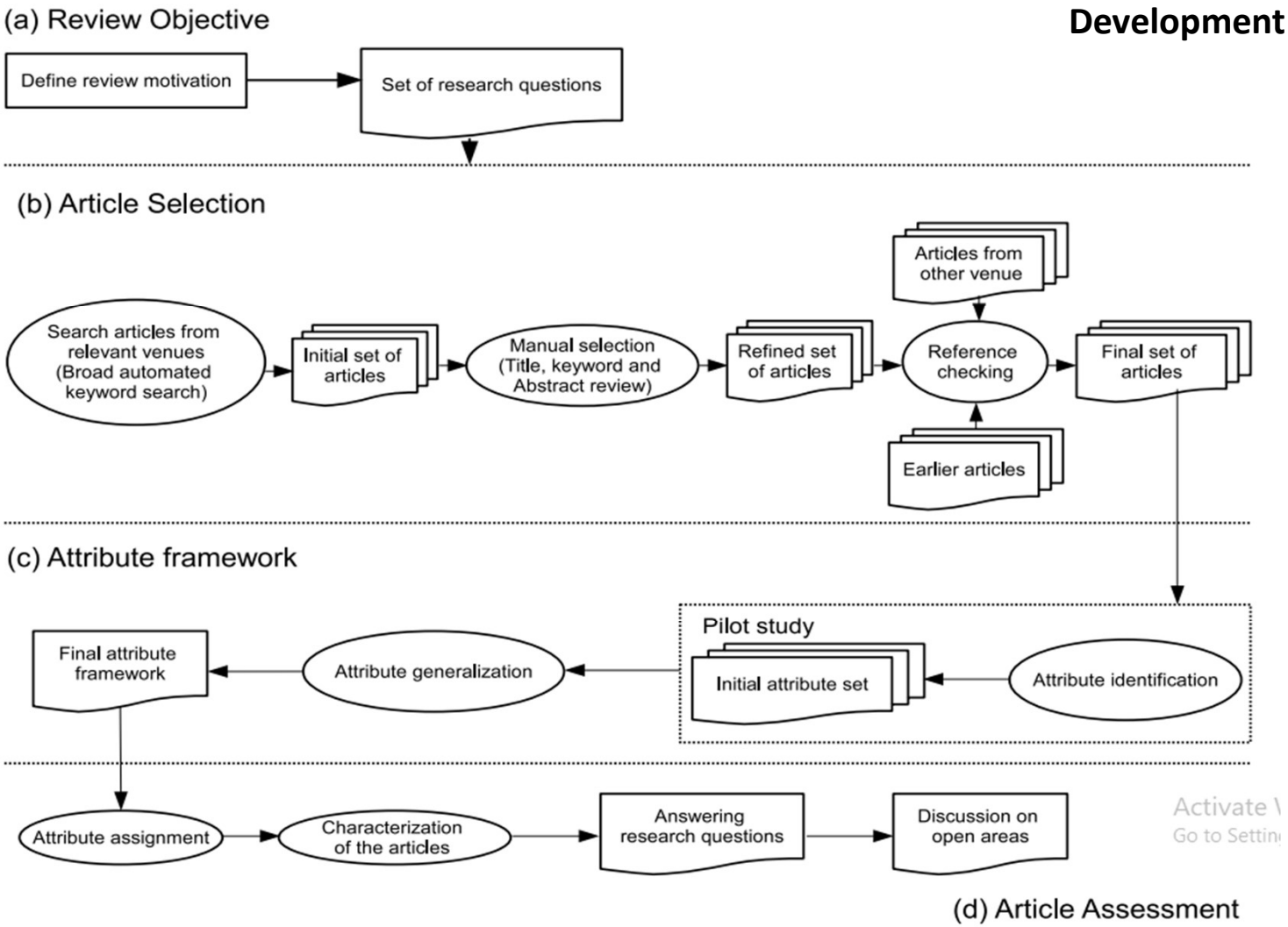
- ✓ Planning the Review
- ✓ Conducting the Review
- ✓ Reporting the Review.

Planning the review are:

1. Identification of the need for a review

2. Development of a review protocol.

Development of a review protocol



Title: Evolution of Open Source Software Projects: A Systematic Literature Review

Identification of need:

This paper reports on a systematic literature survey aimed at the identification and structuring of **research on evolution of OSS projects**. In this review we systematically selected and **reviewed 101 articles** published in relevant venues. The study outcome provides insight in what constitutes the main **contributions of the field, identifies gaps and opportunities**, and **distills several important future research directions**.

