

Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering

Oliver Karras*, Felix Wernlein†, Jil Klünder† and Sören Auer*†

*TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany

Email: {oliver.karras, soeren.auer}@tib.eu

†Leibniz University Hannover, Hannover, Germany

Email: felix.wernlein@stud.uni-hannover.de, jil.kluender@inf.uni-hannover.de, auer@l3s.de

Abstract—[Background.] Empirical research in requirements engineering (RE) is a constantly evolving topic, with a growing number of publications. Several papers address this topic using literature reviews to provide a snapshot of its “current” state and evolution. However, these papers have never built on or updated earlier ones, resulting in overlap and redundancy. The underlying problem is the unavailability of data from earlier works. Researchers need technical infrastructures to conduct sustainable literature reviews. [Aims.] We examine the use of the Open Research Knowledge Graph (ORKG) as such an infrastructure to build and publish an initial Knowledge Graph of Empirical research in RE (KG-EmpiRE) whose data is openly available. Our long-term goal is to continuously maintain KG-EmpiRE with the research community to synthesize a comprehensive, up-to-date, and long-term available overview of the state and evolution of empirical research in RE. [Method.] We conduct a literature review using the ORKG to build and publish KG-EmpiRE which we evaluate against competency questions derived from a published vision of empirical research in software (requirements) engineering for 2020 – 2025. [Results.] From 570 papers of the IEEE International Requirements Engineering Conference (2000 – 2022), we extract and analyze data on the reported empirical research and answer 16 out of 77 competency questions. These answers show a positive development towards the vision, but also the need for future improvements. [Conclusions.] The ORKG is a ready-to-use and advanced infrastructure to organize data from literature reviews as knowledge graphs. The resulting knowledge graphs make the data openly available and maintainable by research communities, enabling sustainable literature reviews.

Index Terms—Knowledge graph, empirical research, requirements engineering, infrastructure, sustainability, literature review

I. INTRODUCTION

Empirical research in requirements engineering (RE) is a constantly evolving topic, with a growing number of publications [1]–[3]. Several publications examined how empirical research in RE is conducted and how it should be conducted in the future [2], [4]. Over the years, they presented snapshots of the “current” state and evolution of empirical research in RE [1]–[3], [5], [6] and, more generally, in software engineering (SE) [7]–[15]. They share the same goal of synthesizing a comprehensive, up-to-date, and long-term available overview of the state and evolution of empirical research in RE and SE. Although they share the same goal, use similar methods, i.e.,

(systematic) literature reviews, and even examine overlapping periods, venues, and themes (cf. Table I) [16], they have not collaborated to build on and update earlier works, which are known challenges of literature reviews [17]–[20]. Overcoming these challenges is critical to ensure the quality, reliability, and timeliness of research results from literature reviews [19], [21].

Recent research addresses these challenges by focusing on when and how to update (systematic) literature reviews in SE and its subfields [4], [21]–[23]. While these works mainly provide social and economic decision support and guidance for updating literature reviews [4], [20], the underlying problem is the unavailability of the extracted and analyzed data, corresponding to open science in SE [23], [24]. Unavailable data complicates collaboration among researchers and updating literature reviews, as the entire data collection, extraction, and analysis must be repeated and expanded for comprehensive results. Researchers need support in the form of technical infrastructures and services to conduct sustainable literature reviews so that all data is openly available in the long term [5], [17], [18], [20] according to the Findable, Accessible, Interoperable, and Reusable (FAIR) data principles [25], [26]. For this purpose, the data must be organized in a flexible, fine-grained, context-sensitive, and semantic representation to be understandable, processable, and usable by humans and machines [5], [13], [27]. Over the last decade, Knowledge Graphs (KGs) have become an emerging technology in industry and academia as they enable this versatile data representation [28]–[30]. Besides well-known KGs for encyclopedic and factual data, such as *DBpedia* [31] and *WikiData* [32], using so-called Research Knowledge Graphs (RKGs) for scientific data is a rather new approach [28], [29], [33]. RKGs include bibliographic metadata, e.g., titles, authors, and venues, as well as scientific data, e.g., research designs, methods, and results [34]–[39]. They are a promising technology to sustainably organize scientific data so that the data is openly available for long-term collaborations [27], [40].

We examine the use of RKGs as technical infrastructure by building, publishing, and evaluating an initial KG of Empirical research in RE (KG-EmpiRE). Similar to Frattini et al. [41], our long-term goal is to continuously maintain,

(re-)use, update, and expand KG-EmpIRE with the research community to synthesize a comprehensive, up-to-date, and long-term available overview of the state and evolution of empirical research in RE. In this way, we can *divide* the efforts to *conquer* the EmpiRE. We use the *Open Research Knowledge Graph* (ORKG), a RKG with services that combine manual crowdsourcing and automated approaches to organize scientific data [27]. Karras et al. [42] have successfully used the ORKG to organize qualitative and quantitative data [43], [44] from two systematic literature reviews in CrowdRE [45], [46].

Based on 570 papers from the *IEEE International Requirements Engineering Conference* (2000 – 2022), we show how scientific data on empirical research in RE can be consistently organized in the ORKG. In this way, we build and publish the initial KG-EmpIRE that the research community can constantly maintain, (re-)use, update, and expand. Similar to Abualhaija et al. [47], we evaluate KG-EmpIRE by analyzing its data to provide initial insights into the state and evolution of empirical research in RE. In particular, we answer competency questions [48], [49] derived from the vision by Sjøberg et al. [50] of how researchers should conduct empirical research in all fields of SE, including RE, in the period 2020 – 2025. The initial insights show a positive development towards the vision [50], but also the need for future improvements. We provide the following contributions:

Contribution:

- 1) The openly available KG-EmpIRE which the research community can maintain, (re-)use, update, and expand.
- 2) A reusable and expandable ORKG *template* for organizing scientific data on empirical research.
- 3) A set of 77 competency questions, the answers to which provide insights into the state and evolution of empirical research in SE and its subfields for the period 2020 – 2025.
- 4) A reproducible data analysis of KG-EmpIRE to answer 16 of 77 competency questions, providing initial insights into the state and evolution of empirical research in RE.

This paper is structured as follows: Section II explains the background. Section III discusses related work. While Section IV describes the approach, Section V reports the results. Section VI discusses threats to validity, and we discuss the findings in Section VII. Section VIII concludes the paper.

II. BACKGROUND

A Research Knowledge Graph (RKG) represents scientific data semantically, i.e., explicitly and formally, by linking (meta-)data of scientific artifacts (publications, datasets, and software) and entities (persons and organizations), which offers several benefits [28]. According to Auer et al. [28], the semantic representation leads to *better identification*, *traceability*, and *reduced ambiguity* of concepts and relationships of scientific data through terminological and conceptual clarity across disciplines. These improvements result in *easier (re-)use* of scientific data and thus *less redundancy and duplication*. For example, extracted and analyzed data of literature

reviews can be continuously (re-)used, updated, and expanded over time [26]. In addition, access to scientific data is easier for humans and especially for machines, as machines can grasp and understand the structure and semantics of scientific data in publications, so-called *machine actionability*. This improved access enables far-reaching opportunities for the *development of novel digital services* in science, such as customizable visualizations [51] and question answering systems [52], [53].

There are *generic* and *specific* RKGs [29], [40]. While generic RKGs focus on bibliographic metadata, e.g., titles, authors, and venues, specific RKGs focus on scientific data, e.g., research designs, methods, measurements, and results.

Generic RKGs focus on bibliographic metadata of scientific artifacts and entities. There are several well-known generic RKGs, such as *Microsoft Academic Knowledge Graph* [54], *OpenAlex* [55], *Springer Nature SciGraph* [56], *Semantic Scholar Literature Graph* [33], *OpenAIRE Research Graph* [57], [58], *Research Graph* [59], and *Scholarly Link Exchange* (Scholix) [60]. These RKGs have in common that they use bibliographic metadata to organize scientific artifacts, entities, and their relationships to enable, for example, their search, visualization, and processing [40], [61].

Specific RKGs focus mainly on scientific data combined with bibliographic metadata to describe and link scientific artifacts and entities. Specific RKGs are either specific to certain topics or more general to certain domains. Some well-known examples of topic-specific RKGs are *CovidGraph* [62], *COVID-19 Air Quality Data Collection* [38], [63], and *SoftwareKG* [64]–[66]. The first two RKGs address the topic of COVID-19. However, the *CovidGraph* looks more generally at scientific data on COVID-19 to explore publications, patents, existing treatments, and drugs around the coronavirus family [62]. In contrast, the *COVID-19 Air Quality Data Collection* is more fine-grained, focusing only on scientific content from publications about the impacts of COVID-19 lockdowns on air quality [38], [63]. The *SoftwareKG* deals with the topic of software that is mentioned in scientific publications. This RKG enables users to explore and understand the role of software in science. Besides topic-specific RKGs, there are several examples of domain-specific RKGs [40], such as *Computer Science Knowledge Graph* (CS-KG) [29], *Papers-with-Code* [36], *Hi Knowledge* [35], *Cooperation Databank* (CoDa) [39], and *OpenBiodiv* [37]. These RKGs organize scientific data from a specific domain, including computer science [29], machine learning [36], invasion biology [35], social sciences [39], and biodiversity [37].

