Project Summary

Keywords: Systematic Literature Reviews, Software Engineering, Web-based Infrastructure

The Systematic Literature Review (SLR) process is a means for analyzing published evidence to draw conclusions about about a phenomenon of interest. The prevalence of empirical research in software engineering (SE) makes it a domain well-suited for the application of the SLR process. While the popularity of SLRs has been increasing within the SE community in recent years, there are still significant barriers to more widespread adoption. In particular:

- The process of systematically identifying relevant papers is a labor-intensive, largely manual process. Current SE literature databases lack adequate tools to assist in the extraction and analysis of a large body of literature. A researcher must manually manipulate search strings and parse through large numbers of search results, many of which are irrelevant
- There is little support for collaboration among multiple researchers. SLRs require the interaction of multiple researchers to review papers, extract data, and analyze the results. There is no infrastructure to support this critical task
- There is no mechanism to store the results of the SLR process to allow for reuse. Each paper analyzed during an SLR results in a completed data extraction sheet. There is currently no infrastructure to allow some or all of this data to be reused in subsequent SLRs
- There is no mechanism to allow SLRs to evolve over time. As new research is published, there needs to be a way for SLRs to evolve and stay current

These barriers remain due to the lack of infrastructure support for the labor-intensive, largely manual process of performing an SLR. The primary goal of this CI-P project is to **create an infrastructure to support the SLR process**. The community targeted by this project comprises SE researchers who already conduct SLRs, as well as SE researchers who will be more likely to conduct SLRs if the barriers to adoption are removed. The primary objectives of this CI-P project build on the investigators' preliminary work and are to:

- Evolve the community needs and proposed infrastructure identified in preliminary work
- Create a detailed plan for the development and deployment of the infrastructure

The ***intellectual merit*** of this project is in the fact that literature reviews, whether systematic or *ad hoc* are an integral part of the research process. By enabling the execution of *systematic* literature reviews, this infrastructure will help to ensure that researchers identify a complete and unbiased set of candidate papers when performing such a review.

The ***broader impacts*** of this project stem from lowering the barrier to performing SLRs. By removing the barriers currently faced by SE researchers, the proposed infrastructure will allow a larger portion of the SE community to participate in the conduct of SLRs. The infrastructure will be especially helpful to PhD students who will conduct a literature review as part of their thesis development. Because SLRs are often publishable as a stand-alone article, the infrastructure will also help increase the publication rate for these students.

Part I Computing Research Infrastructure Concept

Systematic Literature Reviews (SLR) are useful for drawing conclusions about about a phenomenon, based upon the current state of published evidence. Although SLRs are increasingly recognized as vital to the software engineering (SE) discipline, there are still barriers inhibiting widespread adoption of SLRs within the SE community. These barriers remain due to the lack of infrastructure support for the effort-intensive, largely manual SLR process. Given the frequency of empirical studies in SE, the domain is well-suited for SLRs. The primary goal of this CI-P project is to propose an infrastructure to support the SLR process. Our target community ("the community") comprises SE researchers who already conduct SLRs, as well as SE researchers who will be more likely to conduct SLRs if the barriers are removed. Building on our preliminary work, the primary objectives of this CI-P project are to:

- Elicit, validate, and prioritize the requirements for an SLR support infrastructure
- Propose a new SLR support infrastructure based on those requirements
- Create a detailed plan for the development and deployment of the infrastructure

After completion of these planning activities, we estimate that it will take approximately three years and $500,000 to develop and deploy the infrastructure. Once deployed, it should require minimal maintenance, which will consist primarily of ensuring compatibility with software upgrades. Therefore, the ongoing operation costs should be easily managed with internal resources. We anticipate the infrastructure to have an indefinite lifetime because review of existing literature is, and will continue to be, an integral part of any new research endeavor.

1 Introduction

The review of existing literature is the foundation of all new research. In SE, these reviews have traditionally been performed in an *ad hoc* manner. To provide some structure to the literature review process, medical researchers defined the SLR process. While SLRs have been commonplace in medicine, they are a recent phenomenon in SE largely due to significant barriers that inhibit their wider adoption, e.g., the process is labor intensive and tool support is lacking. The goal of the proposed infrastructure is to lower the barrier for researchers who want to perform SLRs. The proposed infrastructure will also provide researchers with a platform for community interaction and data dissemination. We expect that the infrastructure will increase not only the quantity, but also, more importantly, the quality of SLRs and SE research in general.

The initial step of a new research endeavor is a review of previous work to properly ground the new research. The most common method of examining previous work is via literature review. A literature review can also have other goals, such as, summarizing the current state of knowledge about an area as a service to the community. Regardless of the goal of the literature review, a researcher can perform the review in the traditional, unsystematic, fashion, that is, by conducting database searches and following references or the researcher can use a more systematic method.

An SLR is a formal, repeatable method by which a researcher can identify, evaluate, and interpret the available research about a question or topic area. The primary difference between an SLR and and *ad hoc* review is the level of advanced planning in an SLR. Prior to conducting the review, the researchers develop a protocol that documents: the research question(s) to guide the review, the search strategy (including specific databases and keywords), the criteria for choosing appropriate papers, a quality assessment method for the papers, the specific information to be

extracted from each paper, and a plan for synthesizing the information from the set of papers to draw a conclusion. By using this systematic process, researchers are much less likely to accidentally omit important papers from the literature review.

