

Promotion of Open Science in Requirements Engineering – Leveraging the Open Research Knowledge Graph for FAIR Scientific Information

Oliver Karras, Alessio Ferrari, Davide Fucci, and Davide Dell'Anna

oliver.karras@tib.eu, alessio.ferrari@isti.cnr.it, davide.fucci@bth.se, d.dellanna@uu.nl

32nd IEEE International Requirements Engineering 2024 Conference – Exploring New Horizons:
Expanding the Frontiers of Requirements Engineering

June 24th, 2024, Reykjavik, Iceland

Welcome!



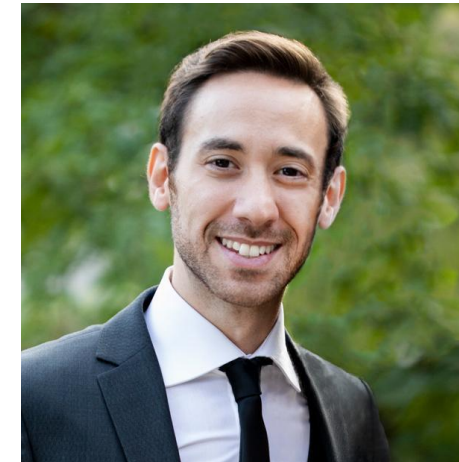
Oliver Karras



Alessio Ferrari



Davide Fucci



Davide Dell'Anna

Schedule and Table of Content

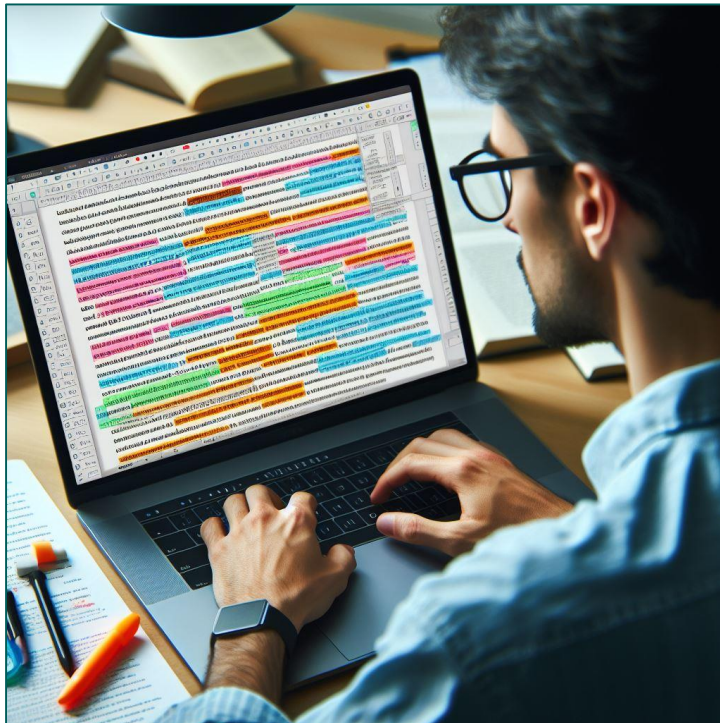
Session	Time	Table of Content	Style	Speaker
Theoretical	09:00 - 09:25	1. Welcome (5 min) 2. Introduction to open science in RE (10 min) 3. Introduction to the ORKG (10 min)	Presentation Presentation Presentation	All organizers Alessio Ferrari Oliver Karras
Practical	09:25 - 10:15	4. Create a FAIR-annotated publication for the ORKG (50 min) 4.1 Set up an Overleaf project for an exemplary publication 4.2 Use the LaTeX package SciKGT _E X to annotate the publication 4.3 Generate PDF with embedded FAIR scientific information 4.4 Optional: Upload the FAIR-annotated publication to the ORKG	Exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise	Oliver Karras All organizers All organizers All organizers All organizers
Break	10:15 - 10:45	Coffee break		
Practical	10:45 - 11:45	5. Use the ORKG based on a RE use case (60 min) 5.1 Add an exemplary publication to the ORKG 5.2 Describe the scientific information of the publication in the ORKG 5.3 Create an ORKG comparison of the publications added by participants 5.4 Publish the created ORKG comparison as a citable digital artifact 5.5 Optional: Create visualizations for the created ORKG comparison 5.6 Optional: Retrieve the information with the SPARQL endpoint	Exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise	Oliver Karras All organizers All organizers All organizers All organizers All organizers All organizers
Feedback	11:45 - 12:15	6. Reflection of the tutorial with the participants (25 min) 7. Farewell and closing (5 min)	Discussion Presentation	All organizers All organizers

Teaser: REFSQ'25 Open Science Competition

Challenge 1:

Annotate your REFSQ'25 paper with SciKGTeX.

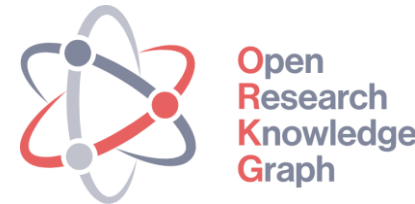
The accepted paper, best annotated with SciKGTeX, will be awarded the **Best ORKG Annotation Award** (prize: 100€).



Challenge 2:

Enrich your RESFQ'25 paper with an ORKG comparison.

The accepted paper, enriched with the best ORKG comparison, will be awarded the **Best ORKG Comparison Award** (prize: 200€).



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

Oliver Karras*, Felix Wernlein[†], Jil Klinder[†] 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.klinder@inf.uni-hannover.de, auer@t3s.de

Abstract—(Background.) Empirical research in requirements engineering (RE) is a constantly evolving topic, with a growing

(systematic) literature reviews, and even examine overlapping periods, venues, and themes (cf. Table 1) [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], [29] 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 *Dispedia* [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-EmpRE). Similar to Frattini et al. [41], our long-term goal is to continuously maintain,

A Comparison of Scientific Publications on the State of Empirical Research in Requirements Engineering and Software Engineering ★

November 2023 Oliver Karras Felix Wernlein Jil Ann-Christin Klinder Sören Auer

This comparison provides an overview of scientific publications that have investigated primary studies in requirements engineering and software engineering to give a snapshot of the “current” state of empirical research in requirements engineering and software engineering. In particular, the comparison shows for each publication (1) which research fields and topics were investigated, (2) whether and where the extracted and analyzed data is available, and (3) which method was used to determine the state, including further details about the respective method.

DOI: <https://doi.org/10.48366/RES0023>



Properties	Empirical research in requirements engineering: trends and opportunities <i>Empirical research - 2016</i>	Empirical research methodologies and studies in Requirements Engineering: How far did we come? <i>Empirical research - 2014</i>	A Survey on Empirical Requirements Engineering Research Practices <i>Empirical research - 2012</i>	Evidence-Based Structuring and Evaluation of Empirical Research in Requirements Engineering: Fundamentals, Framework, Research Map <i>Empirical research - 2010</i>	An Anal. Requires Data <i>Empirical research - 2010</i>
research problem	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering
research field investigated	Requirements Engineering	Requirements Engineering	Requirements Engineering	Requirements Engineering	Requirements Engineering
topic investigated	bibliographic metadata context data collection	bibliographic metadata research topic theory	context data analysis data collection	context research method result	context research method result

your personal use. Not for redistribution. The definitive version of record was published in the proceedings of 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), <https://doi.org/10.1109/ESEM56168.2023.10304795>.

Introduction: Open Science and FAIR in Software Engineering and Requirements Engineering

Oliver Karras, Alessio Ferrari, Davide Fucci, and Davide Dell'Anna

oliver.karras@tib.eu, alessio.ferrari@isti.cnr.it, davide.fucci@bth.se, d.dellanna@uu.nl

32nd IEEE International Requirements Engineering 2024 Conference – Exploring New Horizons: Expanding the Frontiers of Requirements Engineering

June 24th, 2024, Reykjavik, Iceland

What is Open Science?

Open science is the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination **accessible** to all levels of society, amateur or professional.

- Open methodology
- Open source
- Open data
- Open access
- Open peer review (identities, reports, community participation)
- Open educational resources



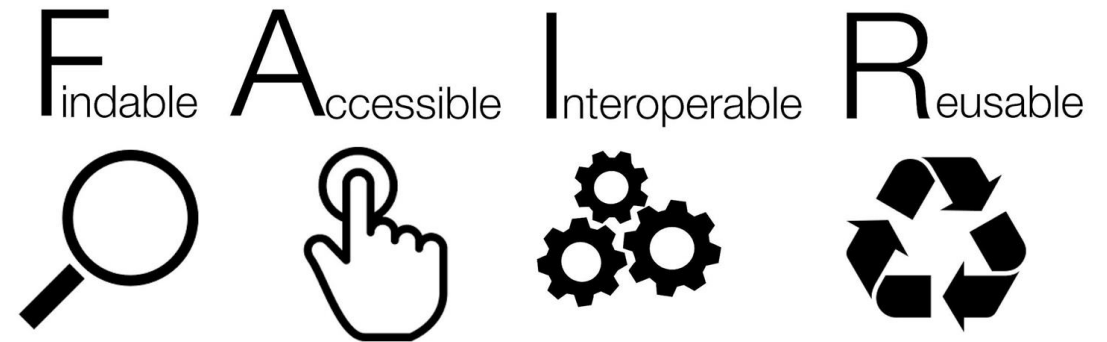
What are the FAIR Principles?

Scientists increasingly rely on computational support to deal with increasing amounts of data.

Focus on **Digital** Assets (Data & Knowledge)

FAIR Principles^[2]:

- **F**indability: Rich metadata, reachable by humans and machines, unique identifiers (DOI)
- **A**ccessibility: Open protocols for data access
- **I**nteroperability: Integration with other data
- **R**euse: Replicable experiments, understandable data (through metadata)



Why Do We Need Open Science and FAIR?

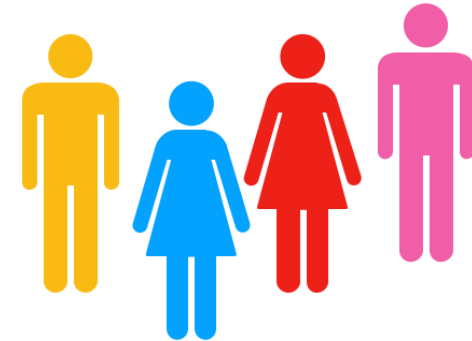
1. Transparency and Reproducibility:

- Open science helps verify the validity of research results
- Scientists can reproduce experiments, which is crucial for the **credibility**



2. Collaboration and Innovation:

- Collaboration among scientists, and **interdisciplinary** communication
- Accelerate discoveries and novel **ideas**



3. Equity and Inclusion:

- Democratize access to knowledge, and enrich viewpoints through **global access**
- Science accessible to the general public can increase scientific literacy, leading to an **informed society**



Current Initiatives in Software Engineering

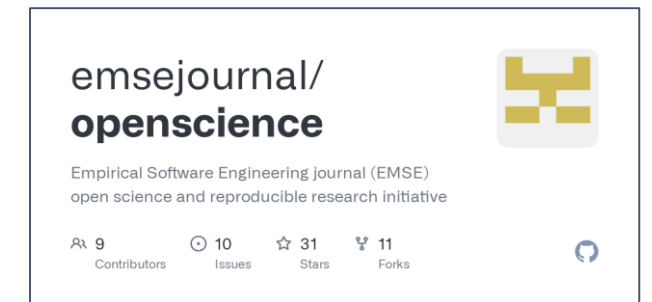
Emphasis on Data, **Tools**, Research Results



The ACM badge system (2010)



Artifact Evaluation Tracks



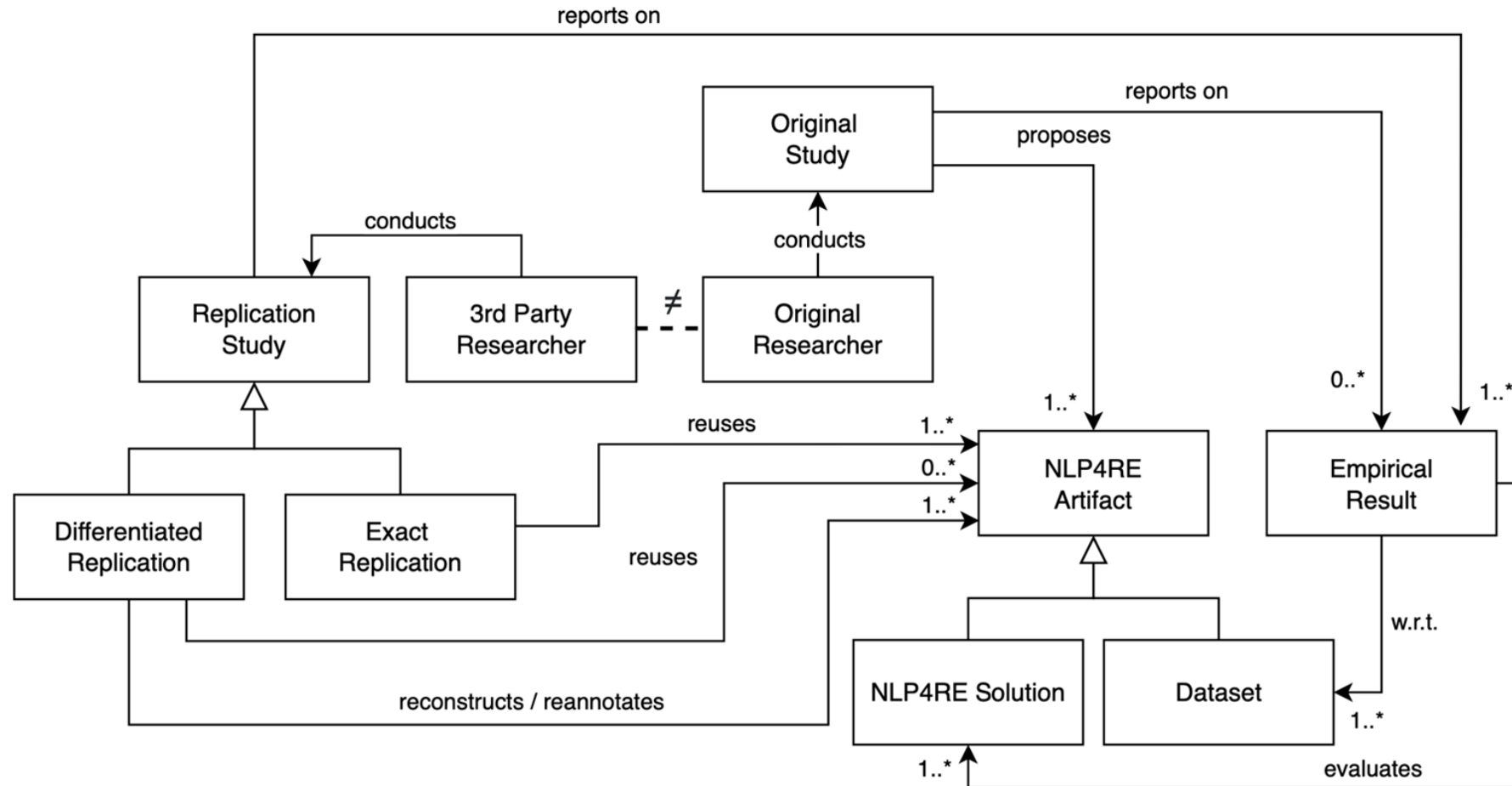
Home > ACM Journals > ACM Transactions on Software Engineering and Methodology > CFP