In contrast to all RKGs mentioned, the ORKG is a special case as it is a RKG organizing any topic-specific scientific data across all research domains. Thus, the ORKG is a cross-domain and cross-topic RKG. The ORKG organizes scientific data provided by a publication as a collection of so-called *contributions*. A contribution consists of a semantic description of scientific data. Selected contributions can be compared in so-called *comparisons*. A comparison is a table where the columns denote the selected contributions by publication and the rows denote the semantically described scientific data.

Properties	Status of Empirical Research in Software Engineering <i>Empirical research - 2007</i>	The type of evidence produced by empirical software engineers <i>Empirical research - 2005</i>	Research in software engineering: an analysis of the literature <i>Empirical research - 2002</i>
method	Literature review	Literature review	Literature review
data availability	✗	✗	✗
time interval/time interval			
↳ has beginning*	1996-01-01	1997-01-01	1995-01-01
↳ has end*	2006-06-30	2003-12-31	1999-12-31
number of papers	133	119	369

Figure 1: *Comparison* of related publications on the “current” state and evolution of empirical research in RE and SE [16].

Figure 1 shows an excerpt from a comparison that we created to get an overview of related publications on the “current” state and evolution of empirical research in RE and SE [16]. For three publications, the excerpt shows the method used, the data availability, as well as the period and the number of papers examined. We use the ORKG due to its cross-domain and cross-topic characteristics, as well as its successful application for CrowdRE by Karras et al. [42].

III. RELATED WORK

Below, we review 14 publications that provide snapshots of the “current” state and evolution of empirical research in RE and SE (see Table I) [16]. We only consider publications that address the topic in general and are not limited to specific aspects, such as a method [67], [68] or a context [69], [70].

We found five publications on empirical research in RE published between 2005 and 2016 and nine on empirical research in SE published between 2002 and 2021. While one publication [1] examined empirical research in RE using a survey with 42 respondents, the other 13 publications [2], [3], [5]–[15] used (systematic) literature reviews or systematic mapping studies to analyze on average 402.9 papers (minimum: 20, median: 154, and maximum: 2237 papers) published between 1977 and 2019 with overlapping periods. In total, these 13 publications examined papers from a total of 60 different venues on 18 different themes. Nine of the 60 venues and ten of the 18 themes were examined by more than two publications. These facts show that there is considerable overlap and redundancy between these publications in terms of

goals, methods used, periods, venues, and themes examined. This overlap and redundancy could have been avoided if researchers had collaborated to build on and update earlier works. However, only four out of 14 publications offer their data at all, with only one publication [15] using a public data repository [71], [72]. The other three publications only offer links that no longer work [5], [12], [13].

In terms of key findings, the 14 publications show consistent results, although not all 18 themes were examined in all publications. For example, eleven of the 14 publications reported on the most commonly used research methods. Until 2000, the most common research methods were conceptual analysis and concept implementation [7]. Between 2000 and 2015, the most commonly used research methods changed to case studies and experiments [3], [5], [8]–[11], which were expanded after 2015 to also include surveys and systematic literature reviews [12]–[15]. While this change shows an evolution of research methods used, we also note that experiments and case studies have been the two main research methods for empirical research in RE and SE for more than 20 years. Although these two research methods have been used for a long time, seven publications concluded that there is a need to develop, expand, and use standardized terminology and theories (from other disciplines) to more consistently represent the empirical research conducted and better explain the results found [1], [2], [6]–[8], [12], [13]. In this regard, seven publications also analyzed the information reported for a comprehensive description of a research design. This information includes details about the research question(s) [1], contextual factors [6],

Table I: Details of related publications on the “current” state and evolution of empirical research in RE and SE [16].
Legend: Literature Review (LR), Systematic Literature Review (SLR), and Systematic Mapping Study (SMS)

Paper	Year	Field	Method	Period	Data basis	Dataset	Venues (Frequency > 2)	Themes (Frequency > 2)
[6]	2005	RE	LR	1986 – 2002	35 papers	Unavailable	1) Empirical Software Engineering Journal (8) 2) IEEE Software (4) 3) Requirements Engineering Journal (4) 4) ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (4) 5) IEEE Transactions on Software Engineering (3) 6) Information and Software Technology Journal (3) 7) IEEE International Requirements Engineering Conference (3) 8) Journal of Systems and Software (3) 9) International Conference on Software Engineering (3)	1) Data collection (12) 2) Research method (11) 3) Bibliographic metadata (10) 4) Data analysis (8) 5) Research paradigm (7) 6) Research design (7) 7) Research topic (5) 8) Research context (4) 9) Sample of population (4) 10) Theory (3)
[5]	2010		SLR	Unknown	154 papers	Broken link		
[1]	2012		Survey	19.03.2012 – 30.03.2012	42 respondents	Unavailable		
[2]	2014		LR	1983 – 2013	2237 papers	Unavailable		
[3]	2016		SMS	Open – 2012	270 papers	Unavailable		
[7]	2002	SE	LR	1995 – 1999	369 papers	Unavailable		
[8]	2002		LR	1996 – 2002	68 papers	Unavailable		
[9]	2005		LR	1997 – 2003	119 papers	Unavailable		
[10]	2006		LR	1977 – 2005	63 papers	Unavailable		
[11]	2007		LR	1996 – 2003	133 papers	Unavailable		
[12]	2015		SMS	1996 – 2013	891 papers	Broken link		
[13]	2018		SMS	2013 – 2017	538 papers	Broken link		
[14]	2019		SMS	1991 – 2014	341 papers	Unavailable		
[15]	2021		SMS	Open – 2019	20 papers	Available		

object of study [1], population/sample [3], [10], threats to validity [13], data collection [15], measurements/metrics [11], and data analysis [10], [13]. Overall, six publications [2], [3], [6], [10], [12], [13] explicitly state that the use of empirical research in RE and SE is constantly increasing. This increase is accompanied by the need to have a comprehensive, up-to-date, and long-term available overview of the state and evolution of empirical research [1], [5], [6]. This need is the underlying motivation of all 14 related publications.

We build on the related work by considering their examined venues and themes for our data collection, extraction, and analysis. Unlike related work, we do not conduct a full systematic literature review or systematic mapping study. Instead, we conduct a literature review to illustrate how researchers can use RKGs, specifically the ORKG, as a technical infrastructure for organizing scientific data in an openly available and long-term way to build and publish KGs that the research community can constantly maintain, (re-)use, update, and expand. We do not claim to provide a comprehensive overview of the state and evolution of empirical research in RE. Our research approach aims to lay the foundation for such an overview by building, publishing, and evaluating the initial KG-EmpIRE.

IV. RESEARCH APPROACH

The research approach and reporting essentially follow the *Empirical Standards for Software Engineering Research* [73].

We first define the research goal and research question to ensure that the scope of our research approach is clearly defined before presenting its details. We defined the research goal in detail using the goal definition template [74]:

Goal definition: We analyze the ORKG for the purpose of organizing scientific data in an openly available and long-term way with respect to building, publishing, and evaluating an initial KG of empirical research in RE that the research community can constantly maintain, (re-)use, update, and expand, from the point of view of ORKG users in the context of enabling sustainable literature reviews to synthesize a comprehensive, up-to-date, and long-term available overview of the state and evolution of empirical research in RE.

Based on this goal, we ask the following research question:

Research question: How can we use the ORKG as a technical infrastructure to organize scientific data in an openly available and long-term way by building, publishing, and evaluating a KG of empirical research in RE that the research community can constantly maintain, (re-)use, update, and expand to enable sustainable literature reviews?

We frame our research approach using the design science paradigm [75]. We collect papers from the field of RE and extract data on the empirical research reported. In this way, we build and publish the initial KG-EmpIRE as a solution design, which we analyze for evaluation. For this reason, our research approach consists of the three main steps: Data collection, data extraction, and data analysis (see Figure 2).

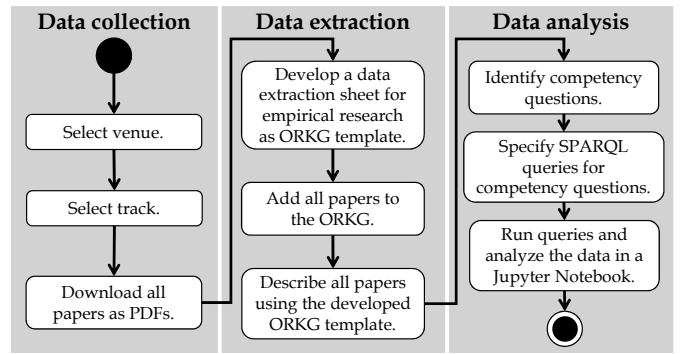


Figure 2: Activity diagram of the research approach.