Medical researchers, practitioners, and policy makers have long relied on SLRs, because they integrate up-to-date, reliable, and critical information that support important decisions. Seeking these same benefits, the SE community has recently begun publishing SLRs. Indeed, with the growing emphasis on empirical research in SE, SLRs are of critical importance because they allow researchers to bring together disparate evidence to understand the effects of various SE tools, techniques and methods. Unfortunately, though production and dissemination of SLRs is key to the maturation of SE and to the adoption of SE research practices by industry, conducting an SLR is a difficult and time-consuming process. Based on our own SLR experiences, a review of over 200 SLRs, and a survey of over 50 SLR authors, we have identified four key barriers to wider adoption. Section 2 discusses each of these barriers in more detail.

First, the process of systematically identifying relevant papers is largely manual and thus very labor intensive. This process is more difficult when research topics cross traditional disciplinary boundaries, as many interesting topics increasingly do. While some advancement has been made in database functionality, as a discipline, we still lack adequate tools to assist in the extraction and compilation of relevant information from existing research. Using most common search engines for SE literature (i.e., IEEExplore, ACM, or Google Scholar), a search may yield thousands of results, of which a large percentage may be irrelevant for various reasons (i.e., overloaded terminology or simply contain the right word combinations by chance). In addition, because databases are not mutually-exclusive, the result set will likely contain duplicates, which must be manually parsed by the researchers.

Second, there is little tool support for collaborative SLRs. After identifying the relevant set of articles, multiple researchers must extract important information from each paper and compare the results for consistency. Again, this step is typically performed manually. There is no tool support to ease this process and to aid in the inter-rater reliability assessment necessary to evaluate the accuracy of the extracted information.

Third, there is no mechanism to store the data extracted from papers so that it can be updated and reused. It is quite likely that the same paper may be relevant to multiple SLRs. While the data extracted from a paper may differ somewhat depending on the research questions, there will likely be a lot of common data items. Because there is no central repository for storage of extracted data, researchers must fully repeat this extraction process for each new SLR. Such a repository would not only reduce effort by enabling a researcher to extract only the additional data relevant to the new research question(s), but also facilitate collaboration by allowing researchers to identify others working on similar topics.

Finally, there is no mechanism to enable SLRs to evolve over time. Ideally, an SLR should be a "living" document that could evolve as new research results become available. Because current publications are static, appropriate infrastructure is needed to support the evolution of SLR results by allowing researchers to easily create new versions or fork off related topics. Making SLRs living documents that incorporate the latest research results will allow them to be more useful both to researchers and practitioners.

To address these barriers and to provide a platform to encourage more researchers to conduct SLRs, there is a need for new infrastructure. The goal of this planning grant is to build upon our initial experiences to develop a complete picture of the community needs for this infrastructure. By interacting with the community SLR authors, we will clarify the set of problems for which infrastructure is particularly needed. In addition, we will also interact with the community to ensure that by the end of the planning grant we have a concrete proposal for the implementation of the infrastructure that the community will find useful and beneficial.

Intellectual Merit Literature reviews are an integral part of any new research activity. As such, it is critical that researchers are able to identify as complete a set of related literature as possible. This project will produce a plan to develop and deploy such infrastructure. Specifically, the resulting infrastructure will facilitate current and future research in that it will:

- Enable researchers to perform SLRs more easily by reducing the manual effort required and by improving the accuracy of the result
- Serve as a repository of all related literature about a research topic that can be kept current through the addition of newly published articles
- Foster additional research by providing a repository of peer-reviewed data extraction sheets to the research community for exploration and use in meta-analysis
- Serve as a community hub to facilitate geographically distributed collaborations and to enable social networking regarding the results of SLRs

Broader Impacts By lowering the barriers to performing SLRs and enabling more SLRs to be produced, the proposed infrastructure will:

- Enable a larger portion of the SE community, especially new researchers, to conduct SLRs
- Increase the prevalence of summarized results that can inform research and practice
- Make the results of a review to be accessible to a larger audience

In addition, because SLRs are often publishable in their own right, and because all PhD students must perform some type of literature review as part of their work, the proposed infrastructure will enable PhD students to obtain an additional publication in the course of their work.

The proposal is organized as follows. The remainder of Part I provides background on SLRs and existing tools. Part II describes our preliminary work to identify community needs with regards to infrastructure requirements (Section 3) and proposed solutions (Section 4). Part III details the work we plan to conduct as part of this planning project to validate the information in Part II. Part IV lays out the plan for completing the proposed work. Finally Part V describes the qualifications of the investigators.

2 Systematic Literature Review Process

Kitchenham ported the SLR process from the medical field to SE. The process, as prescribed by Kitchenham [**?**, **?**], consists of three primary phases: review planning, review execution, and review documentation.

During the **planning phase**, the researcher defines a protocol that guides SLR execution. The goal of the protocol is to reduce researcher bias and provide a repeatable, transparent process for conducting the SLR. The protocol should contain, at a minimum, the following information:

P1. Motivation for conducting the SLR
P2. Research question(s)
P3. Search strategy - including databases to be searched and search string(s)
P4. Strategy for identifying primary studies (i.e. inclusion and exclusion criteria)
P5. Quality assessment criteria
P6. Data extraction form

An independent expert or panel reviews the protocol for completeness and validity. If at any point during the SLR execution the researchers must change the protocol, the expert or panel should re-review the revised protocol.