Sections

Replicated Computational Results (RCR) Report

Journal Initiatives

Replication and Reuse: Terminology



[3]

Example Problems in RE: Replication and Tool Reconstruction

ID	Description	
Rec1	The reconstruction-relevant information and implementation details of the original approach can be ambiguous, imprecise, and incomplete.	Poor information
Rec2	If a tool was partially or fully developed and/or evaluated using proprietary data, then there is no guarantee that the reconstructed tool would be identical to the original one since the used data cannot be accessed for reconstruction purposes.	Proprietary data
Rec3	Communication with the original authors is not always useful since the actual information sources may not be available anymore.	Communication
Rec4	The continuous evolution of the NLP ecosystem entails that some libraries become outdated, unavailable, or not maintained anymore.	Updates and maintenance
Rec5	Tools are typically developed as prototypes and not maintained in the long term.	
Rec6	Tool reconstruction is not (yet) valued as a self-standing research contribution, and hence researchers are discouraged to replicate tools overtime.	Poor value

Example Problems in RE: Dataset Annotation

ID	Description
Ann1	Some RE-specific categorization tasks lack solid theories that can guide the annotation process.
Ann2	Besides annotation experience and theoretical knowledge, the lack of domain knowledge can limit the accuracy of the annotations.
Ann3	The annotation activity is time consuming due to factors such as language barriers, different individuals' background, and fatigue.
Ann4	The annotation protocol can evolve and thereby necessitate the re-annotation of the data which might, again, cause additional time and effort.
Ann5	Theoretical and practical training resources and opportunities are limited and not adequate for training novice annotators who are often trained during the annotation task by more experienced annotators.
Ann6	The lack of benchmarks entails that annotated datasets enabling comparison against the state-of-the-art are scarce.
Ann7	Available imbalanced datasets pose the challenge of both understanding the minority class and consequently the annotation of new examples thereof.
Ann8	Determining the right amount of context to be shared alongside the annotation raw data is essential and can significantly affect the annotation results.
Ann9	Motivating the annotators poses another challenge since an immediate observation of the impact of a given annotation task is not always possible.
Ann10	Annotators are often not experienced in managing the social aspects or resolving conflicts originating from power, authority, or other social relations.

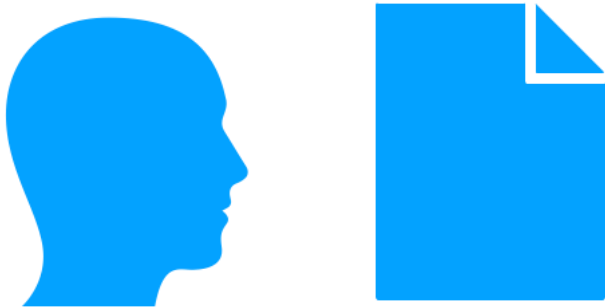
Domain Knowledge

Guidelines

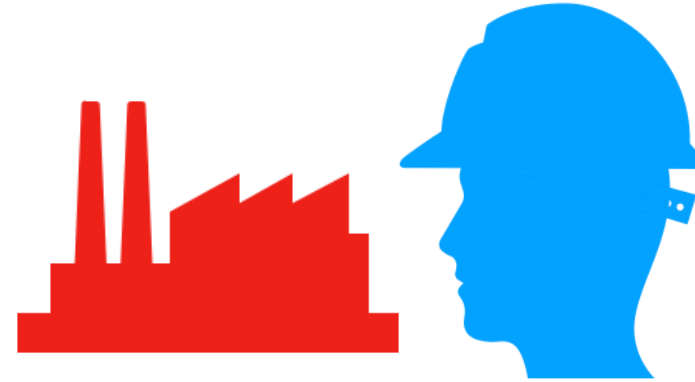
Data Quality

Lack of Benchmarks

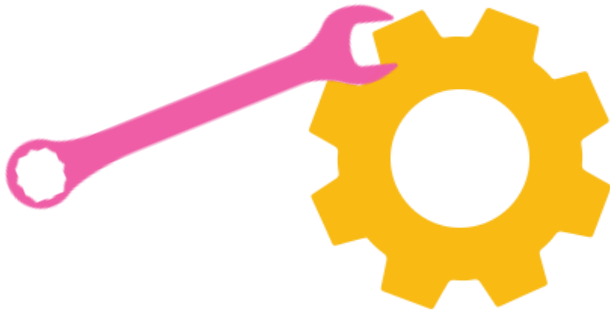
What do We Need?



Clear papers, Clear Communication



Involvement of companies and experts



Tool maintenance



Reward for open science and replication

Thanks to...



Sallam Abualhaija



Fatma Basak Aydemir



Fabiano Dalpiaz



Davide Dell'Anna



Xavier Franch



Davide Fucci

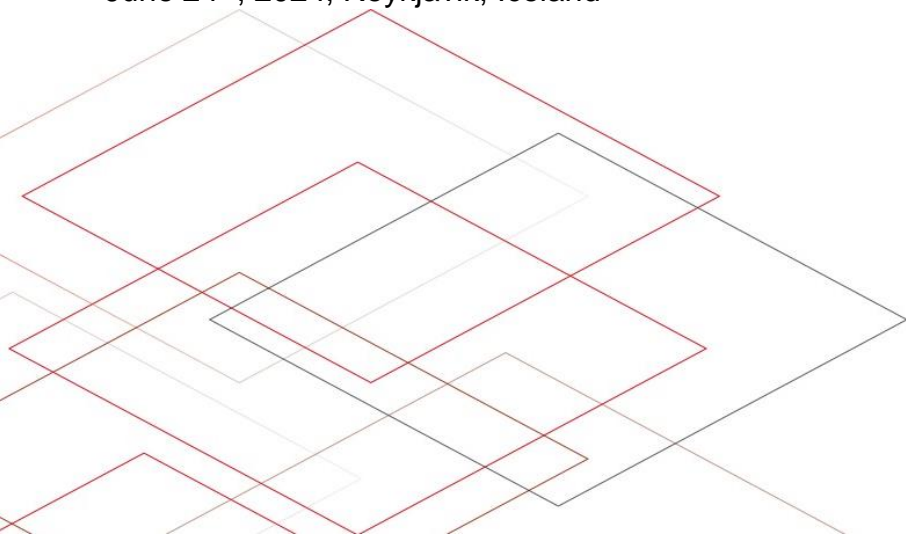
Introduction: Open Research Knowledge Graph

Oliver Karras, Alessio Ferrari, Davide Fucci, and Davide Dell'Anna

oliver.karras@tib.eu, alessio.ferrari@isti.cnr.it, davide.fucci@bth.se, d.dellanna@uu.nl

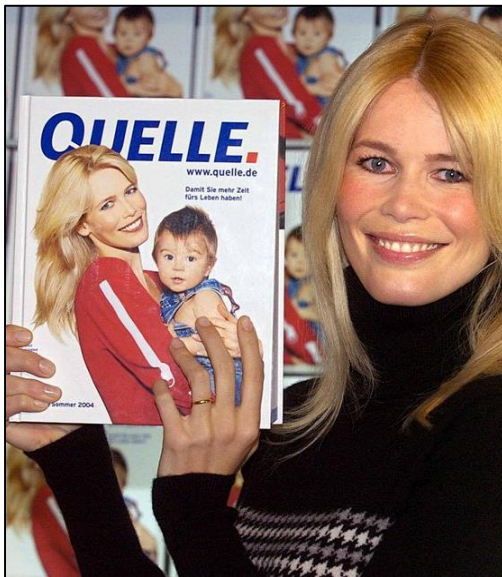
32nd IEEE International Requirements Engineering 2024 Conference – Exploring New Horizons: Expanding the Frontiers of Requirements Engineering

June 24th, 2024, Reykjavik, Iceland

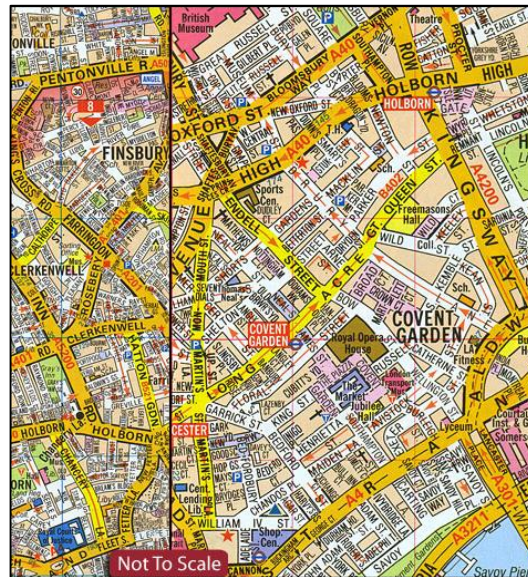


Once Upon a Time, we Communicated with Paper...

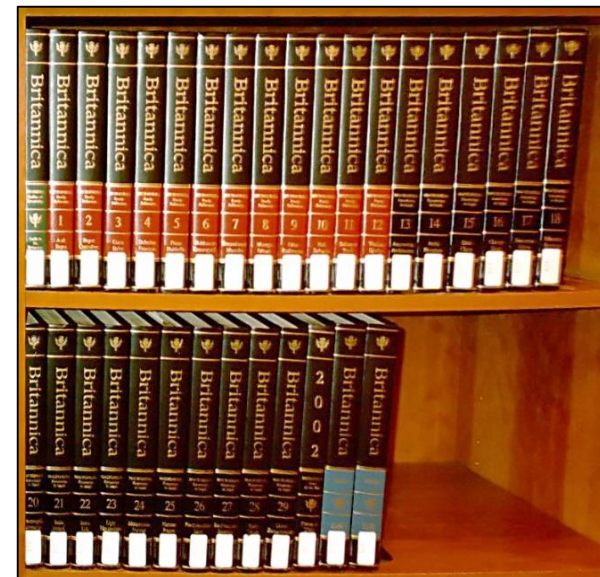
Who still remembers?