A. Data Collection

Many related publications collected data on empirical research in RE and SE from papers (cf. Section III). However, the only available data set is small and covers only a subset of the frequently examined themes (cf. Table I) [16]. For this reason, we conducted another data collection (see Figure 2).

We considered only papers from one venue similar to other related publications [5], [8]–[11], [15] to simplify the search and selection. The selected venue is the *IEEE International Requirements Engineering Conference*, as authors from related publications reported that most of the papers they identified as relevant came from this conference [2], [3]. We selected the research track of the conference, as it is the main track where we expect most papers applying empirical research¹.

For data collection, we downloaded the papers as PDF files from *IEEE Xplore*, where all proceedings of the *IEEE International Requirements Engineering Conference* from 1994 – 2022 can be found. So far, we downloaded all 570 papers from the research track of the conference from 2000 – 2022 for data extraction. The collection of the missing years (1994 – 1999) is future work. We did not need to exclude any paper, as each paper reported at least partial information about empirical research that we could extract.

B. Data Extraction

The data extraction is the essential step in building and publishing the initial KG-EmpIRE. The basis of data extraction are the related publications with their data extraction sheets, themes, and contents (cf. Table I) [16].

We focused on the themes examined in more than five related publications. These themes are *data collection*, *research method*, *bibliographic metadata*, *data analysis*, *research paradigm*, and *research design*. For each theme, we analyzed the related publications to determine the analyzed content and possible values (see Table II). In this way, we determined the content for the data extraction covering the most frequently examined themes. Table II provides an overview

¹The selected venue and track are only a starting point for our work. We know that the other tracks of the *IEEE International Requirements Engineering Conference* are also relevant and that there are other important venues on the topic of empirical research in RE, such as the *Empirical Software Engineering* journal, *IEEE Software* journal, or *Requirements Engineering* journal, etc. (cf. Table I) [16].

of a subset of the content identified for data extraction. For a complete overview of all content identified for data extraction, refer to our [supplementary materials](#) [76].

Instead of a spreadsheet, we implemented the data extraction sheet as a so-called ORKG *template* [77] to organize the scientific data (see Figure 2). ORKG templates are an implementation of a subset of SHACL [78] and allow specifying the structure of ORKG *contributions* to describe a paper (cf. Section II). In this way, we determined which data to extract and ensured that all the semantic descriptions of scientific data are consistent and comparable across all considered papers.

The first two authors of this paper developed the *ORKG template* in four iterations over a period of two months. When developing the ORKG *template*, we focused on a generic design to ensure its reusability. Starting from an initial draft, we applied the (revised) template to five randomly selected papers from the data collection in each iteration. Based on our experiences in data extraction, we continuously adapted the template and always updated the descriptions of all papers from which we had previously extracted data. After the fourth iteration, there were no more changes. The remaining two authors reviewed the final version of the template and confirmed its suitability for data extraction. Figure 3 shows an excerpt from the ORKG *template* to illustrate the structure for describing the data collected (theme: *data collection*) and research question(s) posed (theme: *research design*) in a paper. Our ORKG *template* excludes bibliographic metadata, as the ORKG (semi-)automatically compiles the bibliographic metadata of a paper when the paper is added. For an overview of the ORKG *template*, refer to our [supplementary materials](#) [76].

For data extraction, we added each paper from data collection to the ORKG using its Digital Object Identifier (DOI). In this way, the ORKG automatically compiles the bibliographic metadata of the papers. The second author applied the developed ORKG *template* to each paper and extracted the corresponding data from all papers, using the terminology used in the paper to ensure an accurate and consistent description. The first and the third author reviewed each description by comparing the extracted data with the respective paper. In the case of inconsistencies or ambiguities, the three authors discussed and resolved the issues identified.

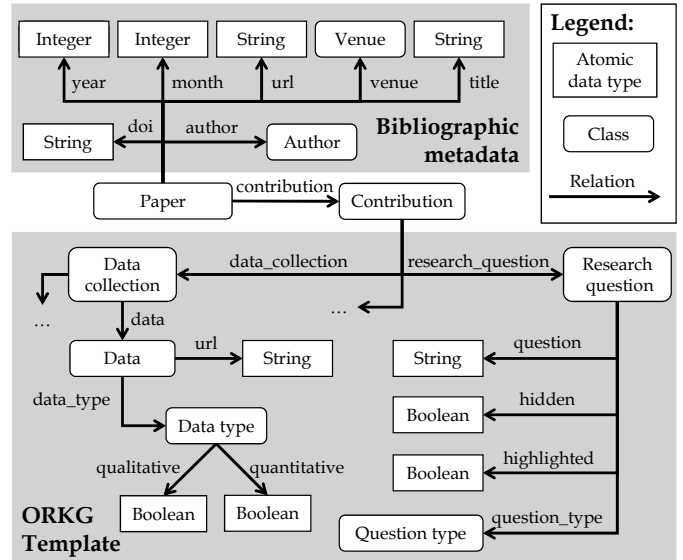


Figure 3: Excerpt from the ORKG *template* for data extraction.

C. Data Analysis

The data analysis serves two purposes: (P1) We evaluate the coverage of the curated topic of empirical research in RE by the initial KG-EmpIRE, and (P2) We get initial insights into the state and evolution of empirical research in RE.

Competency questions are an established method for analyzing and evaluating KGs [48]. A competency question is a natural language question that represents an information need related to the content of a KG and for which a KG must provide relevant information to answer the question [49]. Thus, the number of questions answered reflects the coverage of the curated topic in a KG (P1), and the answers to competency questions provide insights into the curated topic (P2).

Following our approach (see Figure 2), we identified relevant competency questions about the state and evolution of empirical research. We selected the vision of Sjøberg et al. [50] regarding the role of empirical methods in all fields of SE, including RE, to identify competency questions, as the vision precisely targets the current period of 2020 – 2025. Sjøberg et al. [50] present their vision by describing and comparing the “current” state of practice (2007) and their targeted state (2020

Table II: Overview of a subset of the content identified for data extraction.

Related Publications	Theme	Analyzed content	Possible values
[2], [3], [6]–[15]	Data collection	Data	Type of data: Qualitative, Quantitative Location of the data: URL
		Collection method	Case study, Survey, Interview, Experiment, etc.
[3], [5], [7]–[15]	Research method	Analysis method	Comparative analysis, Content analysis, Grounded theory, etc.
[1], [6], [7], [9]–[11], [13], [14]	Data analysis	Inferential statistic	Statistical test: t-test, ANOVA, Logistic regression, etc.
		Descriptive statistic	Type of measure: Frequency, Central tendency, Position, etc.
[3], [6]–[9], [13], [15]	Research paradigm	Type of paradigm	Exploratory, Explanatory, Descriptive, Predictive, etc.
[1], [3], [6], [10], [13]–[15]	Research design	Research question	Formulated question Presentation: Explicitly highlighted, Implicitly hidden Type of question: Exploratory, Explanatory, Descriptive, etc.
		Threats to Validity	Type of validity: Internal, External, Construct, Conclusion, etc.
		Metadata	Title, Authors, Venue, Publication date, URL, DOI
[2], [3], [7]–[14]	Bib. metadata	Metadata	Title, Authors, Venue, Publication date, URL, DOI

– 2025). The first three authors analyzed these descriptions by manually coding them in terms of textual elements that led to a question related to the state and evolution of empirical research. In this way, we derived the competency questions and captured their associated origins so that third parties better understand our analysis and its results. Subsequently, the first three authors matched each identified competency question with the analyzed questions and content of the related publications (cf. Section III) to determine how many times similar questions have been analyzed. Table III shows three examples of identified competency questions, including excerpts from the vision [50] with light gray highlighting of the coded text elements that led to the question and the references to related publications with a similar question. In total, we identified 77 competency questions about the state and evolution of empirical research in SE, including RE, of which 42 questions have been asked similarly in at least one related publication. For the detailed list of all 77 competency questions, refer to our [supplementary materials](#) [76].

For data analysis, we specified queries with a query language for KGs to retrieve the extracted data from the ORKG. This specification requires knowledge of the data structure, i.e., the ORKG *template*, so we can only specify queries for competency questions that can be answered with the extracted data. We used the query language SPARQL [79] as the ORKG provides a SPARQL endpoint for accessing all its data. Listing 1 exemplary shows the specified SPARQL query for competency question 1 (cf. Table III). In this case, we present the query with human-readable identifiers to facilitate understanding the query. This human-readable query is not executable in the ORKG as the ORKG uses alphanumeric identifiers (similar to *WikiData* [32]). For the executable queries, refer to our [supplementary materials](#) [76].