During the **execution phase**, researchers proceed through five steps:

E1. Identify relevant research by executing the defined search strategy
E2. Select primary studies by applying the inclusion and exclusion criteria
E3. Assess study quality using the quality assessment criteria
E4. Extract required data into data extraction forms
E5. Synthesize data to draw conclusions

Researchers apply the search terms to multiple databases. Then the researcher(s) use the inclusion and exclusion criteria to reduce the results of the search process, using titles first, then abstracts, then full text. During each iteration, the researcher(s) eliminate prospective studies only when it is clear that the study is not relevant. After selecting the primary studies, the researchers perform a quality assessment of each selected study. The studies are then weighted based upon the results of the quality assessment. Finally, the researchers extract important data from all included studies. The resulting data set then forms the basis for the data synthesis. To reduce researcher bias in the process, members of the research team perform each step independently and then meet to review the results and resolve any conflicts.

Lastly, during the **documentation phase** the researchers use all of the information described in the protocol, along with the results of the execution of the protocol to document the review in some type of publication. This phase consists of the following steps:

D1. Specify dissemination strategy
D2. Format SLR report

2.1 Existing tools

The SLR process originated with the The Cochrane Collaboration, the primary organization that organizes and disseminates SLRs in medicine. With the long history of SLRs, we expected to find some sophisticated suport tools. We did locate one toolset, the RevMan/Archie combination utilized by the Cochrane Collaboration [**?**]. RevMan focuses primary on the documentation phase of SLRs performed under the guidelines of the Cochrane Collaboration. RevMan includes facilities for the preparation of formatted tables used in the reviews and for tracking the revisions of a review over time. Archie is the central repository, or backend for the RevMan application and is used to store completed reviews. Neither tool appears to provide much functionality to support the planning or execution phases of the SLR process.

With the increasing prevalence of SLRs in SE, we were surprised to find only two tools that support the SLR process within SE. First, StArt [**?**] is a desktop based application design to assist

an individual researcher in designing and conducting an SLR that follows the process originally described by Kitchenham [**?**]. To use StArt, a researcher inputs protocol elements, including: research questions, databases, and study selection criteria. These elements then become attributes of the steps in the execution phase of the SLR. StArt also provides some assistance for "snowballing," the process in which researchers trace the citations of a paper forward and backwards, by comparing identified studies with the references of the studies. StArt also includes visualizations that assist in the development of the search strings and the documentation of the review.

Second, SLuRp [**?**] is a web based tool that functions as the central repository for an SLR research team. The repository provides a common storage location for the SLR protocol, PDF copies of the studies under review, and associated reference information. Additionally, SLuRp provides facilities to manage the assignment of studies for review to the different researchers on the team and to collect the results of these reviews.

Part II Preliminary Work

We have performed initial work to gather community needs and identify necessary characteristics and features for the infrastructure. Section 3 describes the community needs we have identified through our own experiences and our interactions with the community. Section 4 describes our current view of the infrastructure required to address the identified community needs.

3 Identification of Community Needs

As a starting point for planning the infrastructure, we have identified a set of important community needs, which will be iterated upon during the course of the project. The important community needs we identified include, improved support for:

1. Coordination among multiple (possibly distributed) researchers
2. Protocol review by the community
3. Interaction with different literature databases
   (a) Federated search
   (b) Search string manipulation and translation
   (c) Duplicate result elimination
   (d) External citation management tool interoperation
4. Quality assessment
5. Data extraction

Section 3.1 describes the three data sources used to gather these needs. Section 3.2 provides a detailed analysis of the data sources to illustrate the origing of these community needs.

3.1 Overview of Data Sources

To gather community needs, we used three data sources: a graduate course by PI Carver (Section 3.1.1), a review of published SLRs (Section 3.1.2), and a survey of SLR authors (Section 3.1.3).

3.1.1 PI Carver's SLR Course

In Spring 2012, PI Carver taught a graduate "Advanced Empirical Software Engineering" course. There were eight PhD students enrolled in the course (four Computer Science PhD students and

four Management Information Systems PhD students). The main focus of this course was for the students to learn about and conduct SLRs. As such, the course had two primary goals:

1. Each student conducted an SLR as a semester-long project. Realizing that most SLRs cannot be completed within one semester, it was expected that work would continue beyond the semester to make these SLRs publishable. At this point, two of these papers have been accepted in conferences [**?**,**?**], at least five other papers will be submitted to various journals in the near future, including *European Journal of Information Systems* and *Information and Software Technology*, and most of them will become part of the students' dissertations.
2. Second, throughout the semester, class time was devoted to discussing each student's SLR and to evaluating the SLR process. Each student wrote a report at the end of the semester describing their SLR process, noting where they had difficulties, and suggesting improvements to the SLR process.

In addition, each student acted as a second reader for the SLR of one of their classmates. PI Carver oversaw all SLRs, provided input on the protocols and helped to resolve any conflicts during the paper selection/data extraction process. Due to the effort required to conduct these reviews, it was not possible for every paper to be reviewed by two reviewers. So, at each elimination stage of the SLR process (i.e. title elimination, abstract elimination and full paper elimination), the second reader reviewed a random subset of 10-20% of the excluded papers to ensure that the primary author was not prematurely excluding papers that could be relevant. When the primary author reached the data extraction stage, the second reader reviewed the full data extraction for a subset of the papers.

In addition to the interaction among the primary author and the second reader throughout the semester, a large portion of the class meeting time was devoted to discussing each SLR. Each student had the opportunity to present his protocol and received feedback from his classmates. The class also spent a substantial amount of time discussing the logistics of the SLR process and identifying common problems encountered by multiple students. By the end of the course each student produced two deliverables:

1. An initial version of the SLR, which in most cases needed further revision to become publication-ready; and
2. A report describing his experiences with following the standard SLR process, including specific areas in which he encountered difficulties. It is these reports that provided the bulk of the information described Section 3.2.