Mail order catalogs



Maps

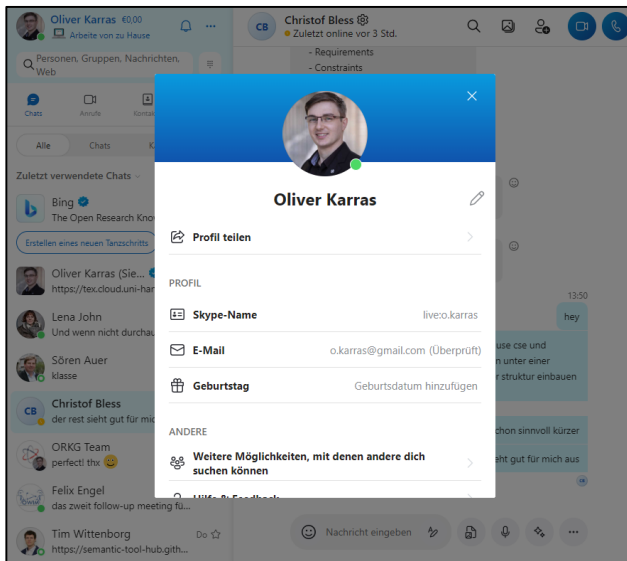
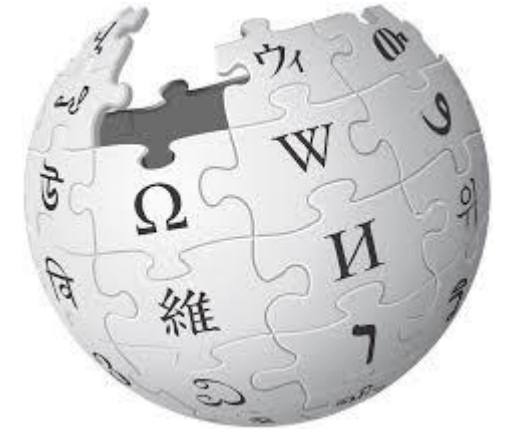
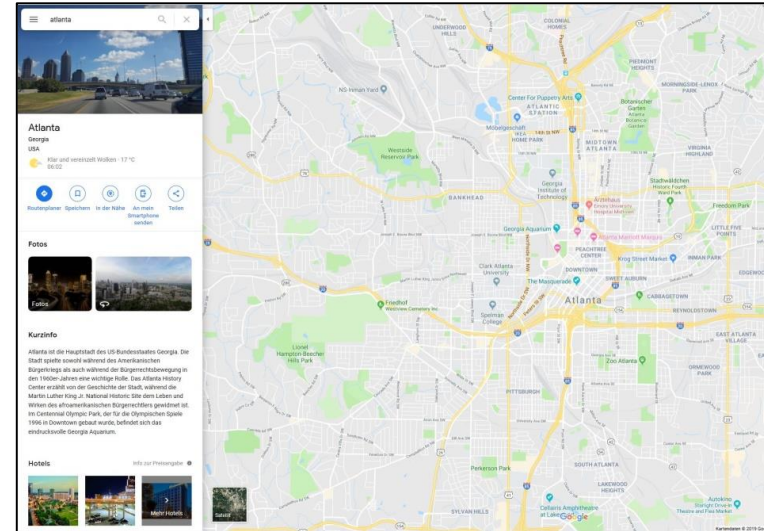
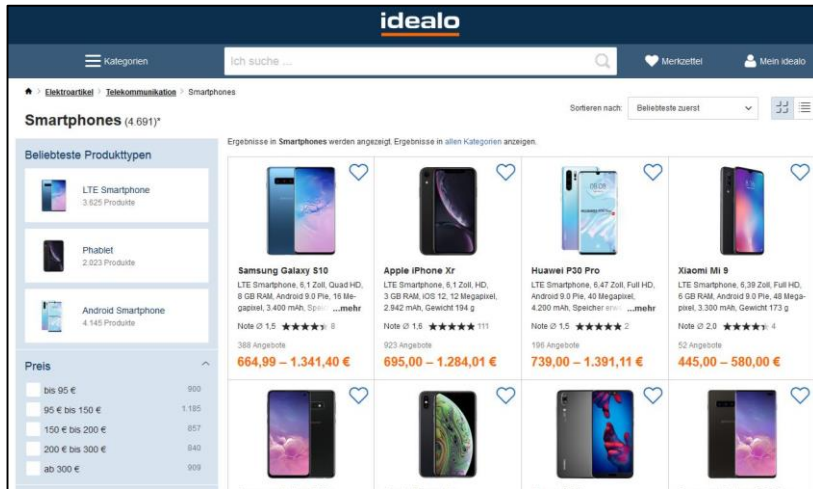


Encyclopedia



Phone books

... until Digital Transformation (Digitalization)!



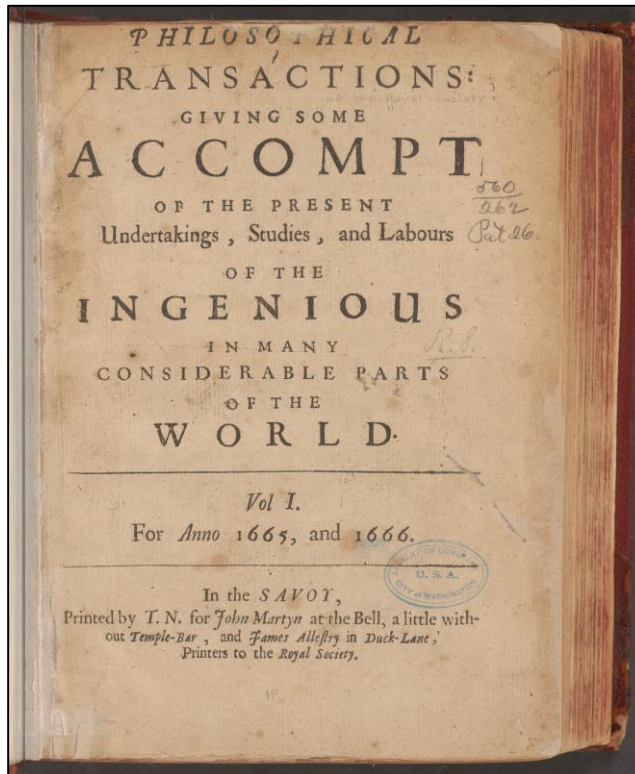
The World of Publishing & Communication has Profoundly Changed!

- **New means adapted to the new possibilities**, e.g., platforms
- **Completely new business models**
- **More focus on data, interlinking, services, and search**
- **Integration, crowdsourcing, and data curation** are important

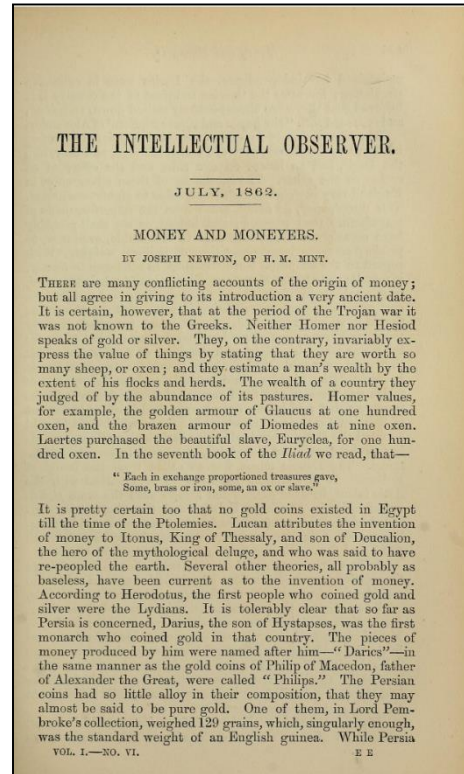
What has happened in academia in terms of scholarly publishing & communication?

Let's Take a Look

17th century



19th century



20th century

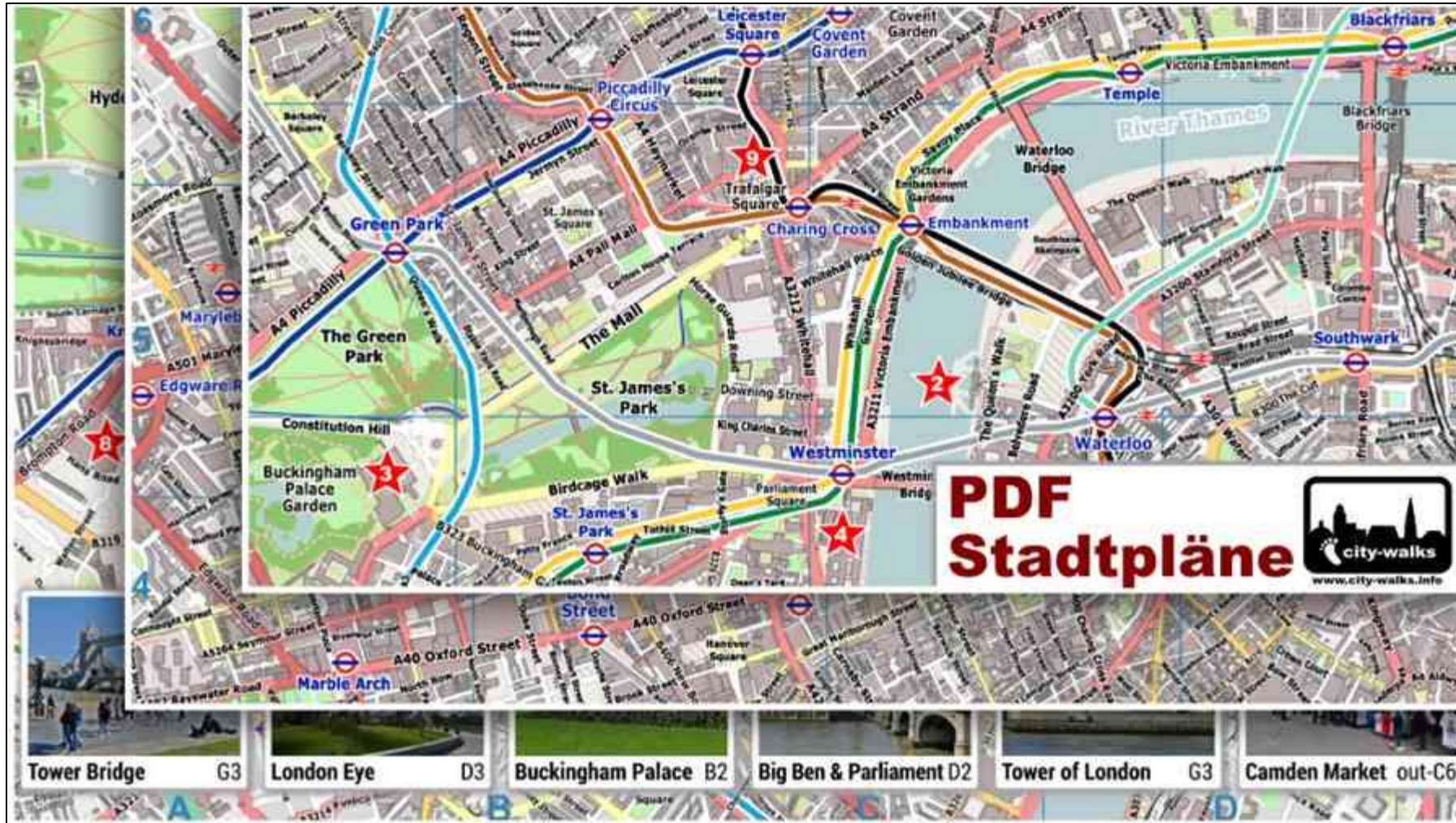


21st century



Scholarly publishing & communication has **not changed** (much)!

Let's Take a Look



21st century

AGDISTIS - Graph-Based Disambiguation of Named Entities Using Linked Data

Ricardo Usbeck^{1,2}, Axel-Cyrille Ngonga Ngomo¹, Michael Röder^{1,2},
Daniel Gerber¹, Sandro Athaide Coelho³, Sören Auer⁴, and Andreas Both²

¹ University of Leipzig, Germany

² R&D, Unister GmbH, Germany

³ Federal University of Juiz de Fora, Brazil

⁴ University of Bonn & Fraunhofer IAS, Germany
(usbeck,ngonga}@informatik.uni-leipzig.de

Abstract. Over the last decades, several billion Web pages have been made available on the Web. The ongoing transition from the current Web of unstructured data to the Web of Data yet requires scalable and accurate approaches for the extraction of structured data in RDF (Resource Description Framework) from these websites. One of the key steps towards extracting RDF from text is the disambiguation of named entities. While several approaches aim to tackle this problem, they still achieve poor accuracy. We address this drawback by presenting AGDISTIS, a novel knowledge-base-agnostic approach for named entity disambiguation. Our approach combines the Hypertext-Induced Topic Search (HITS) algorithm with label expansion strategies and string similarity measures. Based on this combination, AGDISTIS can efficiently detect the correct URIs for a given set of named entities within an input text. We evaluate our approach on eight different datasets against state-of-the-art named entity disambiguation frameworks. Our results indicate that we outperform the state-of-the-art approach by up to 29% F-measure.

1 Introduction

The vision behind the Web of Data is to provide a new machine-readable layer to the Web where the content of Web pages is annotated with structured data (e.g., RDFa [1]). However, the Web in its current form is made up of at least 15 billion Web pages.¹ Most of these websites are unstructured in nature. Realizing the vision of a usable and up-to-date Web of Data thus requires scalable and accurate natural-language-processing approaches that allow extracting RDF from such unstructured data. Three tasks play a central role when extracting RDF from unstructured data: named entity recognition (NER), named entity disambiguation (NED), also known as entity linking [16], and relation extraction (RE). For the first sentence of Example 1, an accurate named entity recognition approach would return the strings Barack Obama and Washington,

¹ Data gathered from <http://www.worldwidewebsize.com/> on January 4th, 2014.

21st century



Rethink How Scientific Knowledge is Communicated



*“The lightbulb was **not** invented by improving the candle.”*

Oren Harari

Digitalization is **more** than just Digitization!

Current and future scientific challenges can not be tackled with an outdated communication system.

**Digitalize Knowledge,
Not Documents!**

Example: Requirements Engineering and Empirical Research

The screenshot shows the Google Scholar interface. At the top, the search bar contains the text "requirements engineering and empirical research" with a magnifying glass icon to the right. Below the search bar, a blue box highlights the text "Ungefähr 3.970.000 Ergebnisse (0,16 Sek.)". On the left side, there are several filters: "Artikel" (with a blue diamond icon), "Beliebige Zeit" (with a dropdown menu showing "Seit 2024", "Seit 2023", "Seit 2020", and "Zeitraum wählen..."), "Nach Relevanz sortieren" (with a red arrow icon), "Nach Datum sortieren", "Beliebige Sprache" (with a dropdown menu showing "Seiten auf Deutsch"), and "Alle Typen" (with a dropdown menu showing "Übersichtsarbeiten"). At the bottom left, there are two checkboxes: "Patente einschließen" (unchecked) and "Zitate einschließen" (checked). The search results are displayed in a list. The first result is titled "[HTML] Empirical research in requirements engineering: trends and opportunities" and is from "springer.com". The second result is titled "[HTML] Empirical research methodologies and studies in Requirements Engineering: How far did we come?" and is from "sciencedirect.com". The third result is titled "A systematic literature review of empirical research on quality requirements" and is from "springer.com". Each result includes the authors, the journal name, the year, and a brief abstract.