Listing 1: SPARQL query for competency question 1.

```

1 PREFIX r: <http://orkg.org/orkg/resource/>
2 PREFIX c: <http://orkg.org/orkg/class/>
3 PREFIX p: <http://orkg.org/orkg/predicate/>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema/#>
5
6 SELECT ?paper, ?year, ?dc_label, ?da_label
7 WHERE {
8   ?paper p:contribution ?contri;
9         p:publication_year ?year.
10  ?contri. a c:C27001.
11  OPTIONAL{?contri p:data_collection_method ?dc.
12            ?dc rdfs:label ?dc_label.}
13  OPTIONAL{?contri p:data_analysis_method ?da.
14            ?da rdfs:label ?da_label.}
15 }

```

Table III: Three examples of derived competency questions (cf. [supplementary materials](#) [76]).

ID	Competency question	Excerpts from the vision of Sjøberg et al. [50]		Related publications with a similar question
		State of practice (2007)	Target state (2020 – 2025)	
1	How has the proportion of empirical studies evolved over time?	There are relatively <i>few empirical studies</i> .	A <i>large number of studies</i> covering all important fields of SE and using different empirical methods are conducted and reported.	[2], [3], [10]
26	How have the proportions of case studies and action research in the empirical methods used evolved over time?	One may question the <i>industrial relevance of most SE studies</i> .	<i>More case studies and action research</i> should be carried out.	[3], [8], [10], [12], [13]
39	How has the provision of data (the materials used, raw data collected, and study results identified) evolved over time?	<i>Few studies provide results that enable efficient cumulative research</i> , [...]	<i>More research studies</i> are designed with the goal of <i>enabling efficient use of its results</i> by other researchers.	None.

We implemented the analysis using a *Jupyter Notebook* [76] with *Python*, published on *GitHub* [76] with the latest release archived on *Zenodo* [76]. We also hosted the repository on *mybinder* [76]. In this way, the analysis is always available to anyone for interactive reproduction and (re-)use, retrieving the latest data from the ORKG. Due to the uniform data structure provided by the developed ORKG *template*, we can always retrieve newly added papers that use the ORKG *template* and include them in the analysis by simply rerunning the script.

V. RESULTS

Below, we present the results of the data analysis. First, we address the coverage of the curated topic by the initial KG-EmpIRE (P1), followed by initial insights into the state and evolution of empirical research in RE based on the competency questions answered (P2).

A. Coverage of the Curated Topic by KG-EmpIRE

Overall, we answered 16 of the 77 competency questions (21%) using KG-EmpIRE. 13 of these 16 competency questions were asked similarly in at least one related publication.

Given the initial stage of KG-EmpIRE, the number of competency questions answered represents an acceptable coverage of the curated topic. So far, we have focused only on one track of one venue, and the ORKG *template* covers only six of the 18 different themes examined in related publications (cf. Section III). Therefore, the need to expand KG-EmpIRE by adding more papers from different venues and more data is evident. In particular, the organization of additional data to answer the open competency questions is necessary, which requires expanding the ORKG *template*. However, our goal was not to build and publish an already extensive KG of empirical research in RE to answer as many competency questions as possible. Instead, we aimed to lay its foundation by building, publishing, and evaluating the initial KG-EmpIRE. We conducted a literature review to illustrate how researchers can use RKGs, specifically the ORKG, as a technical infrastructure for organizing scientific data in an openly available and long-term way to build and publish KGs that the research community can constantly maintain, (re-)use, update, and expand.

B. State and Evolution of Empirical Research in RE

Due to the initial stage of KG-EmpIRE, we cannot provide an analysis of the general state and evolution of empirical research in RE. Nevertheless, we show some results from the

data analysis so far. Due to space limitations, we present only specific results for the three competency questions from Table III before reporting more generally on our initial insights.

CQ1: How has the proportion of empirical studies evolved?

According to the *Empirical Software Engineering* journal, “*Empirical studies presented here usually involve the collection and analysis of data and experience [...]*”. For this reason, we define that an empirical study always includes data collection and analysis. For each year, we examine the relative proportion of all papers that meet the definition to all papers collected, as the absolute number of papers varies per year.

Figure 4 shows that the proportion of papers that report an empirical study is always above 58%, averaging 79.6%, and increases slightly over time. While before 2010 the average proportion is 69.5%, the average proportion for the period 2010 – 2019 is 85.2%. For the target state of the vision (2020 – 2025) [50], the average proportion is 94.3%. Based on these results, we conclude a positive development towards the vision so that a large number of empirical studies can be achieved.

CQ26: How have the proportions of case studies and action research in the empirical methods used evolved?

For each year, we examine the relative proportion of all papers that report *case study* or *action research* as a data collection method to all papers reporting a data collection method, as the absolute number of papers varies per year.

In Figure 5, we present the proportion of papers per year that report *case study* or *action research*, each of which we address individually below. The proportion of papers using *case study* decreases over time, averaging 41.1%. While before 2010 the average proportion of papers is 53.7%, the average proportion for the period 2010 – 2019 is 34.1%. For the target state of the vision (2020 – 2025) [50], the average proportion is 23.6%. We assume this decrease is due to a better understanding of the term “*case study*” among researchers. A recent study by Wohlin [80] found that researchers often misused the term “*case study*” in software engineering research. We can confirm this finding as several papers analyzed use the term “*case study*”, although, at best, they report an experiment or a larger use case. Despite the decrease, this finding represents a positive development of the empirical research in RE, as researchers make better use of the term “*case study*”. The proportion of papers using *action research* is constantly low over time. In only eight years, papers use *action research* at all and the proportions are at most 10%, averaging 2%. Since 2018, no paper uses *action research*, so the average proportion for the target state of the vision (2020 – 2025) [50] is 0%. Based on these results, we conclude that the increased use of case studies and action research, as envisioned for the target state, has not yet been achieved.

CQ39: How has the provision of data (the materials used, raw data collected, and study results identified) evolved?

For each year, we examine the relative proportion of all papers that report at least one URL to their *data* to all papers reporting a data collection method, as the absolute number of papers varies per year.

Figure 6 shows that the proportion of papers that report at least one URL to their *data* increases remarkably over time, averaging 42%. While before 2010 the average proportion of papers is 25.4%, the average proportion for the period 2010 – 2019 is 49.8%. For the target state of the vision (2020 – 2025) [50], the average proportion is 71.3%. Based on these results, we conclude a positive development towards the vision that more empirical studies will provide their *data*.

Initial insights from the data analysis

Overall, the data analysis shows a positive development of the state and evolution of empirical research in RE towards the vision of Sjøberg et al. [50]. For the following statements, we provide all associated analyses with visualizations and explanations as an interactive *Jupyter Notebook* [76].

We found that the proportion of papers reporting an empirical study increases over time, with an average proportion of 94.3% for the target state (2020 – 2025). Regarding the use of empirical methods, we observed that the number of empirical methods used for data collection and analysis in a single paper increases over time, with three to four empirical methods being most frequently used in one paper overall. For the target state, researchers mainly use three to even five empirical methods in one paper with average proportions of 22% for three, 25.3% for four, and 26.7% for five empirical methods. For data collection, researchers frequently and constantly use the established empirical methods *experiment*, *secondary research*, and *survey*, with average proportions of 35.7% (*experiment*), 40% (*secondary research*), and 18.7% (*survey*) for the target state. We also found that the use of *case study* decreases over time, with an average proportion of 22.3% for the target state, and that *action research* is rarely used with an average proportion of 0% for the target state. For data analysis, researchers mainly and constantly use *descriptive statistics* with a proportion of 87.6% overall and 92% for the target state. In contrast, the use of *inferential statistics* with a proportion of 19.2% overall and 26.3% for the target state is low. We also found a positive development in the reporting of research design. The proportion of papers reporting *threats to validity*, providing *data*, and highlighting *research questions and answers* steadily increase over time with average proportions of 91.3% (*threats to validity*), 71.3% (*data*), and 23.7% (*research questions and answers*) for the target state.

Despite the positive developments towards the vision [50], we have also identified the need for future improvements.