3.1.2 SLR Literature Review

For the second source of data, we performed a thorough literature search and identified 214 published SLRs. We analyzed those SLR papers to identify strengths and weaknesses in the current SLR process, including areas in which tool support could be helpful. During the review of these papers, we used the following data extraction process:

1. Determine whether the SLR protocol was explicitly described
2. Review the description of the protocol to:
   (a) Note databases used and the search strategy for those databases
   (b) Note any deviations from Kitchenham's process (whether stated explicitly or not)
   (c) Note any tools used

(d) Note any difficulties described about the process
3. Look for a "lessons learned" section to get more details
4. Note anything else particularly interesting about the SLR process in the paper

The information extracted from this analysis is included in Section 3.2, and the list of 214 published SLRs is included among the cited references at the end of this proposal.

3.1.3 Survey of SLR Authors

Finally, to obtain more detailed insight into the SLR process, we developed a list of SLR authors from the list of published SLRs identified in Section 3.1.2. We sent a survey to all of those authors that asked them to detail: the SLR process followed, where they had difficulties in the SLR process, where they spent their time during the SLR process, and which aspects of the SLR process were most in need of tool support. The results of the 50+ survey responses are included in the discussion in Section 3.2.

3.2 Analysis of SLR Process

This section uses data gathered from the three data sources described in Section 3.1 to analyze the steps in the SLR process. The goal of the analysis is to identify aspects of the SLR process that SLR authors found particularly difficult, highlighting community needs for tool support of the SLR process. This section is organized around the major phases of the SLR process: Section 3.2.1 discusses general issues regarding Protocol Planning and Section 3.2.2 focuses on specific protocol items. For the sake of continuity, we discuss data from all three data sources together under each heading. As we discuss each phase of the SLR process, we specify the identified community needs.

3.2.1 Protocol Planning

This section describes general information pertaining to protocol development. Information about detailed protocol items appears in Section 3.2.2. During PI Carver's course, the students' stated that the frequent discussion of their protocols with each other and with PI Carver helped in the development and refinement of the protocols. The students also found it helpful for their SLR process to be guided by PI Carver, who is experienced in conducting SLRs. These experiences are not unique to PI Carver's course. Our review of the SLR literature suggests that many SLRs are performed by students as part of their thesis work. Guidance from a more experienced researcher/reviewer is crucial to the accuracy and success of the review [**?**]. The community needs identified in this phase are:

- **Protocols may have to be revised and edited during the SLR process**
- **Protocols need to be socialized among peers for review and feedback**
- **Novice researchers need the ability to interact with a more experienced research during the planning and execution of the SLR**

3.2.2 Protocol Items

This subsection describes the results obtained for protocol items P2-P6.

P2. Research Questions The research question(s) are arguably the most important aspect of the protocol because they drive the remainder of the protocol. Based on the experiences in PI

Carver's course, it is clear that the research question(s) have to be appropriately scoped. If they are too broad, they will generate too many papers to reasonably evaluate in one SLR. Conversely, if they are too narrow, they will not generate enough papers to draw useful conclusions. Scoping of the research question(s) is an activity in which feedback from more experienced researchers could be quite beneficial. The community need identified in this section is:

- **Research question(s) must be properly scoped**
- **Expert feedback is beneficial to formulating appropriate research question(s)**

P3. Search Strategy The search strategy includes database selection and creation of one or more search string(s) from the research question(s). Creating the search string(s) can be an iterative process as the researcher attempts to define the appropriate set of keywords and synonyms that cover the research space. This section first discusses issues with the databases and then issues with the search strings.

Regarding databases, SLR researchers typically search multiple databases, which may have different functionality. Based on the experiences in PI Carver's course, we can make the following observations regarding the databases. First, databases have different behavior regarding the search strings. For example, in some cases changing the order of the keywords changes the result set. In other cases, Boolean logic does not work as expected. As a result, researchers must develop a different set of search strings for each database. Second, the advanced search functionality differs across databases. In some cases, the advanced search interface returns different results than the basic search interface, even if the same search string is used. Third, there is a large overlap in the literature covered by the databases commonly used for SLRs, creating the need to identify and remove duplicate studies from the result set. Fourth, databases differ in the content and format of the citation information provided. Finally, the databases are not consistent in their behavior regarding bulk export of references to a citation manager.

Similar problems with the peculiarities of the databases were also reported by the papers in our literature review (e.g., [**?**,**?**,**?**,**?**,**?**,**?**,**?**,**?**]). A relatively small number of researchers used EndNote to facilitate the removal of duplicate papers (e.g., [**?**]) while most performed the task manually.

Regarding the search strings in general (i.e. not including differences among databases), the experiences in PI Carver's course resulted in the following observations. First, adding a '*' at the end of a key term helps to identify variant spellings. Second, when using a common term like "Open Source," restrict the search to the title, abstract and keywords to limit the number of irrelevant hits. Third, in some cases a large number of relevant papers were not returned by the initial search, causing the search string to be refined based upon the results of a secondary search, i.e. reviewing references in the identified papers.

This information suggests the following community needs:

- **The ability to search multiple databases in one tool**
- **Automatic (or semi-automatic) manipulation of search string to accommodate peculiarities of various databases**
- **Automatic elimination of duplicate results**
- **The ability to export all citations in a common format (i.e. BibTex or EndNote)**

P4. Identification of Primary Studies Once the search results are identified, the next step is to determine which papers should be included in the SLR. There are two aspects to this process: the definition of the inclusion/exclusion criteria and the process of actually determining which papers belong in the review. The SLR author survey indicated that selecting appropriate papers was the third most difficult and second most time consuming aspect of the SLR process. It was also the aspect second most in need of tool support.