Google Scholar

requirements engineering and empirical research

Ungefähr 3.970.000 Ergebnisse (0,16 Sek.)

Artikel

Beliebige Zeit
Seit 2024
Seit 2023
Seit 2020
Zeitraum wählen...

Nach Relevanz sortieren
Nach Datum sortieren

Beliebige Sprache
Seiten auf Deutsch

Alle Typen
Übersichtsarbeiten

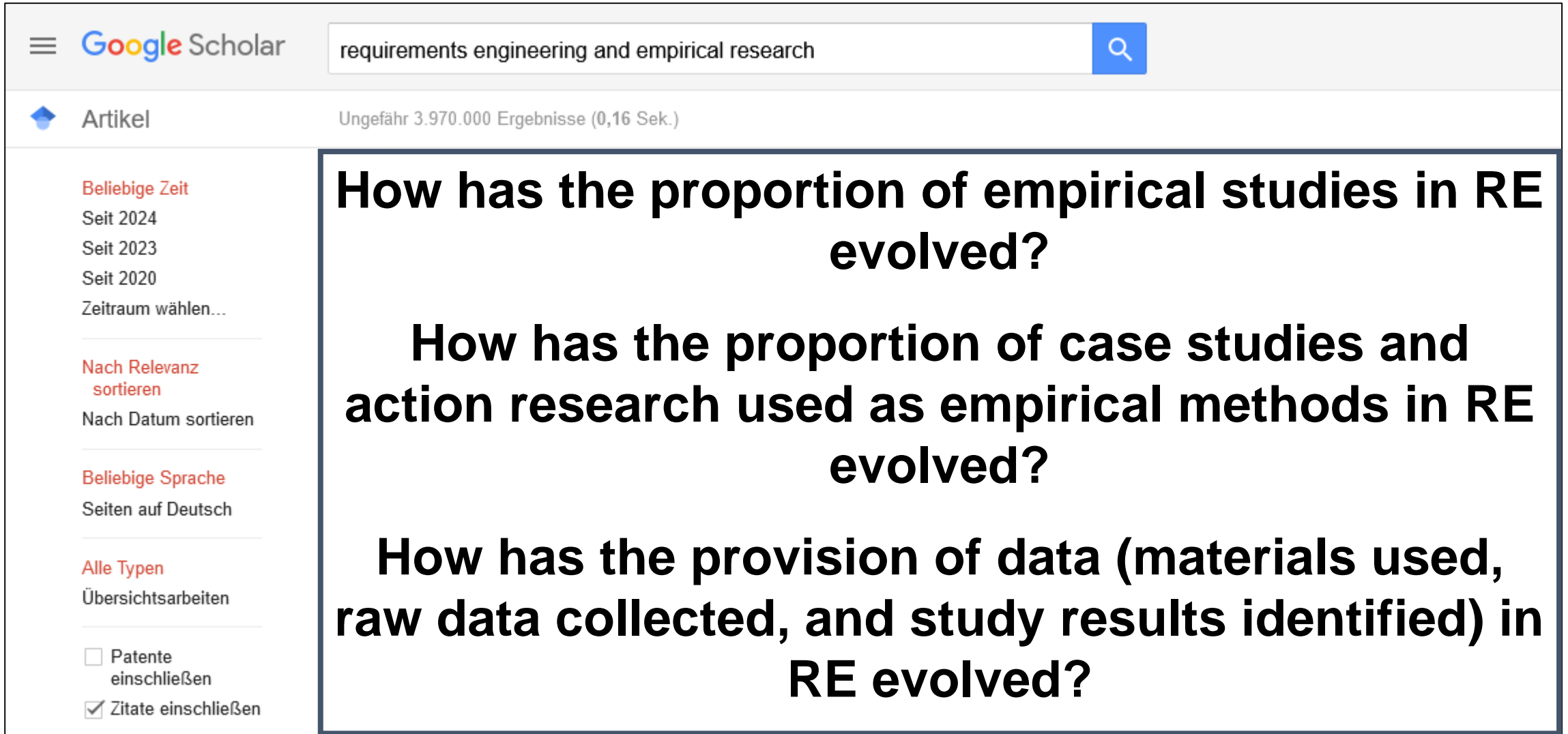
☐ Patente einschließen
☒ Zitate einschließen

[HTML] Empirical research in requirements engineering: trends and opportunities [HTML] springer.com
T Ambreen, N Ikram, M Usman, M Niazi - Requirements Engineering, 2018 - Springer
... trends and future **research** directions. To represent a state-of-the-art of **requirements engineering**, along with various trends and opportunities of **empirical RE research**, we conducted a ...
☆ Speichern 99 Zitieren Zitiert von: 125 Ähnliche Artikel Alle 7 Versionen

[HTML] Empirical research methodologies and studies in Requirements Engineering: How far did we come? [HTML] sciencedirect.com
M Daneva, D Damian, A Marchetto, O Pastor - Journal of systems and ..., 2014 - Elsevier
... Software **Engineering** paradigm. We summarize prior **empirical research** in RE and introduce the contributors to this special issue on **empirical research** methodologies and studies in ...
☆ Speichern 99 Zitieren Zitiert von: 67 Ähnliche Artikel Alle 12 Versionen

A systematic literature review of empirical research on quality requirements [PDF] springer.com
T Olsson, S Sentilles, E Papatheocharous - Requirements Engineering, 2022 - Springer
... We want to understand the **empirical research** on quality **requirements** topics as well as ...
We conclude that more **research** is needed as **empirical research** on quality **requirements** is not ...
☆ Speichern 99 Zitieren Zitiert von: 13 Ähnliche Artikel Alle 9 Versionen

Example: Requirements Engineering and Empirical Research



The image shows a screenshot of a Google Scholar search results page. The search query is "requirements engineering and empirical research". The results are filtered by "Artikel" (Articles) and show approximately 3,970,000 results in 0.16 seconds. The left sidebar contains filters for time (Beliebige Zeit, Seit 2024, Seit 2023, Seit 2020, Zeitraum wählen...), relevance (Nach Relevanz sortieren, Nach Datum sortieren), language (Beliebige Sprache, Seiten auf Deutsch), and type (Alle Typen, Übersichtsarbeiten). There are also checkboxes for "Patente einschließen" and "Zitate einschließen". The main content area displays three research questions in large, bold, black text, each on a new line.

Google Scholar requirements engineering and empirical research

Artikel Ungefähr 3.970.000 Ergebnisse (0,16 Sek.)

How has the proportion of empirical studies in RE evolved?

How has the proportion of case studies and action research used as empirical methods in RE evolved?

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

Beliebige Zeit
Seit 2024
Seit 2023
Seit 2020
Zeitraum wählen...

Nach Relevanz sortieren
Nach Datum sortieren

Beliebige Sprache
Seiten auf Deutsch

Alle Typen
Übersichtsarbeiten

☐ Patente einschließen
☒ Zitate einschließen

How do We Answer These Questions so far?



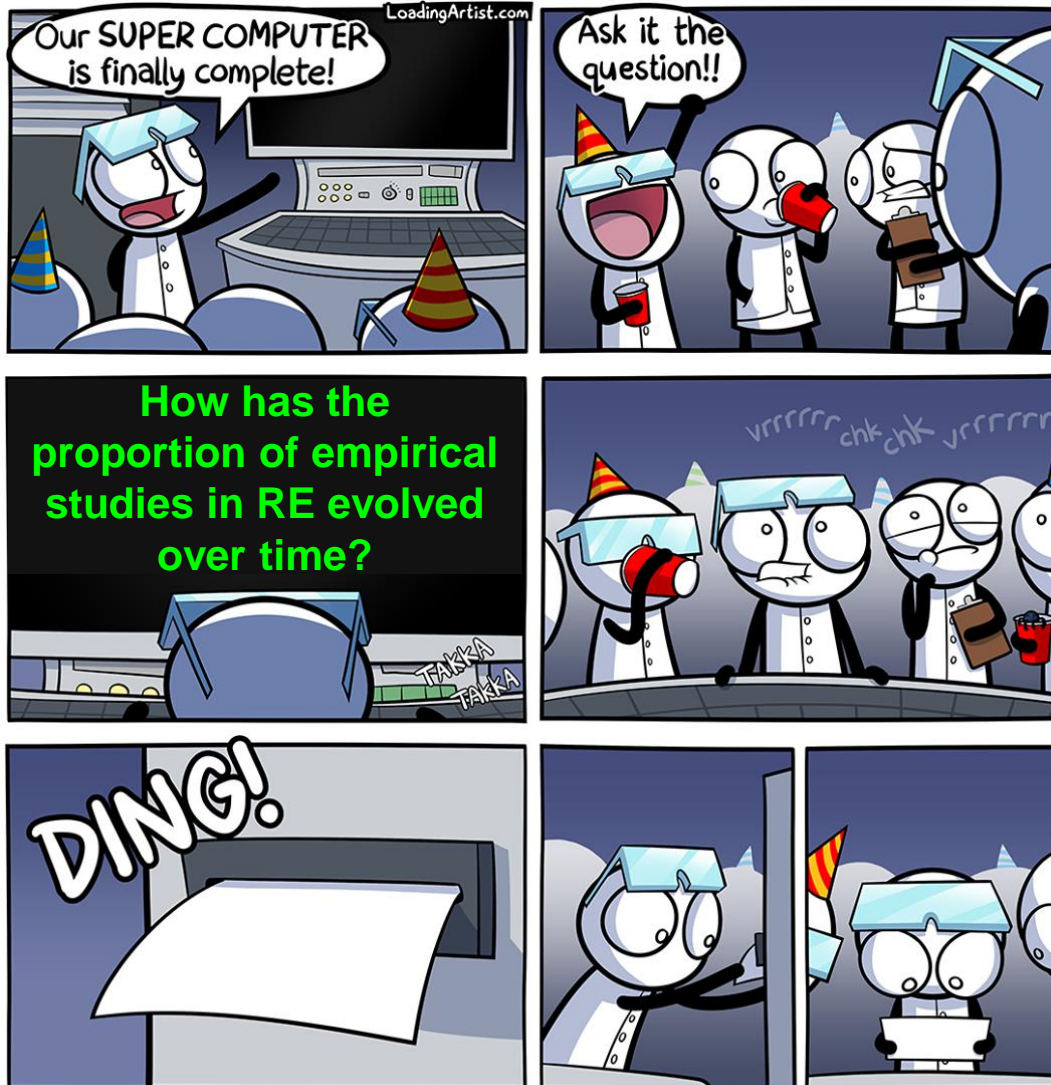
(S)LRs and SMSs for a **comprehensive, up-to-date,**
and **long-term available** overview.



Over **7 million publications** per year^[4]
with an **increasing** (exponential) tendency.

[4] Fire and Guestrin: *Over-Optimization of Academic Publishing Metrics: Observing Goodhart's Law in Action*. GigaScience Vol. 8, No. 6., DOI: [10.1093/gigascience/giz053](https://doi.org/10.1093/gigascience/giz053), 2019.

Wouldn't it be Great if we Could Ask the Computer?



How can we achieve this goal?

Open Research Knowledge Graph (ORKG)

The screenshot displays the ORKG website interface. At the top, there is a navigation bar with the ORKG logo, links for View, Tools, and About, a dropdown menu for NFDI4DataScience, a search bar, and buttons for '+ Add new' and 'Sign in'. Below the navigation bar, a large banner features the text 'Scholarly Knowledge. FAIR.' and a description of the ORKG's purpose: 'The Open Research Knowledge Graph (ORKG) aims to describe research papers in a structured manner. With the ORKG, papers are easier to find and compare.' A 'Play video' button is also present. The main content area is titled 'Browse by research field' and includes a search bar for fields. Below this, five red buttons represent different research fields: Arts and Humanities (432 papers - 33 comparisons), Engineering (3391 papers - 345 comparisons), Life Sciences (4142 papers - 195 comparisons), Physical Sciences & Mathematics (15694 papers - 731 comparisons), and Social and Behavioral Sciences (844 papers - 170 comparisons). The 'Comparisons' tab is selected, showing a list of research papers. The first entry is 'Systematic Literature Review (SLR) Tools analysed based on General Features' in the 'Information Science' field, dated 04-03-2024. The second entry is 'RO estimates for infectious diseases' in the 'Virology' field, dated 04-04-2024. The third entry is 'Machine learning of pre-harvesting crop/fruit parameters to minimize overall losses in farming production' in the 'Plant Cultivation, PL...' field. On the right side, there are three sections: 'Load Tools' with a note about the Mastodon widget, 'ORKG stories' with a link to 'Find out more', and 'Join ORKG!' with a 'Sign up' button. At the bottom right, there is a section titled 'Knowledge base for science' with a note about the increasing volume of research publications.

ORKG

View Tools About NFDI4DataScience

Search... + Add new Sign in

Scholarly Knowledge. FAIR.

The Open Research Knowledge Graph (ORKG) aims to describe research papers in a structured manner. With the ORKG, papers are easier to find and compare. [Play video](#)

Browse by research field

Search for fields...