According to Sjøberg et al. [50], more *case studies* and *action research* are needed for data collection to ensure the industrial relevance of empirical research. However, our results show that the use of *case studies* decreased, and researchers rarely use *action research*. More effort from the research community is required to achieve this part of the vision. For data analysis, the proportion of papers using *inferential statistics* is low (26.2% average proportion for the target state) and diverse. Based on the names, we found 57 different statistical tests in the 570 papers, often apparently using different spellings for partly the supposedly same test. For example,

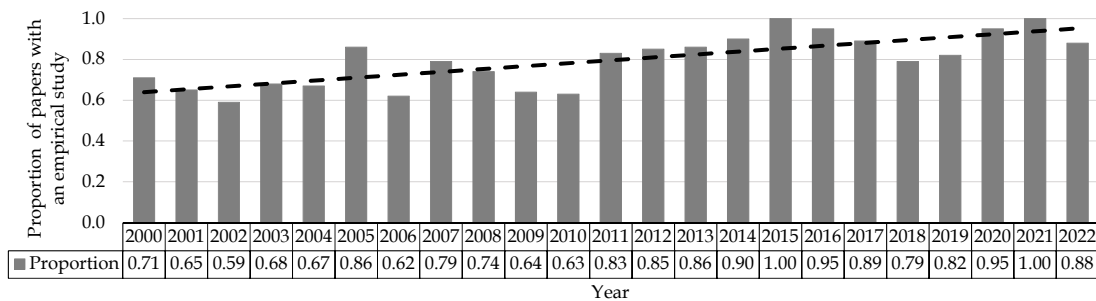


Figure 4: Proportion of papers per year reporting an empirical study with a linear trend line.

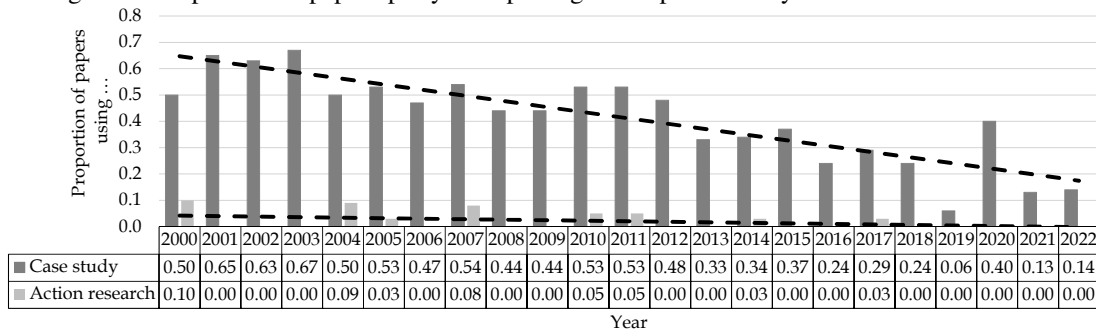


Figure 5: Proportion of papers per year using case study or action research with linear trend lines.

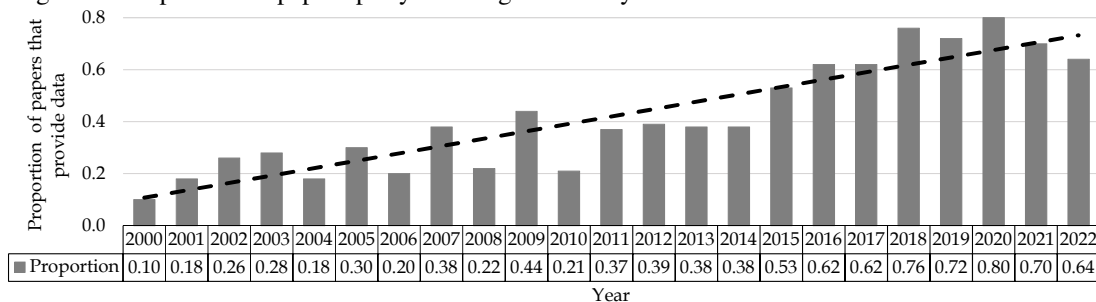


Figure 6: Proportion of papers per year that report at least one URL to their data with a linear trend line.

we found at least six different spellings for the *Mann-Whitney U test* [81], [82]. For a mature use of statistical methods as envisioned by Sjøberg et al. [50], the research community needs to be more concerned with *inferential statistics* and should use references to identify the statistical tests used. Regarding the reporting of *threats to validity*, we found two issues for future improvements. First, a proportion of 33.6% of the papers reporting *threats to validity* did not use any further classification of the types of validity. Although the general reporting of *threats to validity* is useful, the lack of classification makes it difficult for a reader to have a clear overview of whether the *threats to validity* have been comprehensively discussed. Second, a proportion of 18.2% of the papers reporting *threats to validity* addresses *conclusion validity*. This proportion is low compared to the proportions for other types of validity (*external validity*: 60.4%, *internal validity*: 56.1%, and *construct validity*: 47.9%). In the future, the research community needs to communicate threats to validity comprehensively and transparently by discussing all types of validity equally and naming the types of validity addressed.

All of these insights are initial and subject to various threats to validity, which we discuss in more detail in the following.

VI. THREATS TO VALIDITY

We discuss threats to *study selection*, *data*, and *research validity* based on the guideline for managing threats to validity of secondary studies in SE by Ampatzoglou et al. [83].

Study selection validity includes threats to the search process and study filtering. We ensured the *adequacy of initial relevant paper identification* by selecting all papers from the research track of the *IEEE International Requirements Engineering Conference*. This venue is the largest international RE conference, where established RE researchers regularly publish high-quality, peer-reviewed (empirical) research [2], [3]. The *selection of one publication venue* simplified the search and selection process due to the broad topic of empirical research in RE, while leading to a representative subset of relevant papers for building the initial KG-EmpIRE. However, the validity of KG-EmpIRE is limited due to the lack of relevant papers from other important venues. Collecting more papers from other venues is future work. We have been able to *access all papers* from the research track of the *IEEE International Requirements Engineering Conference* as our research institutes provide us access to [IEEE Xplore](#),

where all proceedings of the *IEEE International Requirements Engineering Conference* from 1994 – 2022 can be found.

Data validity includes threats to data extraction and data analysis. We mitigated *data extraction bias* and *researcher bias* by having a clear strategy: 1) The second author extracted the data from the papers, 2) The first and third authors reviewed the extracted data by comparing them with the papers, and 3) The three authors directly discussed and resolved any inconsistencies or ambiguities identified. For the data extraction, the first two authors developed an ORKG *template* in four iterations over two months that the remaining two authors finally reviewed and confirmed. We determined the contents of the ORKG *template* based on the related publications with their data extraction sheets, themes, and analyzed content (cf. Table II) to mitigate *classification schema bias*. *Publication bias* is present in the initial KG-EmpIRE, as we selected only papers from one publication venue. Nevertheless, we assume that the selected papers are representative of empirical research in RE in general and thus ensure the *validity of primary studies*, as the *IEEE International Requirements Engineering Conference* is known for its high-quality (empirical) research.

Research validity includes threats to the entire research process and design. This paper has a *research method bias* resulting from the use of only one method for data collection, namely a literature review. Although this method has its weaknesses, its systematic reflection is a proven means for building an initial, sound knowledge base [84], [85]. We also explicitly state that KG-EmpIRE is initial, and our long-term goal is to continuously maintain KG-EmpIRE with the research community to synthesize a comprehensive, up-to-date, and long-term available overview of the state and evolution of empirical research in RE. For these reasons, we assume that our research approach is appropriate to illustrate how researchers can use RKGs, specifically the ORKG, as a technical infrastructure for organizing scientific data in an openly available and long-term way to build and publish KGs that the research community can constantly maintain, (re-)use, update, and expand. We described our systematic research approach in detail and made all *data and materials* [76] openly available to ensure their *reliability*, *repeatability*, and *reproducibility*. Concerning *generalizability*, our initial insights are consistent with the findings of the related publications (cf. Section III). However, these initial insights are limited. Despite the reputation of the *IEEE International Requirements Engineering Conference* and its representativeness of the RE community, the data analysis mainly provides insights into the state and evolution of empirical research in RE published at this conference, so they do not necessarily reflect the state and evolution of empirical research in RE as a whole.

VII. DISCUSSION

Empirical research in RE is a constantly evolving topic, with a growing number of publications. The ever-growing number of publications is a well-known problem [86], [87], as it becomes almost impossible to keep track of the current state of research [27]. Therefore, there is a need for a comprehensive,

up-to-date, and long-term available overview of the state and evolution of empirical research in RE [1], [5], [6].

Over the years, several publications have addressed this need (cf. Section III), but none of these publications built on or updated earlier ones, resulting in considerable overlap and redundancy between them. The underlying problem is the unavailability of data from earlier work to maintain, (re-)use, update, and expand them (cf. Table I) [16]. Researchers need technical infrastructures and services to conduct sustainable literature reviews so that all data is openly available in the long term [5], [17], [18], [20]. In this paper, we examine the use of RKGs, specifically the ORKG, as such a technical infrastructure for building, publishing, and evaluating an initial KG of empirical research in RE, which is openly available for long-term collaboration among researchers. For this purpose, we collected 570 papers from the *IEEE International Requirements Engineering Conference* and extracted data from them on the six most frequently examined themes (see Table I) [16]. Based on 16 answered competency questions (out of 77), we present initial insights from the analysis of the KG-EmpIRE that are consistent with the findings of related publications.

