Keywords: Web Migration, Systematic Mapping Study, SME, Software Migration, Web Engineering.

Abstract: Legacy systems are business-critical and contain valuable knowledge gathered over years of development. Many of these systems are still non-web desktop applications. Companies are aware of the advantages of web applications. However, Web Migration (WM), i.e. transforming legacy systems into web applications, is still a challenge, in particular for Small and Medium-sized Enterprises (SMEs). A large body of research exists in this field, but the specifics of SMEs have been overseen so far. This survey provides an overview on existing WM approaches with a dedicated focus on the SME perspective. The systematic mapping study analyses 122 primary studies and tools, identifies four major research focuses, reports on common methods and techniques and the consideration of SMEs. We then outline resulting research issues and future research directions.

1 INTRODUCTION

User expectations, advantages of platform-independent deployment and fading away of support for obsolete technologies provide a rationale for renewing legacy software as web applications.

Web browsers are becoming the standard interface for many applications and web applications provide a solution to platform-dependence and deployment issues (Aversano et al., 2001). However, keeping existing valuable knowledge that is contained in legacy systems (LS) is crucial in enterprise contexts. It is a key asset for the company typically representing the result of years of requirements and domain knowledge elicitation. Migration bears the risk of losing this knowledge and is therefore not easily undertaken. These two conflicting motivations are the origin of web migration research, which provides solutions allowing migrating to the Web without losing knowledge.

Web migration (WM) is a challenge for Small and Medium-sized Enterprises (SMEs) in particular, due to their characteristics and constraints (Nussbaumer and Liu, 2013). SMEs have limited human resources, typically less than 250 employees (European Commission, 2003). Therefore, it is not feasible to apply a migration approach which requires an entire development team exclusively dedicated to migration work items. SMEs typically target a very specific market segment with a limited customer base (Rose et al., 2016). In contrast to large enterprises, they have less financial resources (Nussbaumer and Liu, 2013) and no diverse portfolio of products/services to provide them with several sources of income. Thus, the risk of a failed migration is substantially higher. SMEs are not just little big businesses\(^1\). Regarding human resources, collaboration is often better but there is a lack of specialized expert professionals (Yew Wong and Aspinwall, 2004).

The relevance of SME companies has been acknowledged and expressed through various funding programs both in the EU\(^2\) and USA\(^3\). However, we were surprised to not find much work addressing WM for SMEs. Existing surveys (Khadka et al., 2013; Razavian and Lago, 2010; Spanos et al., 2012; Jamshidi et al., 2013; Gipp and Winter, 2007; Kienle and Diztante, 2014) in this field do not sufficiently address the characteristics of SME software companies.

Therefore, this survey aims at providing an overview on existing WM approaches and introducing the SME perspective to this field by providing the following contributions:

- Overview of the state of art in web migration
- Identification of the degree of consideration/support of SMEs in web migration

\(^1\)https://hbr.org/1981/07/a-small-business-is-not-a-little-big-business
\(^3\)https://www.sba.gov/offices/headquarters/oca/resources/6827
The remainder of this paper is organized as follows: In section 2 we outline the research method, in section 3 we report on the results and section 4 concludes the paper and gives an outlook on following work.

2 RESEARCH METHOD

In this systematic mapping study, we follow the guidelines provided by (Petersen et al., 2015).

2.1 Research Questions

To answer the initial question of “How can SMEs feasibly conduct a migration of legacy software to the web?”, we attempt to identify types of Web Migration approaches and assess their applicability for SMEs, resulting in the following research questions:

Q1 Which approaches/methods/models/processes/tools for migration of legacy to web applications exist?

Q2 Where and when were they published?

Q3 Which aspects of WM are addressed by them?

Q4 How can they be categorized?

Q5 Which approaches explicitly consider the situation and constraints of SMEs and how do they do so?

2.2 Data Sources & Search Strategy

In order to identify the primary studies available to answer the research questions, we used ACM Digital Library, IEEE Xplore, Science Direct, Scopus, Web of Science and Wiley Online Library as data sources.