Arts and Humanities
432 papers - 33 comparisons

Engineering
3391 papers - 345 comparisons

Life Sciences
4142 papers - 195 comparisons

Physical Sciences & Mathematics
15694 papers - 731 comparisons

Social and Behavioral Sciences
844 papers - 170 comparisons

Comparisons Papers Visualizations Reviews Lists

Top recent

Systematic Literature Review (SLR) Tools analysed based on General Features
16 Contributions 0 Visualizations 04-03-2024
The non-AI characteristics of SLR tools analyzed based on 23 features in the paper titled "Artificial Intelligence for Literature Reviews: Opportunities and Challenges"

RO estimates for infectious diseases
6 Contributions 0 Visualizations 04-04-2024
RO estimates for infectious diseases

Machine learning of pre-harvesting crop/fruit parameters to minimize overall losses in farming production

Load Tools By loading the Mastodon widget, you agree with the [cookie guidelines](#)

ORKG stories
See how researchers benefit from using ORKG.
[Find out more](#)

Join ORKG! [Sign up](#)

Knowledge base for science
We are flooded with new publications in research every day and it is increasingly challenging to keep up

The ORKG **revolutionizes** how **scientific knowledge** is communicated and processed, making it **actionable** for **machines** and **researchers** to navigate, compare, and review **vast amounts of information** efficiently.

Research Knowledge Graphs

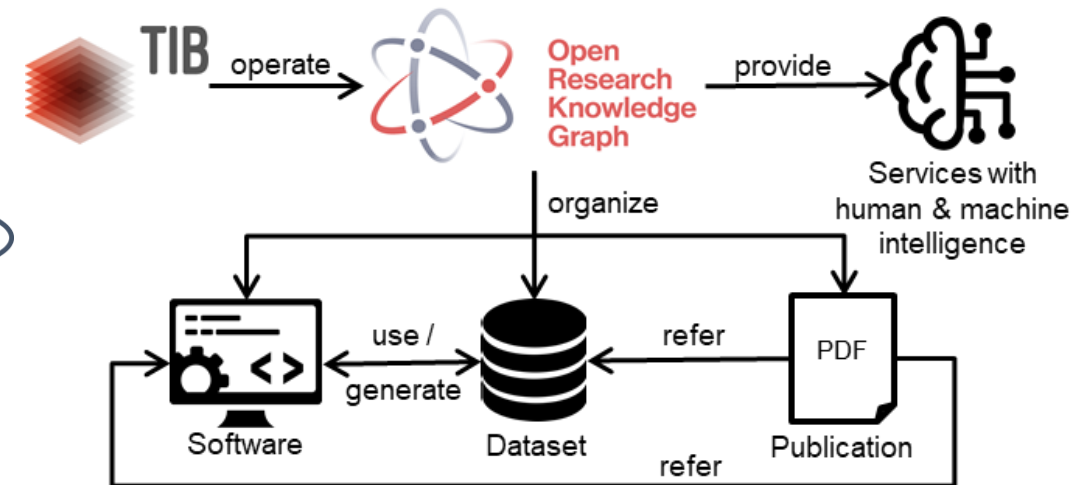
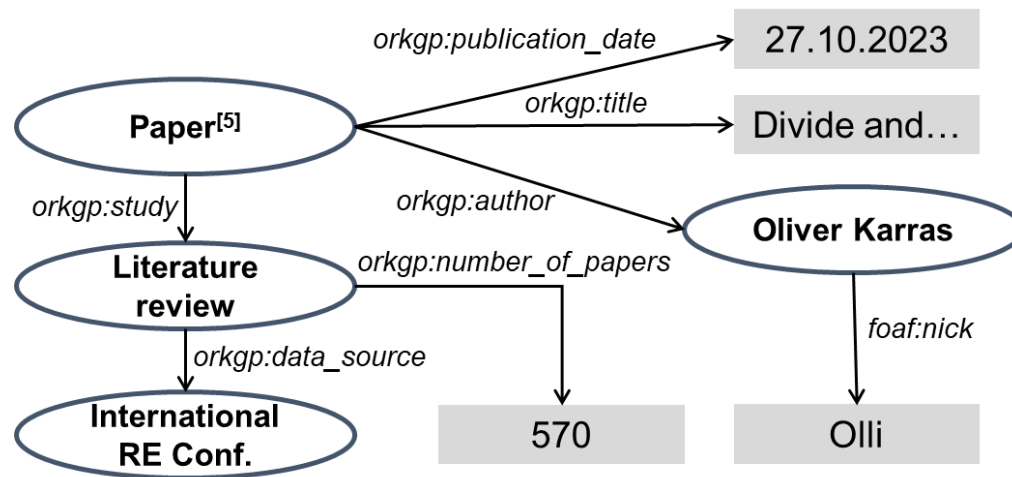
Research Knowledge Graphs (RKGs) are a **technology** for organizing scientific (meta-)data in a

- **Flexible, fine-grained, and semantic** representation
- That is **understandable** and **processable** by humans and machines

Research Knowledge Graphs

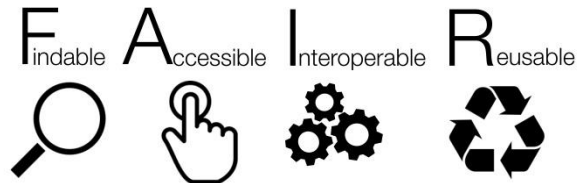
Research Knowledge Graphs (RKGs) are a **technology** for organizing scientific (meta-)data in a

- **Flexible, fine-grained, and semantic** representation
- That is **understandable** and **processable** by humans and machines



The **ORKG** is a **ready-to-use** and **sustainably** operated **infrastructure** with **services** that uses a **cross-discipline RKG** for the **long-term** and **openly available organization** of scientific (meta-)data according to the **FAIR** data principles.

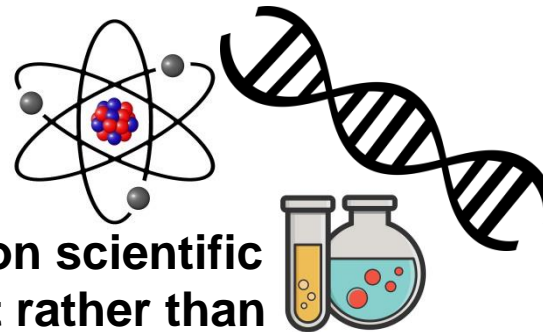
ORKG Objectives



Make research FAIR



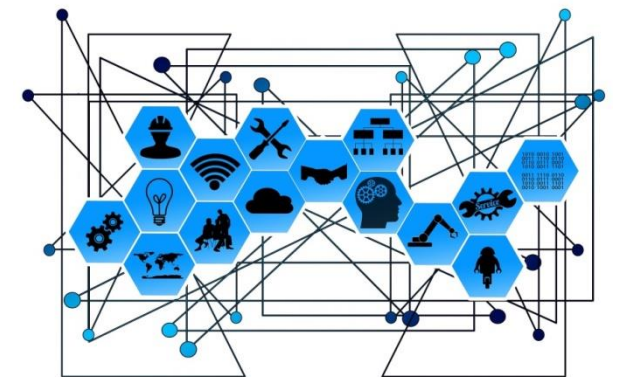
**Provide long-term
available and up-to-date
overviews**



**Focus on scientific
content rather than
pure metadata**



Foster collaboration



**Tackle interdisciplinary
challenges**

Classification category	Feature request
Has dataset	seel.cse.lsu.edu/data/refsq17.zip
Has feature	Bag of words
Has method	Support vector machines
Has result	P = 0.45, R = 0.68, F1 = 0.54, etc.

Behind the Scenes

Mining User Requirements from Application Store Reviews Using Frame Semantics

Nishant Jha and Anas Mahmoud()

The Division of Computer Science and Engineering, Louisiana State University,
Baton Rouge, LA 70803, USA
njha1@lsu.edu, mahmoud@csc.lsu.edu

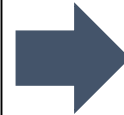
Abstract. *Context and motivation:* Research on mining user reviews in mobile application (app) stores has noticeably advanced in the past few years. The majority of the proposed techniques rely on classifying the textual description of user reviews into different categories of technically informative user requirements and uninformative feedback. *Question/Problem:* Relying on the textual attributes of reviews often produces high dimensional models. This increases the complexity of the classifier and can lead to overfitting problems. *Principal ideas/results:* We propose a novel semantic approach for app review classification. The proposed approach is based on the notion of semantic role labeling, or characterizing the lexical meaning of text in terms of semantic frames. Semantic frames help to generalize from text (individual words) to more abstract scenarios (contexts). This reduces the dimensionality of the data and enhances the predictive capabilities of the classifier. Three datasets of user reviews are used to conduct our experimental analysis. Results show that semantic frames can be used to generate lower dimensional and more accurate models in comparison to text classification methods. *Contribution:* A novel semantic approach for extracting user requirements from app reviews. The proposed approach enables a more efficient classification process and reduces the chance of overfitting.

Keywords: Requirements elicitation · Application stores · Classification

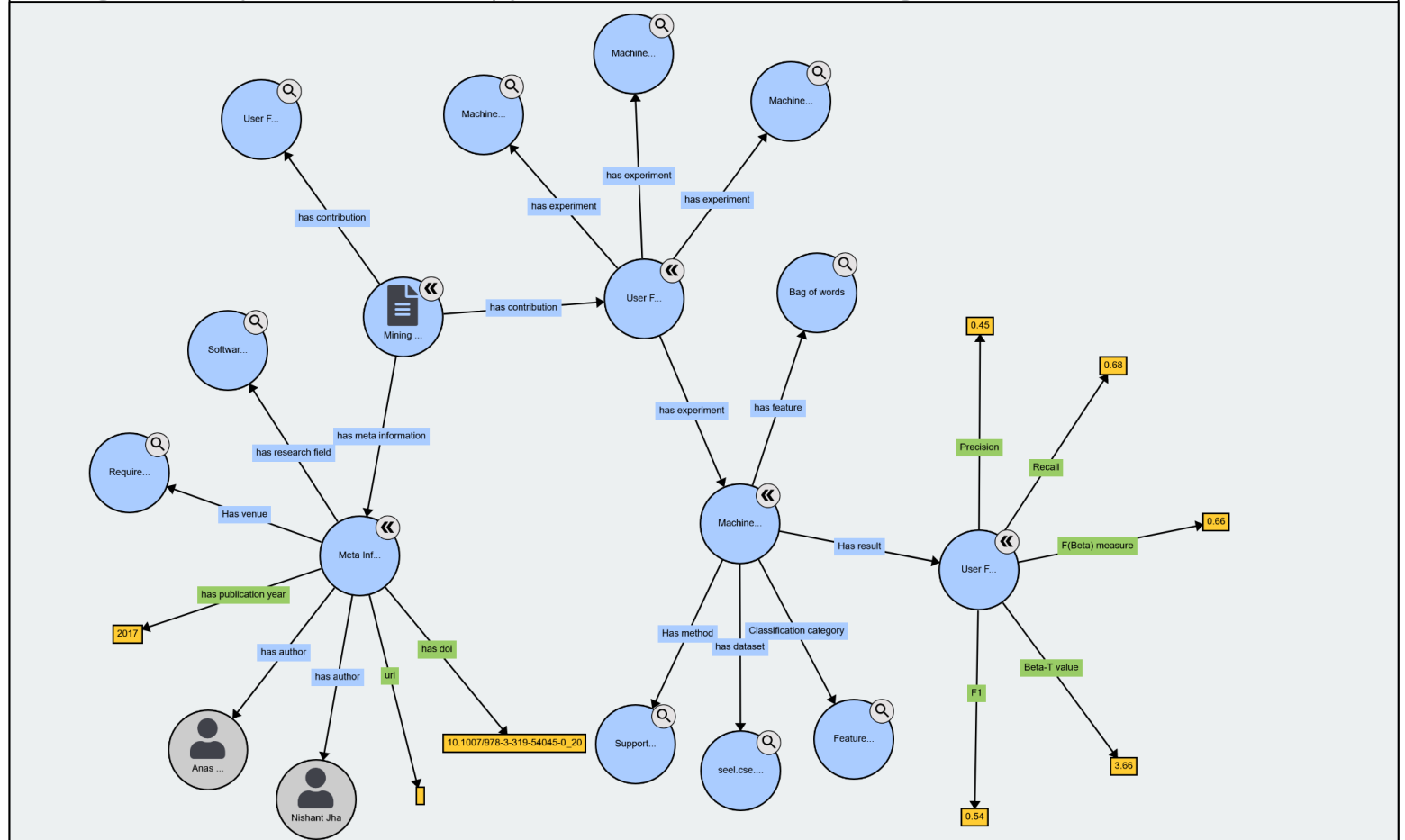
1 Introduction

Mobile application markets, or app stores (e.g., Google Play and Apple App Store), represent a unique model of service-oriented business. Such platforms have created an unprecedented opportunity for app developers to directly monitor the opinions of a large population of end-users of their software [25]. Through app stores feedback services, app users can directly share their experience in the form of textual reviews and meta-data (e.g., star ratings). Analyzing large datasets of app store reviews has revealed that they contain substantial amounts of up-to-date technical information. Such information can be leveraged by app developers to help them maintain and sustain their apps in a highly-competitive

© Springer International Publishing AG 2017
P. Grünbacher and A. Perini (Eds.): REFSQ 2017, LNCS 10153, pp. 273–287, 2017.
DOI: 10.1007/978-3-319-54045-0_20



Mining User Requirements from Application Store Reviews Using Frame Semantics



Using FAIR Scientific Knowledge

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

Mining User Requirements from Application Store Reviews Using Frame Semantics

Nishant Jha and Anas Mahmoud^(✉)

The Division of Computer Science and Engineering, Louisiana State University,
Baton Rouge, LA 70803, USA
njha1@lsu.edu, mahmoud@csc.lsu.edu

Abstract. *Context and motivation:* Research on mining user reviews in mobile application (app) stores has noticeably advanced in the past few years. The majority of the proposed techniques rely on classifying the textual description of user reviews into different categories of technically informative user requirements and uninformative feedback. *Question/Problem:* Relying on the textual attributes of reviews often produces high dimensional models. This increases the complexity of the classifier and can lead to overfitting problems. *Principal ideas/results:* We propose a novel semantic approach for app review classification. The proposed approach is based on the notion of semantic role labeling, or characterizing the lexical meaning of text in terms of semantic frames. Semantic frames help to generalize from text (individual words) to more abstract scenarios (contexts). This reduces the dimensionality of the data and enhances the predictive capabilities of the classifier. Three datasets of user reviews are used to conduct our experimental analysis. Results show that semantic frames can be used to generate lower dimensional and more accurate models in comparison to text classification methods. *Contribution:* A novel semantic approach for extracting user requirements from app reviews. The proposed approach enables a more efficient classification process and reduces the chance of overfitting.