Fundamentally, our research approach (cf. Figure 2) is the same as any literature review and consists of manual data collection, extraction, and analysis. Even if manual collection and extraction still require a lot of effort, there are first approaches [88], [89] that enable authors to describe new papers semantically while writing. For example, authors could combine our ORKG *template* with these approaches to collect the relevant data on their empirical research themselves. The resulting semantic descriptions can then be imported into RKGs, which can simplify the data collection and extraction for literature reviews in the long term and, in the best case, even make them obsolete. Bless et al. [89] have already presented a proof-of-concept by importing and updating semantic descriptions created with their approach into the ORKG.

Despite the similarity of our approach to any literature review, there are important differences in its implementation that leads to decisive advantages. Using the ORKG, the extracted data is not encapsulated in a file as usual, which is, at best, published on a data repository, but in an openly available knowledge graph, which, to put it simply, is nothing more than a graph-based database. Overall, the ORKG offers a ready-to-use and sustainably governed infrastructure that implements best practices, such as FAIR principles [25] and versioning, with services to support researchers in organizing FAIR scientific data [26]. As a result, the FAIR scientific data is openly available in the long term and can be understood, processed, and used by humans and machines. Thus, the research community can constantly maintain, (re-)use, update, and expand the initial KG-EmpIRE, that we have built, published, and evaluated, in a long-term and collaborative manner. For example, in case of errors in data extraction, anyone, and in the best case the authors themselves, can update the data. It is also possible to expand KG-EmpIRE by curating more papers using or even expanding our ORKG *template* to extract more data in a structured, consistent, and comparable way. In

all these cases, anyone can (re-)use our [Jupyter Notebook](#) [76] to reproduce the data analysis and its (updated) results.

Based on our results, the ORKG is a promising option as a technical infrastructure for conducting sustainable literature reviews and thus also for the continuous systematic literature review (CSLR) process envisioned by Napoleão et al. [4, p. 6]. For a full implementation of the CSLR process, however, the ORKG needs additional features, such as *forward snowballing* or *suggestion of potentially relevant papers*. Such future use and expansion of the ORKG is realistic and feasible as the ORKG team seeks to collaborate with others on use cases and new features for their system [42]. As an answer to our research question, we can summarize:

Answer to the research question: The ORKG is an innovative technical infrastructure with services that we can use *directly* for organizing scientific data in an openly available and long-term way to build, publish, and evaluate a KG of empirical research in RE that the researcher community can maintain, (re-)use, update, and expand. Moreover, due to its conception, the ORKG fundamentally enables building and publishing KGs on any topic from any domain, thus laying the foundation for sustainable literature reviews. Despite the existing functionality, there is a need to expand the ORKG to better support researchers in collecting papers, extracting data, and analyzing them. Such an expansion of the ORKG is realistic and feasible so that the ORKG even has the potential to become a suitable platform for the CSLR process.

For our future work, we have a plan with short-, mid-, and long-term actions. As short-term action, we expand KG-EmpIRE by describing more papers with our ORKG *template*. Our goal over the coming months is to cover the entire research track of the *IEEE International Requirements Engineering Conference* from 1994 – 2022 to get a comprehensive overview of the state and evolution of empirical research in RE at this conference. We also establish a more general ORKG *observatory* on [Empirical Software Engineering](#) as a central access point to all curated papers. The observatory is an open group that anyone can join to contribute to the topic. As mid-term actions, we write and publish an ORKG *review* about the state and evolution of empirical research in RE, based on the complete collection of all papers from the research track of the *IEEE International Requirements Engineering Conference*. An ORKG *review* is a special kind of literature review article that the research community can constantly maintain when underlying content in the ORKG changes due to updates or expansions [17], [18]. Besides the ORKG *review*, we also expand KG-EmpIRE by including more papers from other important venues (cf. Table I) [16] to gain a more comprehensive overview of the state and evolution of empirical research in RE. As long-term action, we expand our ORKG *template* to organize more extensive scientific data about empirical research in a structured, consistent, and comparable manner and thus to address the 61 still open competency questions.

With this plan, we work towards maintaining, updating, and expanding the initial KG-EmpIRE together with the research community by *dividing* the efforts to *conquer* the EmpiRE.

VIII. CONCLUSION

Empirical research in RE is a constantly evolving topic. Several publications address this topic, i.e., using (systematic) literature reviews. However, they only provide snapshots of the “current” state and evolution but no comprehensive, up-to-date, and long-term available overview of the state and evolution of empirical research in RE. The underlying problem is the unavailability of data from earlier works to build on and update collaboratively. While recent research addresses these challenges by providing social and economic decision support and guidance, researchers need technical infrastructures and services to conduct sustainable literature reviews.

We examine the use of RKGs, specifically the ORKG, as such a technical infrastructure. We conduct a literature review using the ORKG for organizing scientific data from 570 papers in an openly available and long-term way. As a result, we build and publish the initial KG-EmpIRE that the research community can constantly maintain, (re-)use, update, and expand. We analyze KG-EmpIRE for evaluation by answering 16 out of 77 competency questions derived from a published vision of empirical research in software (requirements) engineering for the period 2020 – 2025 [50]. Besides consistent findings with the related publications, the analysis shows a positive development towards the vision [50], but also the need for future improvements.

We conclude that the use of the ORKG and RKGs, in general, is a step in the right direction to allow researchers to build on and update earlier works, enabling sustainable literature reviews to ensure the quality, reliability, and timeliness of research results for successful long-term collaboration among researchers. Comprehensive, up-to-date, and long-term available overviews of the state and evolution of broad topics such as empirical research in RE are major research challenges that we as a research community can only *conquer* by *dividing* the efforts, true to the principle: *Divide et Impera*.

DATA AVAILABILITY

The data supporting the findings of this study are openly available in the [Open Research Knowledge Graph](#). We provide all [supplementary materials](#) [76], a [snapshot of data in the ORKG](#) [76], and the interactive [Jupyter Notebook](#) [76] of our analysis on [GitHub](#) and [Zenodo](#) [76].

ACKNOWLEDGMENT

The authors thank the Federal Government, the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the NFDI4Ing and NFDI4DataScience consortia. This work was funded by the German Research Foundation (DFG) - project numbers 442146713 and 460234259, by the European Research Council for the project ScienceGRAPH (Grant agreement ID: 819536), and by the TIB - Leibniz Information Centre for Science and Technology.