PI Carver's course resulted in the following observations about the inclusion/exclusion criteria: 1) there is a need to ensure that it conforms to the goals of the current review rather than just being reused from other SLRs; 2) there is a need for expert review; 3) it should be as specific as possible; and 4) it may have to be reviewed and edited during the search process.

PI Carver's course also resulted in the following observations about the paper selection process: 1) review of the titles and abstracts may not be sufficient for eliminating papers; 2) there is a need for a citation manager to keep track of the references; 3) have to manually examine or write a script to handle duplicate results across databases; 4) when searching for specific terms or content that may not be evident in the title/abstract/keywords, the addition of a quick scan of the full text of the paper may be useful for quickly eliminating irrelevant papers; 5) this step was very time consuming, especially having multiple reviews of papers; and 6) it is difficult to coordinate meeting times.

The literature review showed that other authors reported the same or similar problems. For example, some researchers noted the difficulty of duplicate removal [**?**]. Researchers used various tools for managing the papers: a "citation manager," [**?**] or EndNote [**?**, **?**, **?**].

This information suggests the following community needs:

- **There is a need for expert review of inclusion/exclusion criteria**
- **The citations need to be managed within the tool**
- **The tool needs to interoperate with external citation management tools (e.g., BibTeX, EndNote)**
- **There is a need to automate the elimination of duplicate papers**
- **There is a need for management of the review process among the SLR team**

P5. Quality Assessment Once a set of candidate primary studies has been identified, the next step is to perform a quality assessment to determine whether any should be excluded due to the unreliability of their results. The results of the SLR author survey showed that quality assessment was the second most difficult and third most time consuming aspect of the SLR process.

The experiences from PI Carver's course resulted in the following observations. First, the quality assessment checklist should be part of the data extraction form. Second, the quality criteria must be relevant to the specific topic and not just reused from a prior SLR. Third, the researcher must ensure that the quality assessment criteria will actually differentiate between papers of different quality. Finally, the quality assessment should be performed by at least two researchers to avoid bias. These observations suggest the following community needs:

- **There is a need for a mechanism to build appropriate quality assessment criteria**
- **There is a need to facilitate multiple authors performing quality assessments independently**

P6. Data Extraction Once the authors have arrived at a final set of papers that are to be included in the SLR, the next step data extraction. The results of the SLR author survey showed that extracting data from papers was the most difficult and most time consuming aspect of the SLR process. It was also the aspect that was the third most in need of tool support.

The experiences in PI Carver's course resulted in the following observations about the data extraction process. First, the data extraction form needs to be reviewed by an expert in SLRs and an expert in the domain of the review. Second, realize that the form may need to be refined throughout the process as the authors better understand the papers. Third, the extracted information needs to be reviewed by collaborators to eliminate any bias. Finally, there is a need for a tool that allows the extracted data to be easily recorded and analyzed across multiple papers. In our literature review, most researchers did not report the use of a tool for data extraction. We did find a few researchers using EndNote (e.g., [**?**,**?**,**?**]) or a spreadsheet (e.g., [**?**,**?**,**?**,**?**]) to manage the data extraction process. These observations suggest the following community needs:

- **There is a need to facilitate expert review of the data extraction form**
- **There is a need to support multiple authors extracting and reviewing data**
- **There is a need to ease the recording and analysis of the extracted data**

3.3 Existing SLR tools

Section 2.1 described three tools that support various aspects of the SLR process. While RevMan / Archie does include desirable features, it is primarily designed for documenting and maintaining individual medical studies under the auspices of the Cochrane Collaboration. StaRt and SLuRp have a similar focus to our proposed infrastructure. This section briefly analyzes the StArt and SLuRp tools with regards to their match to the identified community needs.

StArt enforces a rigid interpretation of the SLR process. It requires the user to enter elements of the protocol, which are later used to mandate additional inputs before the user may proceed with conducting the review. While true to the original definition of the SLR protocol, this tool lacks the ability to support the iterative nature of the SLR process. As a result, evolution of the protocol during review process is cumbersome. In addition, StArt is designed to be a desktop application that supports a single user, thus, neglecting the collaborative nature of the SLR process.

SLuRp is web-based and allows multiple researchers to collaborate on the same SLR. However, it does not aid in the development of the search string. While it does record the researchers' ratings about primary study selection and quality assessment, it does not support the resolution of disagreements among researchers. SLuRp does record bibliographic data for studies imported into the tool, but removal of duplicates and merging of conflicting entries is still a manual process. Finally, SLuRP lacks facilities to export bibliographic and other extracted data into commonly accepted formats for importation into specialized tools, such as EndNote or statistical packages.

Some other observations about these two tools highlight the need for the creation of a new infrastructure. Both follow a strict interpretation of the SLR process and do not fully support the inclusion, and use, of techniques such as snowballing to recover relevant studies that may have been missed due to the nature of database searches. StaRt does allow these techniques to be used during the piloting phase, but they cannot be included as a formal part of the review. Additionally, neither tool facilitates ongoing maintenance of reviews or the ability for researchers

to build upon previous work. Prior search results, quality assessments and extracted data are not readily available for use in the updating of existing reviews or the construction of new reviews.

4 Infrastructure Characteristics and Features

While we realize that the activities of this planning grant, specifically those defined in Section 5, may provide additional inputs that could change the community needs identified in Section 3, we do not anticipate any significant changes in the focus of the project. As a result, to illustrate the proposed infrastructure, this section provides an overview of our proposed solution. We plan to develop a web-based cyberinfrastructure that will enable: 1) coordination among multiple (possibly distributed) researchers, 2) community review of protocols, 3) automated interaction with different databases, 4) improved quality assessment, and 5) simple, repeatable data extraction.