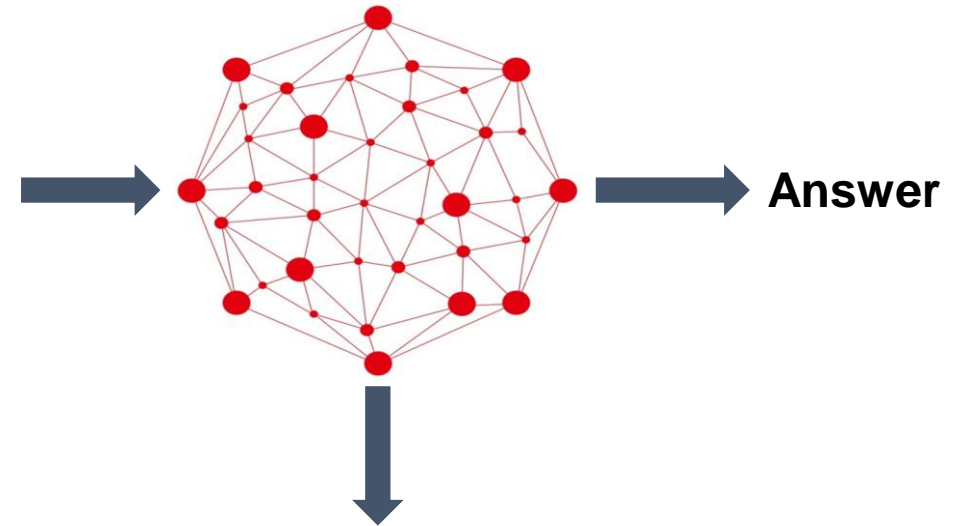
Keywords: Requirements elicitation · Application stores · Classification

1 Introduction

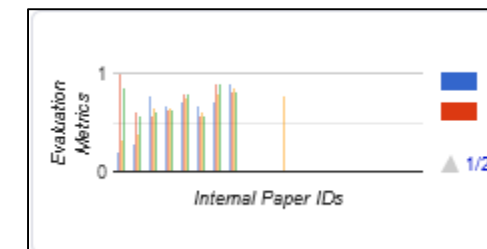
Mobile application markets, or app stores (e.g., Google Play and Apple App Store), represent a unique model of service-oriented business. Such platforms have created an unprecedented opportunity for app developers to directly monitor the opinions of a large population of end-users of their software [25]. Through app stores feedback services, app users can directly share their experience in the form of textual reviews and meta-data (e.g., star ratings). Analyzing large datasets of app store reviews has revealed that they contain substantial amounts of up-to-date technical information. Such information can be leveraged by app developers to help them maintain and sustain their apps in a highly-competitive

© Springer International Publishing AG 2017
P. Grünbacher and A. Perini (Eds.): REFSQ 2017, LNCS 10153, pp. 273–287, 2017.
DOI: 10.1007/978-3-319-54045-0_20

Natural
language
question



Properties	Software Feature Request Detection in Issue Tracking Systems	Mining User Requirements from Application Store Reviews Using Frame Semantics
	2016 - User Feedback Classification	2017 - User Feedback Classification
Has dataset	https://zenodo.org/record/56907#.YKT_NudCRPY	https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip
		https://sites.google.com



ORKG Comparisons

Acknowledgement
of creators

DOI

Visualizations

Interactive filtering

Overview of Approaches that Classify User Feedback as Feature Request ☆👁

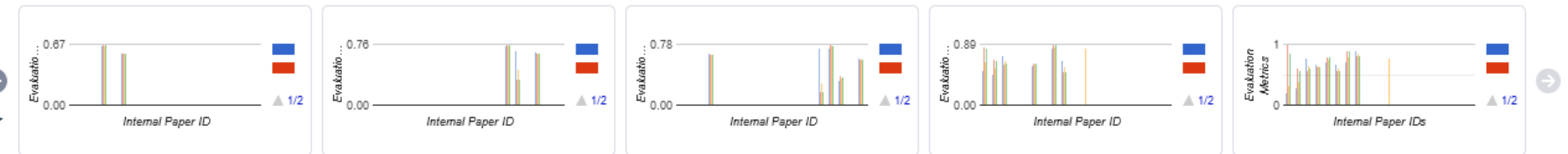
📅 June 2021

👤 Oliver Karras

👤 Eduard C. Groen

This overview shows the classification results of approaches that use the machine learning algorithms Naïve Bayes, Support Vector Machines, and Decision Trees C4.5 in combination with the machine learning features Bag of Words or Term Frequency - Inverse Document Frequency to classify user feedback as feature request.

DOI: <https://doi.org/10.48366/r112387>



Properties

Software Feature Request
Detection in Issue Tracking
Systems

User Feedback Classification - 2016

Mining User Requirements from
Application Store Reviews Using
Frame Semantics

User Feedback Classification - 2017

Mining Twitter Feeds for Software
User Requirements

User Feedback Classification - 2017

Automatic Classification of Non-
Functional Requirements from
Augmented App User Reviews

User Feedback Classification - 2017

Bug reports
simply p

classifying

User Feedback

has dataset

https://zenodo.org/record/56907#.YKT_NudCRPY

https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip

<https://sites.google.com/site/appuserreviews/>
seel.cse.lsu.edu/data/refsq17.zip

seel.cse.lsu.edu/data/re17.zip

Not available

https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip

ORKG Comparisons are Citable

Comparison of Studies on Germany's Energy Supply in 2050

November 2021

Felix Kullmann Jan Göpfert Oliver Karras Patrick Kuckertz Sören Auer Markus Stocker

Peter Markewitz Leander Kotzur Detlef Stolten

This comparison compares electricity generation studies for Germany in 2050. It lists studies from the German Energy Research Consortium (GERC) and the German Energy Research Association (DERA). The comparison is funded by the German Research Foundation (DFG) under the Special Collaborative Program (Sonderforschungsbereich) SFB 1151/B1.

DOI: <https://doi.org/10.26434/chemrxiv-2021-00000>

Comparison of studies on Germany's energy supply in 2050

Autoren Felix Kullmann, Peter Markewitz, Detlef Stolten, Oliver Karras, Patrick Kuckertz, Leander Kotzur, Jan-Maris Göpfert, Sören Auer, Markus Stocker

Publikationsdatum 2021

Ausgabe FZJ-2022-00782

Verlag Technoökonomische Systemanalyse

Beschreibung This comparison compiles the results from various studies analyzing a future low-carbon energy system for Germany. The focus of this study comparison is electricity generation. In the future, however, other essential characteristics of the respective energy system designs in the individual studies will be listed. Installed capacity is given in GW and electricity generation is given in TWh.

Zitate insgesamt **Zitiert von: 2**

Google Scholar Artikel [Comparison of studies on Germany's energy supply in 2050](#)
F Kullmann, P Markewitz, D Stolten, O Karras... - 2021
Zitiert von: 2 [Ähnliche Artikel](#)

Share

NFDI4Ing

Energy System Research

Facebook Twitter LinkedIn

Properties

[has energy sources](#)

Klimaneutralität

Contributions

Den Weg zu einem treibhausgasneutralen Deutschland ressourcenschonend gestalten

Contribution 1 - 2019

[all sources](#)

[bioenergy](#)

[geothermics](#)

[hydropower](#)

[import](#)

[net import](#)

net import

net import

net import

net import

net import

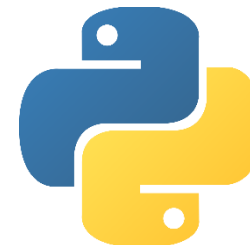
So far so Good, but...

...what can we do with machine-actionable scientific knowledge?

Anything we want!

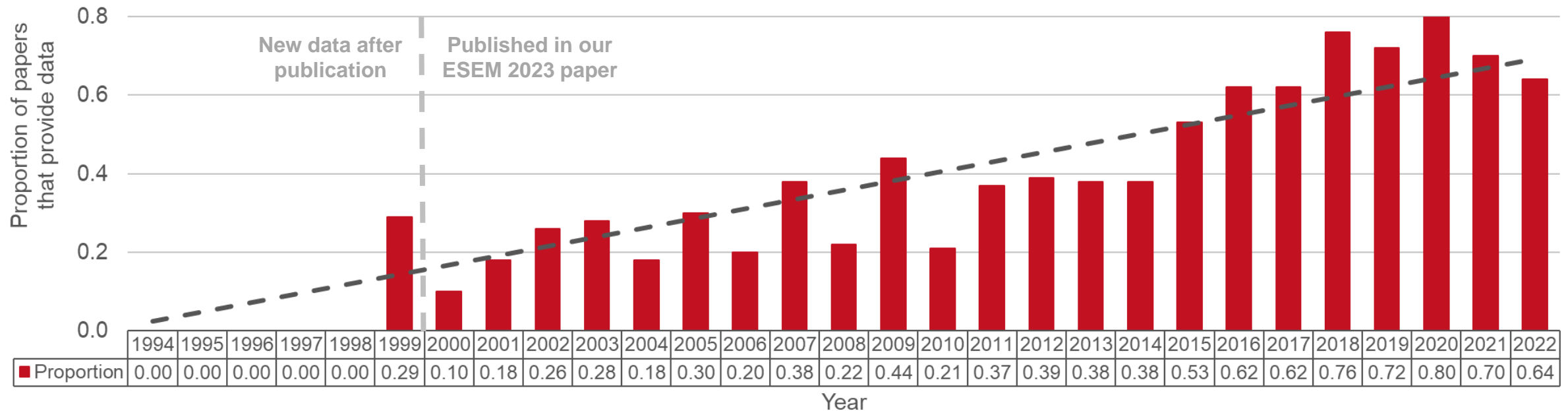
1. Papers, comparisons, and visualizations in the ORKG are **openly available for (re-)use and extension to anyone.**
2. The ORKG provides several **interfaces** for processing the data, e.g., to develop novel search, retrieval, mining, and assistance applications.

{ REST : API }



(Re-)use and Extension of Scientific Knowledge

How has the provision of data (materials used, raw data collected, and study results identified) in RE evolved?^[5, 6, 7]
(Based on 680 papers in the ORKG)



[5] Karras et al.: *Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering*. 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), DOI: [10.1109/ESEM56168.2023.10304795](https://doi.org/10.1109/ESEM56168.2023.10304795), 2023.

[6] Project on GitHub: <https://github.com/okarras/EmpiRE-Analysis>

[7] Interactive Jupyter notebook: <https://mybinder.org/v2/gh/okarras/EmpiRE-Analysis/HEAD?labpath=%2Fempire-analysis.ipynb>

ORKG Observatory

Empirical Software Engineering | Observatory

Edit

This observatory works to provide a community-maintainable knowledge graph of empirical research in software engineering. Our goal is to continuously acquire and curate comprehensive knowledge about empirical research applied in scientific publications in the research field of software engineering and its subfields, such as requirements engineering. In this way, we want to provide a comprehensive, up-to-date, and long-term overview of the state-of-the-art on empirical research in software engineering.

We are currently working on a knowledge graph of empirical research in requirements engineering.


For this purpose, we have developed a corresponding ORKG template (cf. <https://orkg.org/template/R186491>).

Contact: oliver.karras@tib.eu


Research problems

- empirical research in software engineering
- empirical research in requirements engineering

Organizations




Leibniz
Universität
Hannover




TIB
LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

Members



Oliver Karras
TIB - Leibniz Information Centre for Science and
Technology



Jil Kluender
Leibniz University Hannover

Content | 688 items

Show: ☒ Paper ☒ Comparison ☒ Visualization Top recent

- **Open** groups maintain topics in the ORKG
- **Central** access point for the community to all curated contents

Who organizes this scientific knowledge in the ORKG?

Who Organizes this Scientific Knowledge in the ORKG?

The heart of the ORKG is its **CROWD**:
Researchers from any discipline!



How can We Contribute to the ORKG?