REFERENCES

- [1] N. Condori-Fernandez, M. Daneva, and R. Wieringa, "A Survey on Empirical Requirements Engineering Research Practices," in *18th International Working Conference on Requirements Engineering: Foundation for Software Quality*, 2012.
- [2] M. Daneva, D. Damian, A. Marchetto, and O. Pastor, "Empirical Research Methodologies and Studies in Requirements Engineering: How Far Did We Come?" *Journal of Systems and Software*, vol. 95, 2014.
- [3] T. Ambreen, N. Ikram, M. Usman, and M. Niazi, "Empirical Research in Requirements Engineering: Trends and Opportunities," *Requirements Engineering*, vol. 23, no. 1, 2018.
- [4] B. M. Napoleão, F. Petrillo, S. Hallé, and M. Kalinowski, "Towards Continuous Systematic Literature Review in Software Engineering," in *48th Euromicro Conference on Software Engineering and Advanced Applications*. IEEE, 2022.
- [5] M. Goeken and J. Patas, "Evidence-Based Structuring and Evaluation of Empirical Research in Requirements Engineering," *Business & Information Systems Engineering*, vol. 2, no. 3, 2010.
- [6] B. Paech, T. Koenig, L. Borner, and A. Aurum, "An Analysis of Empirical Requirements Engineering Survey Data," in *Engineering and Managing Software Requirements*. Springer, 2005.
- [7] R. L. Glass, I. Vessey, and V. Ramesh, "Research in Software Engineering: An Analysis of the Literature," *Information and Software Technology*, vol. 44, no. 8, 2002.
- [8] R. Jeffery and L. Scott, "Has Twenty-Five years of Empirical Software Engineering Made a Difference?" in *Ninth Asia-Pacific Software Engineering Conference*. IEEE, 2002.
- [9] J. Segal, A. Grinyer, and H. Sharp, "The Type of Evidence Produced by Empirical Software Engineers," in *Workshop on Realising Evidence-Based Software Engineering*, 2005.
- [10] C. Zannier, G. Melnik, and F. Maurer, "On the Success of Empirical Studies in the International Conference on Software Engineering," in *International Conference on Software Engineering*. Association for Computing Machinery, 2006.
- [11] A. Höfer and W. F. Tichy, "Status of Empirical Research in Software Engineering," in *Empirical Software Engineering Issues. Critical Assessment and Future Directions*. Springer, 2007.
- [12] A. Borges, W. Ferreira, E. Barreiros, A. Almeida, L. Fonseca *et al.*, "Support Mechanisms to Conduct Empirical Studies in Software Engineering: A Systematic Mapping Study," in *19th International Conference on Evaluation and Assessment in Software Engineering*, 2015.
- [13] L. Zhang, J.-H. Tian, J. Jiang, Y.-J. Liu, M.-Y. Pu, and T. Yue, "Empirical Research in Software Engineering - A Literature Survey," *Journal of Computer Science and Technology*, vol. 33, no. 5, 2018.
- [14] J. S. Molléri, K. Petersen, and E. Mendes, "CERSE - Catalog for Empirical Research in Software Engineering: A Systematic Mapping Study," *Information and Software Technology*, vol. 105, 2019.
- [15] C. Guevara-Vega, B. Bernárdez, A. Durán, A. Quina-Mera, M. Cruz, and A. Ruiz-Cortés, "Empirical Strategies in Software Engineering Research: A Literature Survey," in *International Conference on Information Systems and Software Technologies*. IEEE, 2021.
- [16] O. Karras, F. Wernlein, J. A.-C. Klünder, and S. Auer, "A Comparison of Scientific Publications on the State of Empirical Research in Requirements Engineering and Software Engineering," 2023. [Online]. Available: <https://orkg.org/comparison/R255464/>
- [17] A. Oelen, M. Stocker, and S. Auer, "SmartReviews: Towards Human-and Machine-Actionable Representation of Review Articles," in *International Conference on Asian Digital Libraries*. Springer, 2021.
- [18] —, "SmartReviews: Towards Human-and Machine-Actionable Reviews," in *International Conference on Theory and Practice of Digital Libraries*. Springer, 2021.
- [19] M. Bano, D. Zowghi, and N. Ikram, "Systematic Reviews in Requirements Engineering: A Tertiary Study," in *IEEE 4th International Workshop on Empirical Requirements Engineering*. IEEE, 2014.
- [20] V. dos Santos, A. Y. Iwazaki, K. R. Felizardo, É. F. de Souza, and E. Y. Nakagawa, "Towards Sustainability of Systematic Literature Reviews," in *15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2021.
- [21] E. Mendes, C. Wohlin, K. Felizardo, and M. Kalinowski, "When to Update Systematic Literature Reviews in Software Engineering," *Journal of Systems and Software*, vol. 167, 2020.
- [22] C. Wohlin, E. Mendes, K. R. Felizardo *et al.*, "Guidelines for the Search Strategy to Update Systematic Literature Reviews in Software Engineering," *Information and Software Technology*, vol. 127, 2020.
- [23] K. R. Felizardo, É. F. de Souza, T. Malacrida, B. M. Napoleão, F. Petrillo *et al.*, "Knowledge Management for Promoting Update of Systematic Literature Reviews: An Experience Report," in *46th Euromicro Conference on Software Engineering and Advanced Applications*, 2020.
- [24] D. Mendez, D. Graziotin, S. Wagner, and H. Seibold, "Open Science in Software Engineering," in *Contemporary Empirical Methods in Software Engineering*. Springer, 2020.
- [25] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton *et al.*, "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data*, vol. 3, no. 1, 2016.
- [26] M. Stocker, A. Oelen, M. Y. Jaradeh, M. Haris, O. A. Oghli *et al.*, "FAIR Scientific Information with the Open Research Knowledge Graph," *FAIR Connect*, vol. 1, no. 1, 2023.
- [27] S. Auer, A. Oelen, M. Haris, M. Stocker, J. D'Souza, K. E. Farfar, L. Vogt, M. Prinz, V. Wiens, and M. Y. Jaradeh, "Improving Access to Scientific Literature with Knowledge Graphs," *Bibliothek Forschung und Praxis*, vol. 44, no. 3, 2020.
- [28] S. Auer, V. Kovtun, M. Prinz, A. Kasprzik, M. Stocker, and M. E. Vidal, "Towards a Knowledge Graph for Science," in *8th International Conference on Web Intelligence, Mining and Semantics*, 2018.
- [29] D. Dessì, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, and E. Motta, "CS-KG: A Large-Scale Knowledge Graph of Research Entities and Claims in Computer Science," in *International Semantic Web Conference*. Springer, 2022.
- [30] H. Hussein, A. Oelen, O. Karras, and S. Auer, "KGMM - A Maturity Model for Scholarly Knowledge Graphs Based on Intertwined Human-Machine Collaboration," in *From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries: 24th International Conference on Asian Digital Libraries*. Springer, 2022.
- [31] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann *et al.*, "DBpedia: A Nucleus for a Web of Open Data," in *The Semantic Web*. Springer, 2007.
- [32] D. Vrandečić and M. Krötzsch, "Wikidata: A Free Collaborative Knowledge Base," *Communications of the ACM*, vol. 57, no. 10, 2014.
- [33] W. Ammar, D. Groeneveld, C. Bhagavatula, I. Beltagy, M. Crawford *et al.*, "Construction of the Literature Graph in Semantic Scholar," in *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3*. Association for Computational Linguistics, 2018.
- [34] M. Y. Jaradeh, A. Oelen, K. E. Farfar, M. Prinz, J. D'Souza, G. Kismihók, M. Stocker, and S. Auer, "Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge," in *10th International Conference on Knowledge Capture*, 2019.
- [35] J. Jeschke, M. Enders, M. Bagni, D. Aumann, P. Jeschke, M. Zimmermann, and T. Heger. (2020) Hi-Knowledge.org. [Online]. Available: <https://hi-knowledge.org/>
- [36] (2020) Papers With Code. [Online]. Available: <https://paperswithcode.com/about>
- [37] L. Penev, M. Dimitrova, V. Senderov, G. Zhelezov, T. Georgiev *et al.*, "OpenBiodiv: A Knowledge Graph for Literature-Extracted Linked Open Data in Biodiversity Science," *Publications*, vol. 7, no. 2, 2019.
- [38] G. I. Gkatzelis, J. B. Gilman, S. S. Brown, H. Eskes, A. R. Gomes, A. C. Lange, B. C. McDonald, J. Peischl, A. Petzold, C. R. Thompson, and A. Kiendler-Scharr, "The Global Impacts of COVID-19 Lockdowns on Urban Air Pollution: A Critical Review and Recommendations," *Elementa: Science of the Anthropocene*, vol. 9, no. 1, 2021.
- [39] G. Spadaro, I. Tiddi, S. Columbus, S. Jin, A. ten Teije *et al.*, "The Cooperation Databank: Machine-Readable Science Accelerates Research Synthesis," *Perspectives on Psychological Science*, 2022.
- [40] M. Stocker, T. Heger, A. Schweidtmann, H. Ćwiek-Kupczyńska, L. Penev, M. Dojchinovski, E. Willighagen, M.-E. Vidal, H. Turki, D. Balliet *et al.*, "SKG4EOSC - Scholarly Knowledge Graphs for EOSC: Establishing a Backbone of Knowledge Graphs for FAIR Scholarly Information in EOSC," *Research Ideas and Outcomes*, vol. 8, 2022.
- [41] J. Frattini, L. Montgomery, J. Fischbach, M. Unterkalmsteiner, D. Mendez, and D. Fucci, "A Live Extensible Ontology of Quality Factors for Textual Requirements," in *IEEE 30th International Requirements Engineering Conference*, 2022.
- [42] O. Karras, E. C. Groen, J. A. Khan, and S. Auer, "Researcher or Crowd Member? Why not both! The Open Research Knowledge Graph for Applying and Communicating CrowdRE Research," in *29th International Requirements Engineering Conference Workshops*. IEEE, 2021.
- [43] O. Karras and E. C. Groen, "Overview of Approaches that Classify User Feedback as Feature Request," Open Research Knowledge Graph, 2021. [Online]. Available: <https://doi.org/10.48366/R112387>