Based on our previous research, we gathered search terms for migration in general, migration types and application types by listing synonyms as shown in Table 1. We constructed search queries combining the alternative search terms as shown in Table 2. These queries were executed on the above data sources considering titles, abstracts and keywords. To limit the time frame, we accepted only publications from year 2000 on. This search strategy yielded 870 initial search results.

2.3 Study Selection

Study selection ensures that results which match the search queries but are not relevant for the research questions are disregarded and only publications of sufficient quality are considered. To perform study selection, we define a list of inclusion and exclusion criteria as presented in Table 3 and Table 4. I1 and E1 - E3 relate to relevance and scope, I2 was added for consideration of quality. The inclusion and exclusion criteria are applied to titles and abstracts of the initial search results. If no clear decision can be made, the full text of the publication is checked. As a result of study selection, 92 publications have been considered as relevant primary studies.

2.4 Migration Tool Snowballing

Q1 also includes tools for migration. However, running the queries against scientific databases did not yield sufficient results for migration tools. Therefore, we applied the snowballing method (Wohlin, 2014) for the identification of tools. It uses traversal of references of a set of publications to identify additional ones and proceeds in three stages: Start Set, Iteration and...
Table 3: Inclusion Criteria.

<table>
<thead>
<tr>
<th>Id</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>Publications which contribute to the migration of user interfaces, databases, code, APIs or of an entire system to web-based systems. <strong>Motivation:</strong> elicit approaches that provide solutions for web migration.</td>
</tr>
<tr>
<td>I2</td>
<td>Peer-reviewed publications which have been published in journals, as conference papers or workshop papers. <strong>Motivation:</strong> assure quality by considering only work that has passed review of peers.</td>
</tr>
</tbody>
</table>

Table 4: Exclusion Criteria.

<table>
<thead>
<tr>
<th>Id</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Publications which do not explicitly contribute to the migration to web applications. <strong>Motivation:</strong> this study aims at finding web migration solutions, publications about other migration topics like migration to 4GL languages or to object-oriented databases are not considered.</td>
</tr>
<tr>
<td>E2</td>
<td>Publications which do not explicitly provide solution for migration scenarios. <strong>Motivation:</strong> this study aims at finding web migration solutions, publications about the development and maintenance of web applications or about the advantages and disadvantages of migration in general are not considered.</td>
</tr>
<tr>
<td>E3</td>
<td>Introductions, summaries, secondary studies or publications with less than three pages. <strong>Motivation:</strong> these publications do not provide sufficient information to provide objective solutions for migration scenarios.</td>
</tr>
</tbody>
</table>

Data Extraction. The primary studies are used as start set. To identify software tools for WM, backward snowballing is applied. We look for references to existing tools instead of referenced publications. Since software tools to not reference other tools, the iteration phase finishes after one step. Similar to study selection, the tools are filtered. Criteria \(I1\) and \(E1\) are also applied to the tools. In addition, we specify one tool specific exclusion criterion \(E4\) in table 5. This snowballing method identified 30 relevant primary tools which are added to the primary studies set.

2.5 Assessment of Approaches

The primary studies and tools are evaluated and results captured in a spreadsheet according to these criteria:

1. Publication/Tool Metadata
2. Migration Type

Table 5: Tool-specific Exclusion Criteria.

<table>
<thead>
<tr>
<th>Id</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>E4</td>
<td>Tools, which cannot be found online and are no longer available on the web either, neither commercially nor for free. <strong>Motivation:</strong> the tools included in this survey should be easily accessible for SME software companies, tools which have been referenced in publications but are no longer available online are therefore excluded.</td>
</tr>
</tbody>
</table>

3. Legacy and Target System
4. Migration Disciplines
5. Consideration of SMEs

**Publication/Tool Metadata:** captures general information on the evaluated approaches and tools, including title/name of the publication/tool, if applicable name of the presented framework, authors or companies involved, year of publication, URL reference, abstract or summary and secondary sources/reports.