1. While writing a paper:

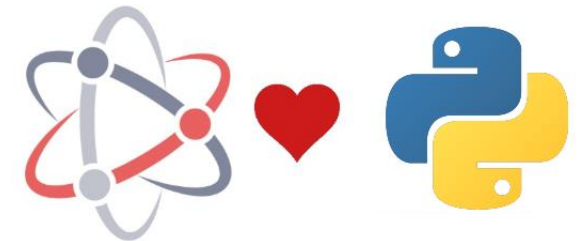
- SciKGTex: LaTeX package for FAIR annotations in a paper (embedded in the PDF) that can be imported into ORKG
- <https://github.com/Christof93/SciKGTex>



SciKGTex

2. While developing an analysis:

- Python package (& R package in development) for FAIR annotations in analysis scripts that can be imported into the ORKG
- <https://pypi.org/project/orkg/>



3. At any time:

- Manually using the ORKG Frontend to describe papers
- (Semi-)automatically using the ORKG REST API



Open
Research
Knowledge
Graph

{ REST : API }

What We Will Learn to Contribute to the ORKG

1. While writing a paper:

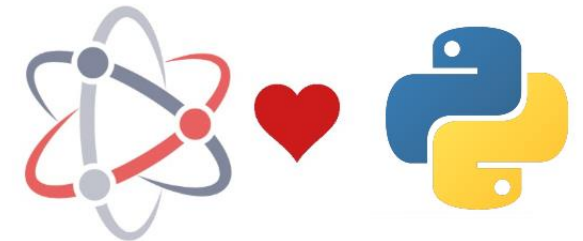
- **SciKGT_{EX}**: LaTeX package for FAIR annotations in a paper (embedded in the PDF) that can be imported into ORKG
- <https://github.com/Christof93/SciKGT_{EX}>



SciKGT_{EX}

2. While developing an analysis:

- Python package (& R package in development) for FAIR annotations in analysis scripts that can be imported into the ORKG
- <https://pypi.org/project/orkg/>



3. At any time:

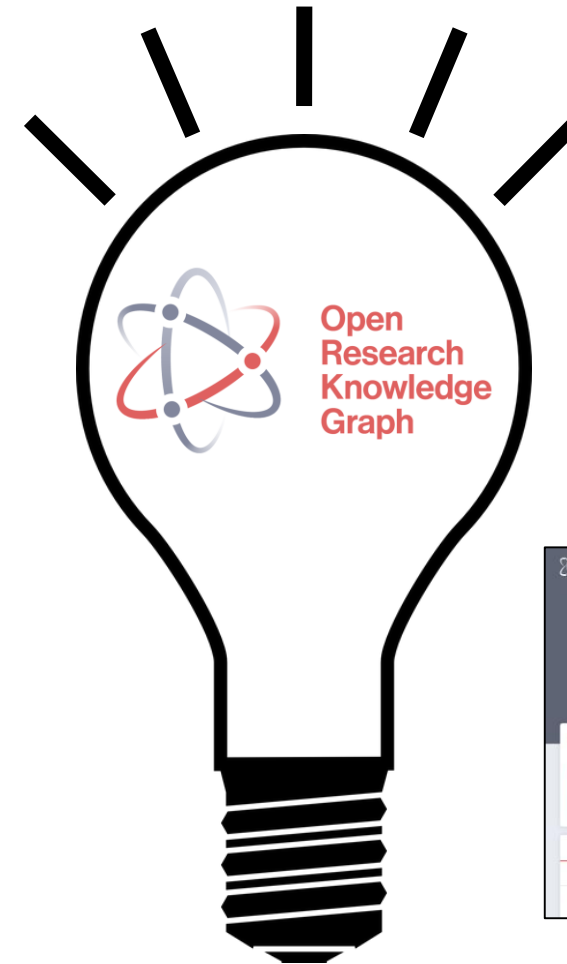
- **Manually** using the ORKG Frontend to describe papers
- (Semi-)automatically using the ORKG REST API



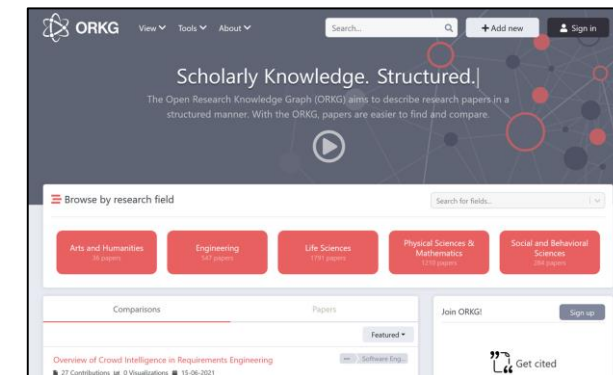
Open
Research
Knowledge
Graph

{ REST : API }

Conclusion



<https://orkg.org>



Let's bring scholarly communication and open science in Requirements Engineering to the 21st century!

Further Reading



Open Access ORKG Book

- Celebrating the 5th anniversary of ORKG
- A practical guide for new and advanced users
 - ORKG's terms and concepts
 - ORKG's approach
 - ORKG's technology
 - ORKG's success stories

<https://cuvillier.de/de/shop/publications/9037-open-research-knowledge-graph>



SciKGT_EX: A LaTeX Package for FAIR-Annotated Publications

Oliver Karras, Alessio Ferrari, Davide Fucci, and Davide Dell'Anna

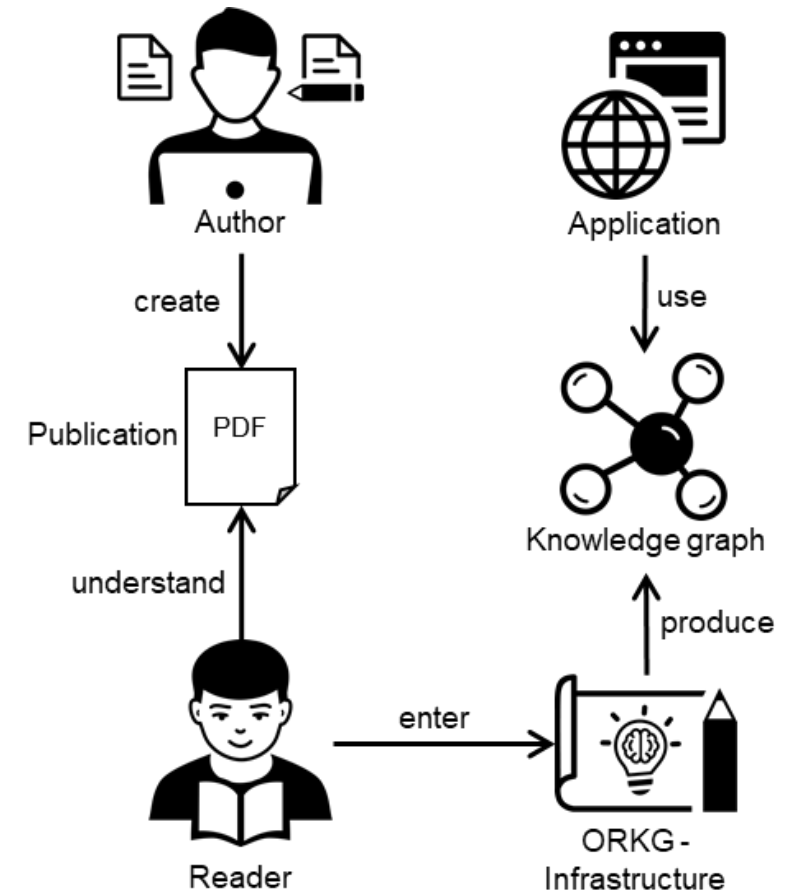
oliver.karras@tib.eu, alessio.ferrari@isti.cnr.it, davide.fucci@bth.se, d.dellanna@uu.nl

32nd IEEE International Requirements Engineering 2024 Conference – Exploring New Horizons: Expanding the Frontiers of Requirements Engineering

June 24th, 2024, Reykjavik, Iceland

Problem: Making Publications FAIR is a Downstream Task

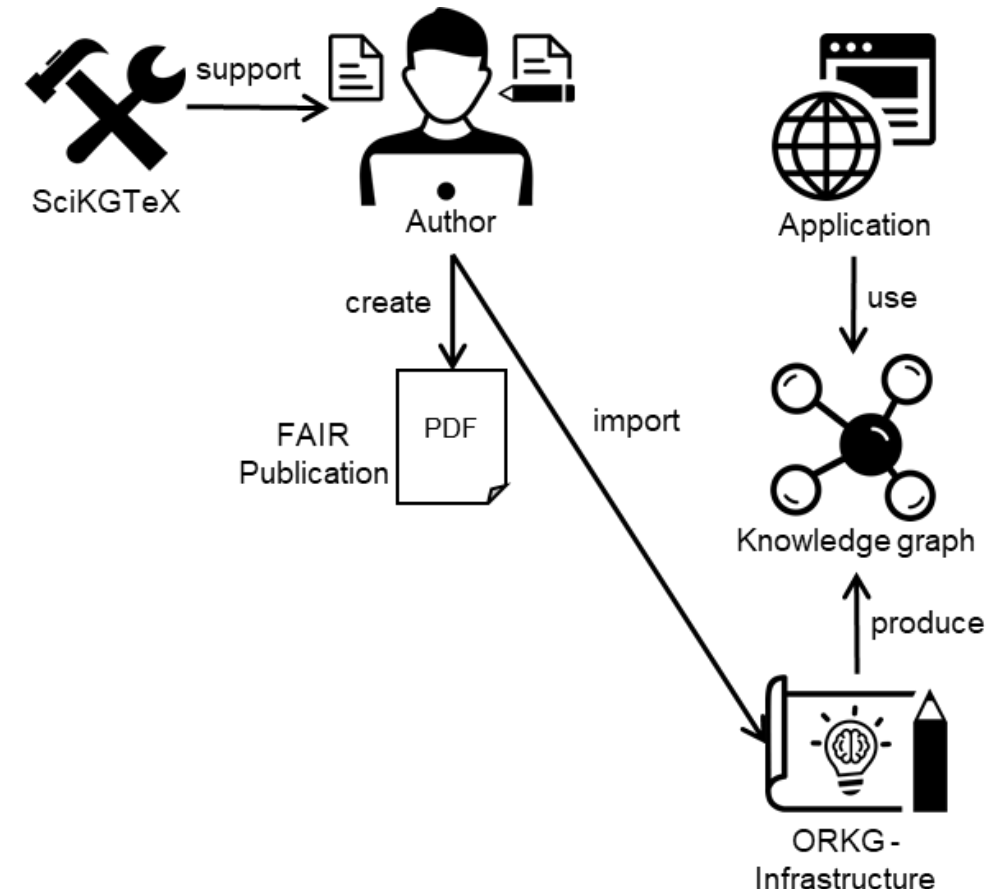
Problem: ORKG focuses on **published** articles, so making them **FAIR** is a **downstream task** and often **not done by the author**.



Solution: “FAIR-by-Design” Artifacts and SciKGT_{EX}^[8]

Solution: “FAIR-by-Design” Artifacts.
SciKGT_{EX} supports authors in making their publication **FAIR** at the time of **creation**.

- **Authors describe** their publication with FAIR information **only once** and **in parallel** at the time of **creation**
- **SciKGT_{EX} embeds** FAIR information into the **PDF metadata** as a knowledge graph
 - **Persistent** over PDF lifetime
 - **Available** for anyone
 - **Reusable**, e.g., import into ORKG



[8] Bless et al.: *SciKGT_{EX} – A LaTeX Package to Semantically Annotate Contributions in Scientific Publications*. 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL), DOI: [10.1109/JCDL57899.2023.00030](https://doi.org/10.1109/JCDL57899.2023.00030), 2023.

SciKGTex – Scientific Knowledge Graph TeX

- Predefined commands for annotation
 - 3 commands for metadata
 - Title
 - Author
 - Research field
 - 5 commands for content
 - Research problem
 - Objective
 - Method
 - Result
 - Conclusion
- Support for own custom annotations
 - REFSQ'24 and REFSQ'25 ask for
 - Code repository
 - Dataset

```
\usepackage{scikgtex}
```

```
\begin{document}
```

The role of `\researchproblem{antibiotic therapy}` is controversial.
The purpose of this study was to `\objective{determine the effectiveness of high-dose amoxicillin/potassium clavulanate in the treatment of children}`.

This was a `\method{randomized, double-blind, placebo-controlled study}`.

`\result{Children receiving the antibiotic were more likely to be cured (50% vs 14%) than children receiving the placebo}`.