- [44] O. Karras and J. A. Khan, "Overview of Crowd Intelligence in Requirements Engineering," Open Research Knowledge Graph, 2021. [Online]. Available: <https://doi.org/10.48366/R114155>
- [45] R. Santos, E. C. Groen, and K. Villela, "An Overview of User Feedback Classification Approaches," in *International Working Conference on Requirements Engineering: Foundation for Software Quality, Workshops*, 2019.
- [46] J. A. Khan, L. Liu, L. Wen, and R. Ali, "Crowd Intelligence in Requirements Engineering: Current Status and Future Directions," in *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, 2019.
- [47] S. Abualhaija, C. Arora, A. Sleimi, and L. C. Briand, "Automated Question Answering for Improved Understanding of Compliance Requirements: A Multi-Document Study," in *IEEE 30th International Requirements Engineering Conference*, 2022.
- [48] M. Grüninger and M. S. Fox, "The Role of Competency Questions in Enterprise Engineering," in *Benchmarking – Theory and Practice*. Springer, 1995.
- [49] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutiérrez, S. Kirrane, J. E. Labra Gayo, R. Navigli, S. Neumaier, A.-C. Ngonga Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, and A. Zimmermann, *Knowledge Graphs*, ser. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool, 2021. [Online]. Available: <https://kgbook.org/>
- [50] D. I. Sjøberg, T. Dyba, and M. Jorgensen, "The Future of Empirical Methods in Software Engineering Research," in *Future of Software Engineering*. IEEE, 2007.
- [51] V. Wiens, M. Stocker, and S. Auer, "Towards Customizable Chart Visualizations of Tabular Data Using Knowledge Graphs," in *Digital Libraries at Times of Massive Societal Transition*. Springer, 2020.
- [52] M. Y. Jaradeh, M. Stocker, and S. Auer, "Question Answering on Scholarly Knowledge Graphs," in *International Conference on Theory and Practice of Digital Libraries*. Springer, 2020.
- [53] S. Auer, D. A. C. Barone, C. Bartz, E. G. Cortes, M. Y. Jaradeh, O. Karras, M. Koubarakis, D. Mourmoutsev, D. Pluukhin, D. Radyush, I. Shilin, M. Stocker, and E. Tsalapati, "The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge," *Nature Scientific Reports*, vol. 13, no. 7240, 2023.
- [54] M. Färber, "The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data," in *International Semantic Web Conference*. Springer, 2019.
- [55] J. Priem, H. Piwowar, and R. Orr, "OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts," *arXiv preprint arXiv:2205.01833*, 2022.
- [56] T. Hammond, M. Pasin, and E. Theodoridis, "Data Integration and Disintegration: Managing Springer Nature SciGraph with SHACL and OWL," in *International Semantic Web Conference*, 2017.
- [57] P. Manghi, A. Bardi, C. Atzori, M. Baglioni, N. Manola, J. Schirrwagen, P. Principe, M. Artini, A. Becker, M. De Bonis *et al.*, "The OpenAIRE Research Graph Data Model," *Zenodo*, 2019.
- [58] J. Schirrwagen, P. Manghi, N. Manola, L. Bolikowski, N. Rettberg, and B. Schmidt, "Data Curation in the OpenAIRE Scholarly Communication Infrastructure," *Information Standards Quarterly*, vol. 25, no. 3, 2013.
- [59] A. Aryani and J. Wang, "Research Graph: Building a Distributed Graph of Scholarly Works Using Research Data Switchboard," 2017.
- [60] A. Burton, H. Koers, P. Manghi, M. Stocker, M. Fenner, A. Aryani, S. La Bruzzo, M. Diepenbroek, and U. Schindler, "The Scholix Framework for Interoperability in Data-Literature Information Exchange," *D-Lib Magazine*, vol. 23, no. 1/2, 2017.
- [61] A. Brack, A. Hoppe, M. Stocker, S. Auer, and R. Ewerth, "Analysing the Requirements for an Open Research Knowledge Graph: Use Cases, Quality Requirements, and Construction Strategies," *International Journal on Digital Libraries*, vol. 23, no. 1, 2022.
- [62] D. Domingo-Fernández, S. Baksi, B. Schultz, Y. Gadiya, R. Karki, T. Raschka, C. Ebeling, M. Hofmann-Apitius, and A. T. Kodumullil, "COVID-19 Knowledge Graph: A Computable, Multi-Modal, Cause-and-Effect Knowledge Model of COVID-19 Pathophysiology," *Bioinformatics*, vol. 37, no. 9, 12 2020.
- [63] (2021) COVID-19 Air Quality Data Collection. [Online]. Available: <https://covid-aqs.fz-juelich.de>
- [64] D. Schindler, B. Zapilko, and F. Krüger, "Investigating Software Usage in the Social Sciences: A Knowledge Graph Approach," in *European Semantic Web Conference*. Springer, 2020.
- [65] D. Schindler, F. Bensmann, S. Dietze, and F. Krüger, "SoMeSci - A 5 Star Open Data Gold Standard Knowledge Graph of Software Mentions in Scientific Articles," in *30th ACM International Conference on Information & Knowledge Management*, 2021.
- [66] —, "The Role of Software in Science: A Knowledge Graph-Based Analysis of Software Mentions in PubMed Central," *PeerJ Computer Science*, vol. 8, 2022.
- [67] D. I. Sjøberg, J. E. Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N.-K. Liborg, and A. C. Rekdal, "A Survey of Controlled Experiments in Software Engineering," *IEEE Transactions on Software Engineering*, vol. 31, no. 9, 2005.
- [68] R. M. Bezerra, F. Q. da Silva, A. M. Santana, C. V. Magalhaes, and R. E. Santos, "Replication of Empirical Studies in Software Engineering: An Update of a Systematic Mapping Study," in *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2015.
- [69] T. Dybå and T. Dingsøy, "Empirical Studies of Agile Software Development: A Systematic Review," *Information and Software Technology*, vol. 50, no. 9-10, 2008.
- [70] T. Zhang, H. Jiang, X. Luo, and A. T. Chan, "A Literature Review of Research in Bug Resolution: Tasks, Challenges and Future Directions," *The Computer Journal*, vol. 59, no. 5, 2016.
- [71] C. Guevara-Vega, B. Bernárdez, A. Durán, A. Quiña-Mera, M. Cruz, and A. Ruiz-Cortés, "80 Initial Data-set Studies SMS Strategy CG," 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4456034>
- [72] —, "20 Primary Studies SMS Strategy CG," 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4455951>
- [73] P. Ralph, N. b. Ali, S. Balthes, D. Bianculli, J. Diaz, Y. Dittrich, N. Ernst, M. Felderer, R. Feldt, A. Filieri *et al.*, "Empirical Standards for software engineering research," *arXiv preprint arXiv:2010.03525*, 2021.
- [74] V. R. Basili, C. Caldiera, and H. D. Rombach, "Goal Question Metric Paradigm," *Encyclopedia of Software Engineering*, vol. 1, 1994.
- [75] P. Runeson, E. Engström, and M.-A. Storey, "The Design Science Paradigm as a Frame for Empirical Software Engineering," *Contemporary Empirical Methods in Software Engineering*, 2020.
- [76] O. Karras, "Analysis of the State and Evolution of Empirical Research in Requirements Engineering," 2023. [Online]. Available: <https://doi.org/10.5281/zenodo.8083529>
- [77] S. Kabongo, J. DSouza, and S. Auer, "Automated Mining of Leaderboards for Empirical AI Research," in *International Conference on Asian Digital Libraries*. Springer, 2021.
- [78] H. Knublauch and D. Kontokostas, "Shapes Constraint Language (SHACL)," W3C, W3C Recommendation, 2017. [Online]. Available: <https://www.w3.org/TR/2017/REC-shacl-20170720/>
- [79] A. Seaborne and S. Harris, "SPARQL 1.1 Query Language," W3C, W3C Recommendation, 2013. [Online]. Available: <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>
- [80] C. Wohlin, "Case Study Research in Software Engineering – It is a Case, and it is a Study, but is it a Case Study?" *Information and Software Technology*, vol. 133, 2021.
- [81] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, 1945.
- [82] H. B. Mann and D. R. Whitney, "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, 1947.
- [83] A. Ampatzoglou, S. Bibi, P. Avgeriou, and A. Chatzigeorgiou, *Guidelines for Managing Threats to Validity of Secondary Studies in Software Engineering*. Springer, 2020.
- [84] M. Glinz and S. A. Fricker, "On Shared Understanding in Software Engineering: An Essay," *Computer Science - Research and Development*, vol. 30, 2015.
- [85] O. Karras, K. Schneider, and S. A. Fricker, "Representing Software Project Vision by Means of Video: A Quality Model for Vision Videos," *Journal of Systems and Software*, vol. 162, 2020.
- [86] A. E. Jinha, "Article 50 Million: An Estimate of the Number of Scholarly Articles in Existence," *Learned Publishing*, vol. 23, no. 3, 2010.
- [87] K. White, "Publications Output: US Trends and International Comparisons. Science & Engineering Indicators 2020. NSB-2020-6." *National Science Foundation*, 2019.
- [88] L. Martin and A. Henrich, "RDFtX: Knowledge Exchange Between LaTeX-Based Research Publications and Scientific Knowledge Graphs," in *Linking Theory and Practice of Digital Libraries*. Springer, 2022.
- [89] C. Bless, I. Baimuratov, and O. Karras, "SciKGTEx - A LaTeX Package to Semantically Annotate Contributions in Scientific Publications," in *23rd ACM/IEEE Joint Conference on Digital Libraries*, 2023.