4.1 Addressing Community Needs

Our approach to addressing each of these needs is described in the remainder of this section.

Coordination among multiple authors The web-based nature of the infrastructure will enable researchers who are not collocated to collaborate on execution of the same SLR protocol. The description of each of the requirements below also highlight features that enable multiple authors to collaborate on the same SLR using the proposed infrastructure.

Community review of protocols The first and arguably most important step of the SLR process is the development, documentation, and validation of the protocol. The protocol includes: the motivation for the study, the research question(s), the sources for the primary studies, the search strings, the inclusion and exclusion criteria, the data extraction information, and the process details for each step of the execution phase. During construction of the protocol, the tool will also allow for the collaboration and feedback among the members of the SLR team. Once completed, the infrastructure will provide a mechanism for the public review of the proposed protocol and allow for community feedback. This feature will help provide researchers with confidence that the protocol is sound and will increase the chances of publication, if it is well-executed.

Automated interaction with databases To support the time consuming process of searching for and identifying primary studies the tool will provide a number of features. First, it will support both manual and automated searching of multiple databases. Second, it will allow for manual importation or automated retrieval of search results. Third it will help researchers create robust search strings by analyzing the provided keywords and suggesting alternatives, based upon the history of previous SLRs. Fourth, the system will adapt the standardized search strings defined in the protocol to the idiosyncrasies of known databases, where possible. Fifth, the tool will automatically detect duplicate papers in the various result sets, perhaps by using DOIs. Upon completion of the search phase, the tool will allow multiple researchers to evaluate the search results against the predefined inclusion and exclusion criteria. As each researcher's evaluation is captured independently, the tool will also provide inter-rater agreement ratings and isolate studies in need of further discussion among the SLR team.

Improved quality assessment Once the final set of candidate studies is determined, the tool will support the quality assessment for each study based on previously defined criteria. As different types of research (e.g., lab experiments, field studies, etc.) should be evaluated based on criteria appropriate for the design, the system will allow researchers to classify and subsequently

evaluate, using the appropriate criteria, each of the identified primary studies. Again, using the history of previous SLRs, the tool will help the researcher refine the quality assessment criteria by suggesting additional criterion that are related to those specified by the researcher. Once again, as each researcher's evaluation is captured independently the tool will provide information regarding agreement among the researchers and facilitate resolution of any conflicts.

Simple, repeatable data extraction For the final set of studies, the tool will provide a mechanism to facilitate data extraction. For strictly quantitative data extraction, the tool will perform an automated comparison to determine agreement among the researchers. To evaluate the extraction of qualitative information, the tool will provide a facility for third-party evaluation of the researchers' data extractions. As with other portions of the system, a means of evaluating agreement among the researchers and resolving conflicts will be made available. As the SLR moves into the analysis and then documentation stages, the tool will permit researchers to export citation information, in various formats, and all data collected during the extraction phase.

4.2 Enabling Future Research

As indicated in the previous section, as each SLR performed utilizing the tool is completed, all of its data will be made available for inclusion in subsequent reviews, with proper attribution to the original researchers. We envision that this feature will support the establishment of research communities focused on a given topic or domain. As new research is completed, studies can be added to the existing knowledge base and processed for inclusion in updated SLRs.

While the infrastructure initially targets the research community, inclusion of practitioners in the target community could provide additional possibilities. Practitioners may provide expert opinions and evaluations of SLR topics, or provide guidance related to important research questions that need to be answered. Furthermore, the system may facilitate the transfer of knowledge gained from research to practitioners in the field.

Part III Remaining Work

In this part we describe the steps that we will take to identify the consensus needs of the community, to identify the major characteristics and features of the infrastructure, and to foster community involvement in the planning process. Specifically, Section 5 describes our plans to supplement and improve the preliminary requirements described in Section 3 and to validate and prioritize the improved requirements. Further, Section 6 describes our plans to update, supplement, and refine the infrastructure proposal described in Section 4 and to prioritize the characteristics and features included in the infrastructure proposal.

The general process we will follow in the remaining work is to first perform data collection and analysis, via surveys or literature reviews, then generate a proposal (of community needs or characteristics and features of the infrastructure), then interact with the community, via workshops and surveys, to gather feedback and evaluate our proposal. This process is iterative and will be repeated multiple times. Of the 15 tasks that constitute the remaining work, 6 tasks foster community involvement in the planning process via workshops and surveys, and 5 tasks implement changes to the requirements or to the infrastructure proposal based on community input, feedback, and evaluation efforts.

5 Identification of Community Needs

This section describes our plan to identify the consensus needs of the community by supplementing and improving the preliminary community needs described in Section 3 and by validating and prioritizing the improved community needs. Our plan comprises the following tasks:

Data Collection and Analysis

T1. Extract additional process details from published SLRs

T2. Code additional survey responses from SLR authors

T3. Analyze the extracted process details and coded survey responses

Evolution of Identified Community Needs

T4. Supplement the preliminary community needs based on the analysis results

T5. Solicit community feedback on the supplemented needs

T6. Improve the supplemented community needs based on the community feedback

Validation and Prioritization of Identified Community Needs

T7. Validate the improved community needs using community evaluation

T8. Solicit community input on the prioritization of the validated community needs

T9. Prioritize the validated community needs based on the community input

Tasks T5 and T7 involve workshop organization. The following paragraphs describe the workshop agendas. Before each workshop we will post an announcement to the SEWORLD mailing list and send invitations to SLR authors. We will also contact survey respondents that have expressed interest in this project. After each workshop we will document activities and outcomes, post them to the project Wiki, and announce their availability.