**Migration Type.** Web migration comprises all areas of software migration: User Interface (UI), Code, Data, System interfaces and System. However, most publications/tools focus only on a subset of these areas. If applied in isolation, user interface, code and data migration require a strong separation of concerns in the source legacy system (LS). System interface migration has become increasingly important in the context of web services, e.g., migrating from SOAP to RESTful web services. As concerns are often not well separated in legacy code, system migration focuses on the joint migration of interfaces, application logic and data.

**Legacy and Target System.** WM approaches provide solutions to move from a legacy to a target system. These systems are defined by their architecture, underlying paradigms and concrete technologies. For UI migration, technologies in terms of programming languages/libraries of the LS and scripting/markup languages of the target system are considered. Code migration is described by legacy and target programming language. For data migration, we specify the database management systems. Legacy and target system interface are described by the interface paradigm (e.g., SOAP, REST) and libraries used. For system migration, architectural aspects (e.g. Web Services, Cloud, MVC) are considered.

**Migration Disciplines.** To describe migration activities, related work identifies several process models with varying phases/disciplines (Khadka et al., 2013; Razavian and Lago, 2010; Jamshidi et al., 2013; Lewis et al., 2006). In the context of this study, we follow the Reference Migration Process (ReMiP) (Gipp and Winter, 2007) metamodel, which provides a generic...
description framework adaptable to all existing migration process models. ReMiP identifies seven migration disciplines: Requirement Analysis, Legacy Analysis, Target Design, Strategy Selection, Implementation, Test, Deployment.

Explicit Consideration of SMEs. Q5 investigates the consideration of SMEs in web migration. This criterion highlights approaches which are explicitly targeting SMEs by addressing the characteristics described in section 1 such as limited human resources.

Appropriateness for SMEs. Even if the majority of approaches does not explicitly consider SMEs, different levels of appropriateness for SMEs can be identified. To achieve this, we use the following sub-criteria:

1. Evidence
2. Industrial Relevance
3. Tool support

For this quality criterion, we employ a simple overall rating which is formed by the sums of ratings for the sub-criteria. Sums from 0 to 2 are rated “(not sufficient), 3 to 5 “(medium) and 6 to 7 “(good). In the following, the partial ratings are stated in brackets.

Evidence. Available evidence of application of the approaches in this study in practice ranges from no evidence of application in practice (0), application in sample scenarios (1), pilot projects (2) to successful migration of entire systems (3).

Industrial Relevance. For evidence of application in practice, this distinguishes between application in industrial (1) or other, e.g. academic, contexts (0).

Tool Support. The available tool support is an important factor for SMEs. Not all approaches offer tool support (0). If there are tools for the given approach, these can be either just described in publications and not publicly available (1), or publicly available either commercially (2) or for free (3). However, a dedicated costs criterion could not be included since hardly any of the investigated approaches provides sufficient information for evaluation. Also, size and desired quality of a migration have a higher impact on migration costs than the chosen tool.

2.6 Threats to Validity

The survey was primarily designed by one researcher and executed and validated by three, including the designer. This introduces the chance of misinterpretation by the executors, a threat to construct validity. To address this, all executors were involved in major design decisions and regular meetings were held to assure a common understanding of the concepts and process, detailed descriptions for the terms were kept in a shared document.

Threats to internal validity are potential subjective biases in the process of the survey. At the beginning of our survey, we derived search terms from the research questions. As WM is a wide research area, different terminologies exist in academia and industry. Therefore, we included synonyms as search terms. However, our selection might not cover all relevant work and different synonyms might have yielded different studies.

To select primary studies, we employed inclusion and exclusion criteria. Their specification influences the scope of the analysis. To achieve transparency, we provide the motivation for each of the criteria. Also, publications and tools are inherently different, so we included an additional tool-specific criterion.

The analysis and categorization of the primary studies and tools is prone to bias. To mitigate this, we applied the categories established by ReMiP (Gipp and Winter, 2007) for migration type and discipline. The analysis was conducted by two researchers and cross-checked by a third. To make the results of the analysis comprehensive, we detailed the criteria. Since “Appropriateness for SMEs” explicitly rates, we try to reduce the subjective bias by breaking it down into three sub-criteria with simple, easy to identify scales.