JOURNAL OF SOFTWARE, VOL. 8, NO. 11, NOVEMBER 2013

Research Questions:
The research questions we have defined fall within the context of **OSS projects and their evolution strategies**.

Total of 11 Questions are investigated.

Research Questions:

Category	Research Question	Motivation
Target	Which facets of OSS projects were explored and what statistical distribution the articles have in those facets?	To decompose the articles according to their study focus and intensity of studies in each focus area
	What are the dimensions of OSS projects explored under each study facet?	To determine the specific aspect(s) of OSS projects explored in evolution studies within each facet.
Approach	What are the research approaches followed in the studies?	To identify the general research approach followed in evolution studies (empirical studies with quantitative or qualitative data analysis).
	What are the datasets or data sources of OSS projects mostly exploited in evolution studies?	To identify the data sources of an OSS project that are used for studies.
	What metric suits are evaluated and what tools are used for metric data collection?	To explore the metric suits used for evolution study and the popularly used tools for data extraction.

Research Questions:

Identification of the need for a review

Category	Research Questions	Motivation
Target Group	What is the portfolio of projects analyzed for evolution studies and what are their domains?	To determine the mode of evolution studies (e.g., horizontal or vertical) by statistically measuring the studied OSS projects and their domains .
Outcome	Does the concern on "OSS evolution study" follow an increasing trend?	To determine the mode of evolution studies (e.g., horizontal or vertical) by statistically measuring the studied OSS projects and their domains
	What contributions are made in literature to analyze the evolution of software?	To explore what results are presented to enhance the understanding of OSS projects evolution. (e.g., do evolution of OSS projects conforms to the theory of software evolution?)
	What contributions are made in literature to analyze the evolution of organization or community?	
	What contributions are made in literature to analyze the interdependency in the evolution of the software and organization?	
	How are the research approaches and results of the articles typically validated?	To identify the approaches employed to evaluate the research approaches and study results (e.g., internal validity, external validity, construct validity)

THE REVIEW PROCESS

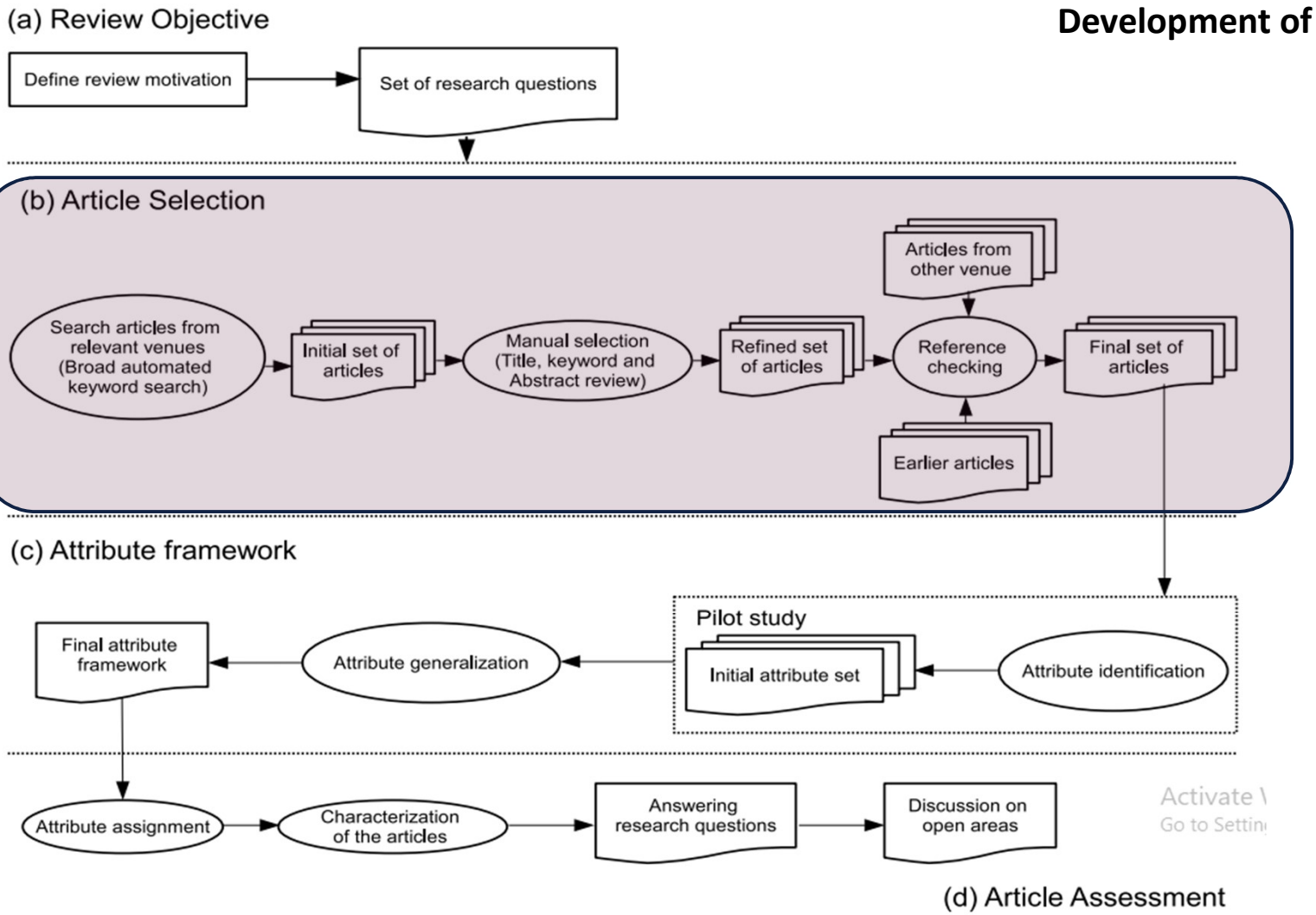
Three main phases:

- ✓ Planning the Review
- ✓ Conducting the Review
- ✓ Reporting the Review.

Conducting the review are:

1. Identification of research
2. Selection of primary studies
3. Study quality assessment
4. Data extraction & monitoring
5. Data synthesis.

Development of a review protocol



Inclusion criteria.

(a) Subject area of the articles must unveil strong focus on evolution of OSS projects. Authors must explicitly state the target of the study (e.g., **software evolution, community evolution, co-evolution, prediction**) and provide detail evidence of **research methodology, data sets, and statistical detail** of case study projects.

(b) Articles must exhibit a **profound relation to OSS projects** and take into consideration those aspects that are particularly attributed to the OSS community and projects. Articles using OSS as a case study are taken into account only if they satisfy the above criterion.

(c) Articles published in **referred journals and conferences** are included for the review. Similar to most SLRs, **books are not considered** for the review.

The suitability of the articles was determined against the above mentioned selection criteria through a manual analysis (discussed later in this section) of title, keywords, abstract. In case of doubt conclusions are checked.

Automated Keyword Search.

Automated keyword search. Automatic keyword search is a widely used strategy in literature surveys.

Thus we performed a **broad automated keyword search** to get the initial set of articles.

First author of this article was responsible for the search process.

Seven digital libraries were searched:

IEEE Computer Society Digital Library;

ACM;

ScienceDirect;

SpringerLink;

Google Scholar;

FLOSShub and

Mendeley.

Automated Keyword Search.

- ✓ All searches were based on the **title, keywords and abstract**.
- ✓ The time period for this search was from **January, 2000 to January, 2013**.
- ✓ Search strings varies among libraries, we first defined **search terms according to our inclusion criteria**.
- ✓ The list of search terms that were used is as follows.
  - ✓ Terms representing OSS: "Open source" or OSS or "Open Source Software" or "Open Source Software projects" or FLOSS or "Libre Software" or "F/OSS".
  - ✓ Terms representing evolution study: "evolution" or "structural evolution" or "evolution of software" or "project evolution" or "project history" or "software evolution" or "community evolution" or "co-evolution".

Automated keyword search ended up with ***181 articles consisting of 46 journal articles and 135 conference articles***.

Manual Selection.

Recent studies [15] [9] pointed out that (a) current digital libraries on software engineering do not provide good support for automated keyword search due to lack of consistent set of keywords, and (b) the abstracts of software engineering articles are relatively poor in comparison to other disciplines.

- ✓ Thus it is possible that the 181 articles identified through automated search process might contain **irrelevant ones and some relevant might be missing**.
- ✓ **First author** performed a **manual selection** on these articles by reviewing the title, keywords and abstract (and in case of doubt, checking the conclusion [15]).
- ✓ To reduce the **researcher bias** in this selection process, the **domain experts** (second and third author) examined the selected articles against the selection criterion.
- ✓ Any **disagreement was resolved through discussion**.

This process ended up with 97 articles consisting of 21 journal articles and 76 conference articles.

Reference checking.

To ensure the inclusion of other relevant but missing articles (as mentioned above), the first author performed a **non-recursive search through the references** of the 97 selected articles. This process identified **4 additional conference articles**.

Final set of articles.

The article selection process finally ended up with **101 articles (21 journal and 80 conference articles).**

A complete list of these articles along with year and venue wise distribution can be found in our review website [14].