Data Collection and Analysis Tasks T1–T3 involve the improvement of our preliminary analysis of process details from the published SLRs and survey responses from the SLR authors. Task T1 is to improve our preliminary analysis of the 214 published SLRs described earlier. We will extract additional process details from each of the published SLRs to permit a more complete analysis of the SLR process, including the difficulties that authors face and the strategies that they use to ease those difficulties. Task T2 is to improve our preliminary analysis of the 50+ responses to our SLR author survey. We will code additional author responses from each of the survey responses to permit a more complete analysis of SLR process data, including labor-intensive, automatable process phases. Task T3 is to analyze all collected data.

Evolution of Identified Community Needs Tasks T4–T6 involve the evolution of the preliminary community needs. Task T4 is to supplement the preliminary community needs based on the results of Task T3. We will both refine existing and formulate new community needs based on the analysis results. Task T5 is to solicit community feedback on the supplemented community needs that result from Task T4. We will organize a half-day workshop at the 2013 International Conference on Software Engineering (ICSE'13). Because an award notification would arrive just before the conference (in May), participation is likely to be limited to those already planning to attend ICSE'13. Thus, we will contact colleagues directly to request their feedback. Task T6 is to improve the supplemented community needs based on the community feedback from Task T5.

Validation and Prioritization of Community Needs Tasks T7–T9 involve the validation and prioritization of the improved community needs that result from Task T6. Task T7 is to validate the improved community needs with help from the community. We will organize a workshop at the

2013 International Symposium on Empirical Software Engineering and Measurement (ESEM). During the day-long workshop, participants will refine and validate the improved community needs via a series of evaluation activities. Time permitting, we will solicit participant input on the prioritization of the validated community needs. Task T8 is to solicit additional community input related to prioritization of community needs. We will solicit this input via a survey. Task T9 is to prioritize the validated community needs based on the community input from Task T8.

6 Infrastructure Characteristics and Features

In this section we describe our plan to identify the major characteristics and features of the infrastructure by updating, supplementing, and refining the infrastructure proposal described in Section 4 and by prioritizing the characteristics and features included in the infrastructure proposal. Our plan comprises the following tasks:

Infrastructure Evolution

T10. Update the infrastructure proposal based on the validated, prioritized community needs
T11. Solicit community feedback on the updated infrastructure proposal
T12. Supplement the updated infrastructure proposal based on the community feedback
T13. Refine the supplemented infrastructure proposal using community evaluation

Characteristic and Feature Prioritization

T14. Solicit community input on the prioritization of the proposed characteristics and features
T15. Prioritize the proposed characteristics and features based on community input

Task T13 involves workshop organization. The following paragraphs describe the workshop agenda. We will publicize the workshop and record its activities and outcomes in the same manner as described in the previous section.

Infrastructure Evolution Tasks T10–T13 involve the improvement of the infrastructure proposal. Task T10 is to update the infrastructural proposal based on the results of Task 9 (from the previous section). We will both refine existing and define new characteristics and features based on the validated, prioritized community needs. Task T11 is to solicit community feedback on the updated infrastructure proposal that results from Task T10. We will request feedback from individuals who contributed to the identification of community needs and participated in the community needs validation and prioritization process. Task T12 is to supplement the updated infrastructure proposal based on the results of Task T11. We will both refine existing and define new characteristics and features based on the community feedback. Task T13 is to refine the supplemented infrastructure proposal with help from the community. We will organize a workshop at the 2014 International Conference on Evaluation and Assessment in Software Engineering (EASE'14) or at ICSE'14. During the day-long workshop, participants will refine the supplemented infrastructure proposal via a series of evaluation activities. Time permitting, we will also solicit participant input on the prioritization of the proposed characteristics and features.

Characteristic and Feature Prioritization Tasks T14 and T15 involve the prioritization of the characteristics and features included in the refined infrastructure proposal that results from Task T13. Task T14 is to solicit additional community input related to the prioritization of the proposed characteristics and features. We will solicit this input via a survey. Task T15 is to prioritize the proposed characteristics and features based on the community input from Task T14.

Part IV Project Plan

In this part we describe the project management plan and project timeline.

7 Project Management

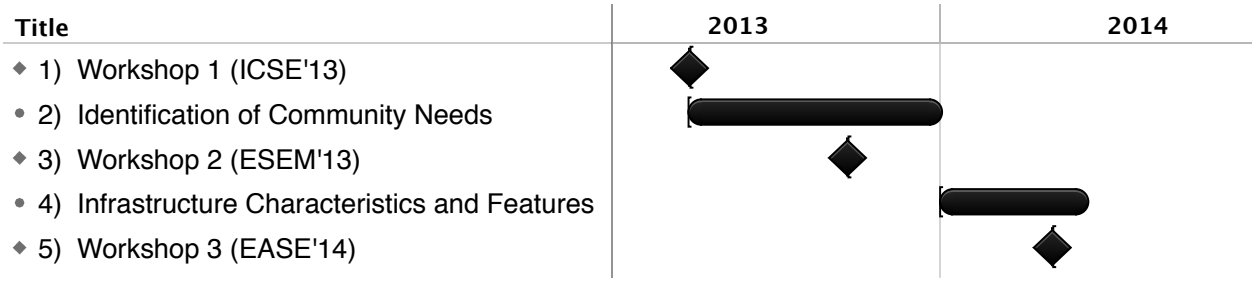
PI Carver has extensive experience with the SLR process. He has planned and executed several SLRs, including two that are published [**?**, **?**], one that is under review, and others that are in preparation. He has also taught an advanced graduate course on SLRs (Section 3.1.1). PI Carver will oversee the project and coordinate the data analysis tasks. He will also participate in planning community interactions.