Considering external validity, i.e. generalizability of results, we attempted to widen the scope of this survey beyond academia by explicitly adding industrial and commercial tools. However, since our start set for snowballing were the primary studies, tools not referred to in publications are not included.

3 RESULTS

For brevity, detailed results, additional figures and the full data of the entire survey is available online4.

Q1. Out of 870 initial search results according to our search strategy, 92 publications have been selected as primary studies using the inclusion and exclusion criteria described above. Additionally, the snowballing identified 30 primary tools.

Q2. An overview of the distribution of the primary studies is shown in fig. 1. Web migration has long been a research interest and the number of publications per year is increasing. The peak of 13 publications in 2013 correlates with increased research in cloud computing and thus web migration in the context of migrating to the cloud. The primary studies starting from 2004 show a focus on migration to SOA. Evolution of existing web systems can be found over the entire timespan.

We consider the distribution of primary studies across venues with a minimum threshold of 3 publica-

4https://vsr.informatik.tu-chemnitz.de/demos/WebMigrationSurvey
tions. Conferences like WSE (8 publications), ICSM (6), WCRE (6) or CSMR (5) from the software maintenance, evolution and re-engineering fields lead, followed by Software Engineering venues like STEP(3) and APSEC(3) and service-oriented computing conferences like SERVICES (3). What surprised us is that none of the Web or Web Engineering conferences like WWW or ICWE appear in this list. Apparently, the Web Engineering community does not consider migration aspects to any significant extent so far.

3.1 Migration Type & Discipline Coverage

Q3. Regarding the different migration types, out of 122 primary studies and tools, 20 addressed UI migration, 15 data migration, 2 interface migration, 7 code migration and 78 system migration. The high percentage of system migration approaches was expected since WM requires paradigm changes affecting all system aspects like towards asynchronous request-response, client-server separation in the spatial and technological dimension and URL-addressable UI states and navigation. Among the discrete aspects, UI is the most prominent, as it differs the most compared to non-web systems. This is followed by the data layer where, in the context of cloud computing, document- and column-oriented NoSQL databases have gained importance in addition to traditional relational databases.

Our survey identified the following coverage of migration disciplines: 32 address requirement analysis, 46 legacy analysis, 31 target design, 10 strategy selection, 86 implementation, 13 test and 10 deployment. This distribution shows high numbers for disciplines also found in traditional (forward) Software Engineering like requirement analysis and implementation. Legacy analysis, which is specific to software migration, also has attracted a high interest. What surprised us is the relatively low number addressing the test discipline. This seems to be in contrast to functional equivalence which is desirable in software migration in general. The tools are mostly dedicated to one or two migration disciplines and therefore limited in their coverage of the entire migration process.

We could find six approaches (Lucia et al., 2008; Khadka et al., 2011; Marchetto and Ricca, 2008; Bernhart et al., 2012; Mohagheghi and Sæther, 2011; Strauch et al., 2013) which provide full coverage of all disciplines and can therefore be considered as complete approaches with regard to ReMiP.

3.2 Methods, Techniques & Research Focuses

Q4. There is a large variety of methods in use which can be grouped, on top level, into three groups:

- **Transformation approaches**, that process the legacy source code by a series of automatic transformations turning it into the target system’s source code.
- **Reengineering approaches**, that employ reverse engineering techniques to extract information from the legacy system and follow up with forward engineering to create the target system and
- **Encapsulation approaches**, that wrap the unchanged legacy system or parts of it and expose a new interface which is then integrated with the target system.