Co-PI Kraft also has experience with the SLR process. He has planned and executed two SLRs, including one that is published [**?**] and one that is in preparation. He also has extensive experience in leading the development of software tools in an academic environment and in managing the deployment of web-based systems in a production environment. Co-PI Kraft will coordinate workshop organization tasks. He will also participate in planning other community interactions and in planning the development of the software tools.

Co-PI Hale is from the Management Information Systems (MIS) domain and previously spent nine years with a joint appointment in a medical school. He will provide input and feedback from an external perspective throughout the planning process. He will also participate in planning community interactions and in data analysis tasks.

Other project personnel include two Ph.D. students, one from MIS and one from CS. Both took PI Carver's SLR course and thus have have experience with the SLR process. The MIS student's dissertation work is on enabling a variety of research processes, of which the SLR process is a specific example. The CS student has planned and executed an SLR. The interactions between the students from MIS and CS will benefit not only the students, but also the overall project.

8 Project Schedule



Part V Results of Prior NSF Support

NSF CCF-0915559 (PI: Kraft, Co-PI: Carver), (2009-present), Title: *SHF: Small: Collaborative Research: Improved Code Clone Categorization*. In this project, PIs Kraft and Carver are developing and evaluating methods for categorizing code clone information to make it useful for developers. This project has supported the work of six graduate students. To date, this project has led to six workshop/conference [**?**, **?**, **?**, **?**, **?**, **?**] papers and three journal publications [**?**, **?**, **?**].

Co-PI Hale has no results in the previous 5 years.

Project Personnel

1. Jeffrey C. Carver; The University of Alabama; PI
2. Nicholas A. Kraft; The University of Alabama; Co-PI
3. David Hale; The University of Alabama; Co-PI

Budget Justification

A. Senior Personnel

The budget includes summer support for the senior personnel as follows: 0.4 months for Dr. Carver, 0.4 months for Dr. Kraft, and 0.09 months for Dr. Hale. All three will be involved in the management of the project activities and the planning of the community interactions.

B. Other Personnel

The budget includes support for the other personnel as follows: three summer months for one graduate research assistant (GRA) from Management Information Systems at $3,000 per month and one calendar year for one GRA from Computer Science at $1,750 per month. The monthly rate of $3,000 is the standard rate for Ph.D. students in the Management Information Systems Program, and the monthly rate of $1,750 is the standard rate for Ph.D. students in the Department of Computer Science.

C. Fringe Benefits

Fringe benefits were calculated based on the salary of the PI and Co-PI at the rate of 32% for the summer. This rate includes FICA, state retirement, SUI, and insurance. Fringe benefits for the graduate student include the FICA rate of 7.7% for the summer period only. Insurance for the graduate students is $1,348/year.

D. Equipment

No equipment is requested.

E. Travel

Travel funds are requested for the PI, Co-PIs and graduate students to attend the workshops in which community interaction will be performed. Those workshops will occur at international software engineering conferences including: The International Conference on Software Engineering (ICSE), The International Symposium on Empirical Software Engineering and Measurement (ESEM), and The International Conference on Evaluation and Assessment in Software Engineering (EASE).

G. Other Direct Costs

1. Materials and Supplies

The materials and supplies budget includes modest support for books and other resources.

6. Other

The budget includes funds to cover the tuition of one GRA for one academic year.

I. Indirect Costs

Indirect costs are calculated at the approved rate of 47% MTDC

L. Amount of This Request

The total amount of this request is $100,000.

Data Management Plan

This project will produce 4 types of data that require management:

1. Results of SLR on SLRs
2. Analysis of data from SLR author survey
3. Analysis of existing SLR tools
4. Output from community workshops

Results of SLR on SLRs In the proposal, we describe a preliminary analysis of the published SLR papers. As part of the proposed work, we will analyze each published SLR in detail. The results of the analysis will be a data extraction sheet for each published SLR. We will make these data extraction sheets available on the project website.

Analysis of data from SLR author survey In the proposal, we describe an analysis of a portion of the data gathered from our SLR author survey. As part of the proposed work, we will perform a much deeper analysis of that data. This analysis will include a the quantitative analysis along with a detailed description of the qualitative results from the survey. We will make the results of this analysis available on the project website.

Analysis of existing SLR tools In the proposal, we provide a high-level description of the existing SLR tools, based upon information available through published papers or websites. As part of the proposed work, we plan to interact with the tools in more detail (downloading them where possible). We will analyze, in detail, how these tools address a portion of the requirements for our proposed infrastructure. We will document the results of this analysis on the project website.

Output from community workshops We plan to conduct two community workshops to gather data about our preliminary requirements and our proposed solution. These workshops will be highly-interactive with the goal of generating concrete results that we can incorporate into our proposed infrastructure. For each workshop we plan to document the following information:

1. Input to workshop (i.e. proposed requirements or proposed solution
2. Workshop attendees
3. Summary of the discussion at the workshop
4. Output of the workshop (i.e. revised requirements or revised solution

Policies for Access and Sharing Data gathered during the project will be made freely available on the project website. Any data that could have personally-identifiable information will be anonymized prior to posting. If workshop attendees do not wish their identity to be known, this information will be redacted from the website.

Policies for Reuse Data extraction sheets from the SLR analysis will be available for other researchers to reuse. Detailed findings will be available as technical reports and published research reports.