We illustrate these groups using the six approaches that are complete wrt. migration discipline coverage. The incremental strategy in (Lucia et al., 2008) belongs to encapsulation approaches, creating *wrappers* for the legacy backend, however, combined with *UI reengineering*. (Khadka et al., 2011) combines the technical and business perspective in serviciFi, using method engineering. At the core, serviciFi employs *concept slicing* (program slicing in combination with concept assignment) followed by manual implementation. It is a reengineering approach. (Marchetto and Ricca, 2008) is an incremental, semi-automatic transformation approach, because functionality is extracted and
transformed into web services, using Axis2\(^5\) to create WSDL descriptions and stubs. (Bernhart et al., 2012) presents another incremental reengineering approach, with dedicated focus on parallel operation and a hard downtime constraint. Based on manual legacy analysis, functional requirements are elicited, rapid prototypes created, tested with users and implemented. Focusing on model-driven migration based on ADM, KDM, SoaML etc., REMICS (Mohagheghi and Sæther, 2011) combines reengineering with semi-automatic model-to-model transformation. Progressing in modernization sprints and employing meetings and artifacts from Scrum, it is incremental and agile. The approach in (Strauch et al., 2013) focuses on decision support for migrating the database layer of an application to the cloud. As it does not describe any concrete methods or tools for transforming data or encapsulating legacy databases, we classify it as a reengineering approach.

Our survey revealed four main research foci in web migration within the survey’s time frame:

- Migration to Web
- Migration to SOA
- Migration to Cloud
- Web Systems Evolution (WSE)

Migration to Web: is the most basic area of web migration. Research addresses the problem of bringing legacy non-web systems into the web. This comprises desktop TUI or GUI applications and mainframe systems. There is a large variety in what is considered the target “Web state”. Some minimal approaches cover only the creation of near-identical copies of legacy user interfaces in HTML, typically adding a middleware for handling HTTP-Requests and wrapping the legacy backend, like (Karampaglis et al., 2014). Others comprise a re-structuring, e.g. towards the MVC architecture (Bohduin et al., 2002). However, most approaches fall short of creating state-of-the-art web application results, e.g. with regard to rich interaction, responsiveness, let alone touch support or easy URL-based information sharing.

Another well researched area is migration to service-oriented architectures (SOA) (Khadka et al., 2013; Razavian and Lago, 2010). Here, the focus is on identification of services in legacy applications (Sosa et al., 2013) and architectural transformation towards a loosely coupled orchestration of these web services. This is mainly motivated by the reuse of valuable business logic (Khadka et al., 2013). SOA migration affects not only the technical but also the business perspective, thus there is a strong focus on business processes and business process modeling. Also, an increased interest in model-driven methods, e.g. based on OMG’s Architecture-Driven Modernization (ADM)\(^6\) and its Knowledge Discovery Meta-model (KDM) can be seen. Comprehensive research projects in the SOA migration area include REMICS considering SOA migration in the context of cloud, MIGRARIA\(^7\) with its focus on model-driven web engineering and SOAMIG\(^8\) based on compiler technology. Many approaches treat the legacy system as black box and therefore create wrappers, e.g. (Canfora et al., 2008) using interaction modeling by finite state automata. Regarding the target technologies, most of the work addresses migration to SOAP-based web services and related standards like WSDL.

Research in migration to the cloud focuses on the challenges from the change in environment and the SaaS paradigm. For instance, cloud platforms often provide specific NoSQL databases (Cai et al., 2015) and message queuing services for which LS have to be transformed to employ. Another main issue is data security, since migration to the cloud means involving a third party and adding data flows which can be compromised. This is one of the reasons why often hybrid architectures – partially on-premise and partially in the cloud – are the target for migration. Here, it is crucial to identify the distribution of components between on-premise and cloud optimized with regard to factors like communication costs (Huang et al., 2014).

Unlike the three research focuses described before, Web Systems Evolution: addresses legacy systems that are already web systems. Kienle et al. (Kienle and Distanze, 2014) provide an excellent overview, in which they identify a chain of evolution from static websites over rich internet applications (RIA), web services/SOA, ajax-based web applications to cloud-based and HTML5-based web applications. WSE approaches describe ways how to move between the different stages of this chain. The majority of recent web migration research belongs to this category.

There is little work on migrating pre-web systems to the latest evolution stages of this chain. Most research into migration to the web was conducted in the late 1990s and early 2000s. Target technologies like JSP or Flash and the resulting migrated systems are outdated by current standards. Modern HTML5-based user interfaces and user interaction patterns are only considered in the context of WSE.

\(^5\)http://ws.apache.org/axis2
\(^6\)http://adm.omg.org/
\(^7\)http://www.eweb.unex.es/eweb/migraria/
\(^8\)http://www.soamig.de/
3.3 SME Consideration

Q5: (Nussbaumer and Liu, 2013) is one of the few approaches to explicitly address SMEs. This work focuses on cloud migration, naming security, reliability, cost, performance, flexibility and service & support as SME cloud provider selection requirements. However, discipline coverage is low, and due to lack of sufficient evidence, industrial relevance and available tool support, it is rated low in SME appropriateness. In spite of a general interest in cloud computing, SME cloud adoption is still low and Nussbaumer et al. identify security concerns, the challenge of migrating existing applications and lack of understanding of business processes as key problems. The study confirms our own findings that SME-specific migration has not yet been discussed in research and is only addressed by proprietary commercial tools associated with high costs. Major non-proprietary frameworks “lack sufficient SME relevance” (Nussbaumer and Liu, 2013). (Bodhuin et al., 2002) state that their approach is in the context of “a research project aimed at defining new technological solutions to be transferred to Small and Medium Enterprises operating in the Information and Communication Technologies”, (Mohagheghi and Sæther, 2011; Lucia et al., 2006; Lucia et al., 2008) report on use cases from SME partners without explicit consideration of SME characteristics in their approaches. (Zhang et al., 2010) mentions SMEs for motivating their work by naming SaaS advantages for SMEs.

The tools in our survey rate generally high with regard to appropriateness for SMEs (average 7.8). The main reason are the high ratings for evidence, most of them provide evidence of successful migration of entire systems, and, evidently, for tool support.

4 CONCLUSIONS

Our survey of web migration and the SME perspective has identified two main research issues:

1. Lack of WM research addressing SMEs
2. Modern Web applications only considered in WSE

Few researchers have addressed WM for SMEs, in spite of their specific characteristics and relevance. From our own collaboration experience with industry partners, we know there are still many non-web legacy applications which could benefit from WM. This field provides challenging research problems and should get more attention in research. Integration of WM processes with agile development and maintenance activities, feasible WM methods for limited human/financial resources and lightweight transformations provide opportunities for researchers to contribute to advancing the state of the art in this field.

Most research into WM from non-Web to Web applications dates back to the early 2000s. Modern web applications based on HTML5 and new technologies like Web Components, Web Sockets, Web Storage, Web Workers are only considered in the context of WSE. In several research projects in collaboration with partners from the eHealth and industrial inspection & certification domains, we experienced that industry is still struggling to bring non-web LS to the Web. However, this problem has been neglected in recent research. Existing approaches are not applicable due to outdated technological foundations and lack of consideration of current web technologies.

We feel that both issues are related as the impact of the lack of SME-centered approaches is further intensified by the missing motivation to migrate to the web due to un-recent target systems of existing approaches, raising the initial hurdle for SMEs to migrate to the web. Therefore, new web migration approaches specifically targeted at SMEs are required, which can be applied with limited resources, integrated into day-to-day development and which help to produce modern HTML5-based web applications from non-web LS.

We plan to elaborate on an approach of this kind and to create supporting infrastructure in future work. This includes a more detailed specification of SME requirements and their impact on WM.

Web migration is an important research topic which is seeing increasing interest. We identified four major research focuses over the last 16 years. WM is a widely overseen topic in the Web Engineering community. Software Engineering and Maintenance communities have shifted focus towards SOA, Cloud and WSE. Legacy non-web to web approaches are scarce. Our survey of 122 studies and tools indicates a lack of consideration of SME characteristics. Based on the identified research issues, we briefly outlined potential research directions, in particular further specifying SME requirements for WM and creating suitable approaches. These should be developed and evaluated in close collaboration with SME software providers of different domains and origins.

ACKNOWLEDGEMENTS

The authors would like to thank Annika Krebs and Sven Maak for their valuable contributions to conducting this survey. This research was supported by the eHealth Research Laboratory funded by medatixx GmbH & Co. KG.
REFERENCES