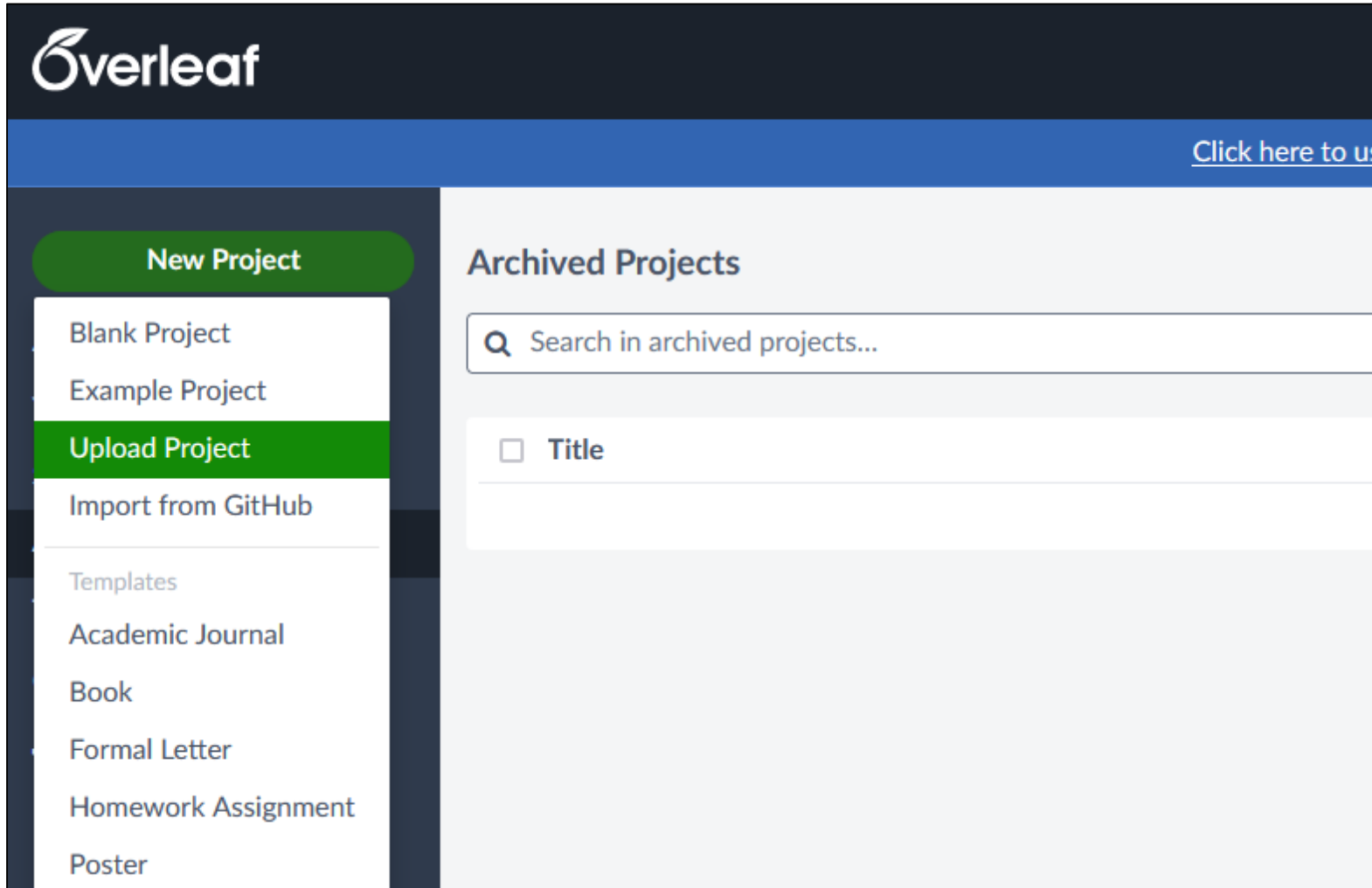
`\conclusion{Amoxicillin/potassium clavulanate results in significantly more cures and fewer failures than placebo}`.

```
\end{document}
```

Complete documentation and latest version:

<https://github.com/Christof93/SciKGTex>

Set up a LaTeX Project (using Overleaf)

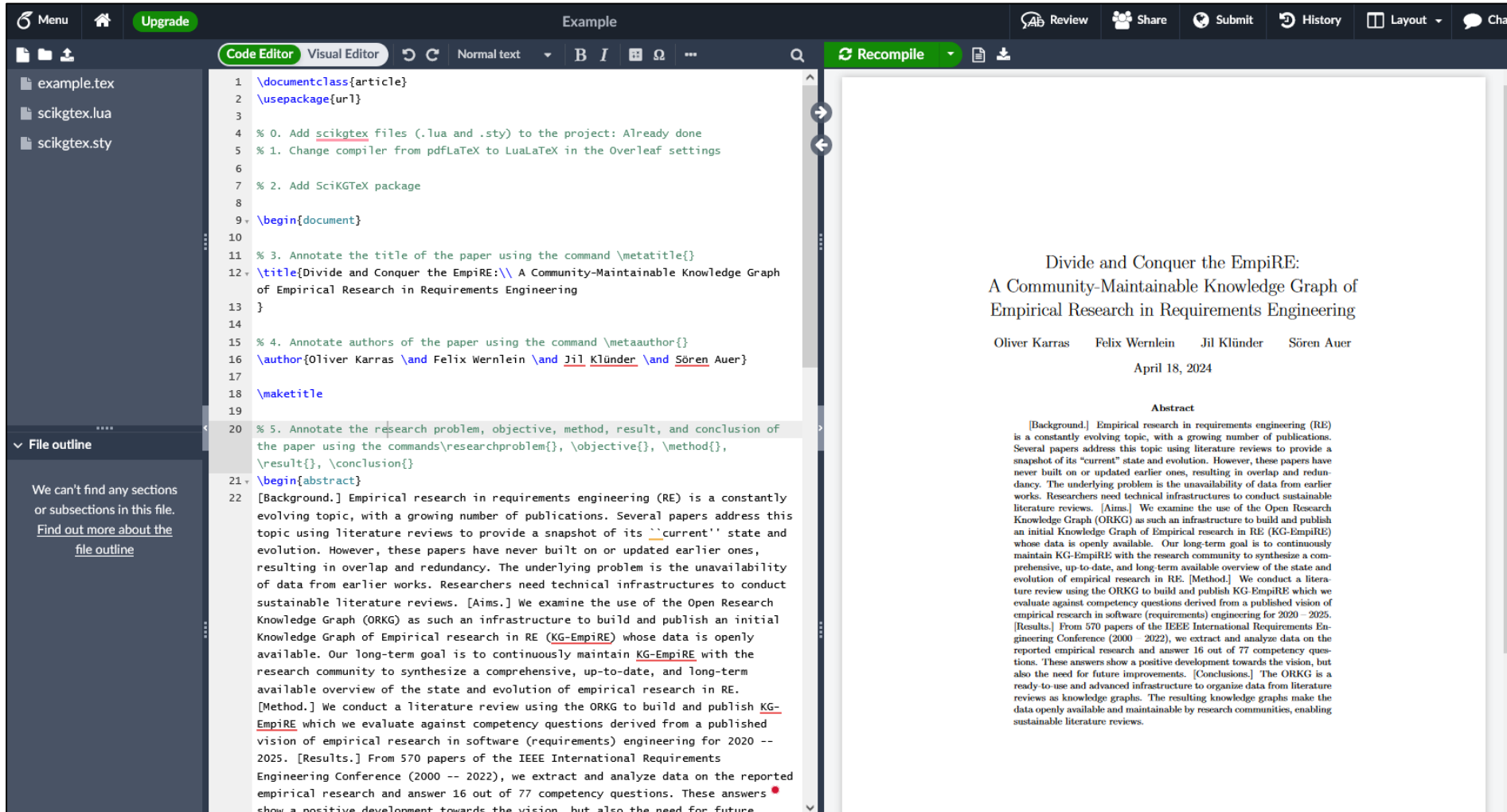


1. Download Example.zip
<https://bit.ly/49RjRbO>
2. Open Overleaf
<https://www.overleaf.com/>
3. Upload Example.zip
New Project → Upload Project
4. Select or drag zip file

Remark:

We only use an abstract for the annotations as a simplified example. Annotations can be used **anywhere** (in one or more LaTeX files).

0. Add SciKGT_EX Files to the Project



The screenshot displays the Overleaf online LaTeX editor interface. The top navigation bar includes 'Menu', 'Upgrade', 'Example', 'Review', 'Share', 'Submit', 'History', 'Layout', and 'Chat'. The left sidebar shows a file explorer with 'example.tex', 'scikgtex.lua', and 'scikgtex.sty'. The main editor area is split into two panes. The left pane shows the LaTeX source code for 'example.tex', which includes comments and commands for document structure, title, authors, and abstract. The right pane shows the rendered PDF output, which displays the title 'Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering', authors 'Oliver Karras, Felix Wernlein, Jil Klünder, Sören Auer', the date 'April 18, 2024', and an abstract section.

```
1 \documentclass{article}
2 \usepackage{url}
3
4 % 0. Add scikgtex files (.lua and .sty) to the project: Already done
5 % 1. Change compiler from pdfLaTeX to LuaLaTeX in the Overleaf settings
6
7 % 2. Add SciKGTEX package
8
9 \begin{document}
10
11 % 3. Annotate the title of the paper using the command \metatitle{}
12 \title{Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph
13 of Empirical Research in Requirements Engineering}
14
15 % 4. Annotate authors of the paper using the command \metaauthor{}
16 \author{Oliver Karras \and Felix Wernlein \and Jil Klünder \and Sören Auer}
17
18 \maketitle
19
20 % 5. Annotate the research problem, objective, method, result, and conclusion of
21 the paper using the commands \researchproblem{}, \objective{}, \method{},
22 \result{}, \conclusion{}
23
24 \begin{abstract}
25 [Background.] Empirical research in requirements engineering (RE) is a constantly
26 evolving topic, with a growing number of publications. Several papers address this
27 topic using literature reviews to provide a snapshot of its "current" state and
28 evolution. However, these papers have never built on or updated earlier ones,
29 resulting in overlap and redundancy. The underlying problem is the unavailability
30 of data from earlier works. Researchers need technical infrastructures to conduct
31 sustainable literature reviews. [Aims.] We examine the use of the Open Research
32 Knowledge Graph (ORKG) as such an infrastructure to build and publish an initial
33 Knowledge Graph of Empirical research in RE (KG-EmpIRE) whose data is openly
34 available. Our long-term goal is to continuously maintain KG-EmpIRE with the
35 research community to synthesize a comprehensive, up-to-date, and long-term
36 available overview of the state and evolution of empirical research in RE.
37 [Method.] We conduct a literature review using the ORKG to build and publish KG-
38 EmpIRE which we evaluate against competency questions derived from a published
39 vision of empirical research in software (requirements) engineering for 2020 --
40 2025. [Results.] From 570 papers of the IEEE International Requirements
41 Engineering Conference (2000 -- 2022), we extract and analyze data on the reported
42 empirical research and answer 16 out of 77 competency questions. These answers
43 show a positive development towards the vision, but also the need for future
```

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

Oliver Karras Felix Wernlein Jil Klünder Sören Auer

April 18, 2024

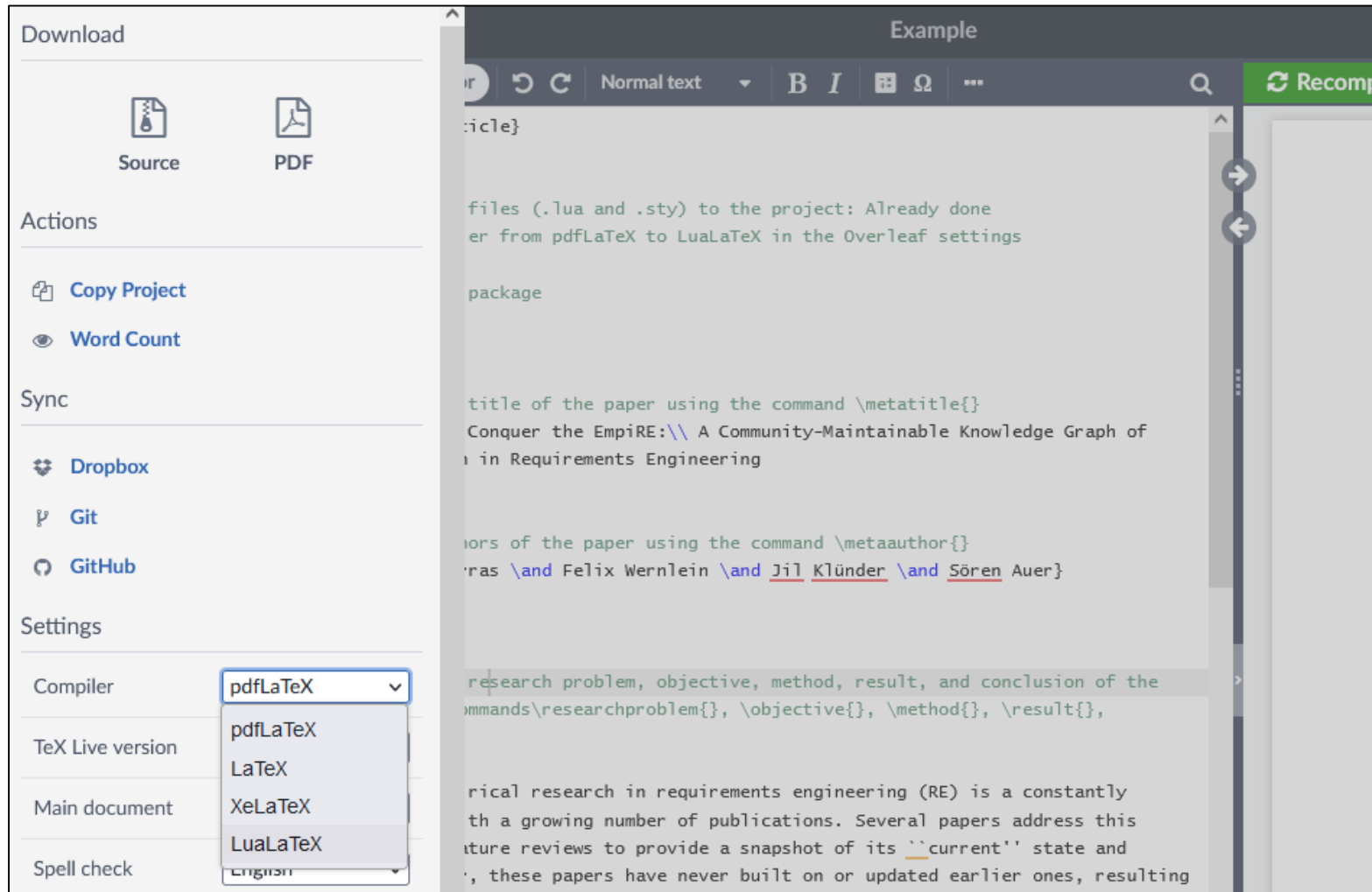
Abstract

[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

The following
two files must
be present:

- scikgtex.lua
- scikgtex.sty

1. Change Compiler from pdfLaTeX to LuaLaTeX



It is necessary to compile your LaTeX source with **LuaLaTeX** for the SciKGT_EX package to work.

2. Add the SciKGT_EX Package to the LaTeX File

```
1 \documentclass{article}
2 \usepackage{url}
3
4 % 0. Add scikgtex files (.lua and .sty) to the project: Already done
5 % 1. Change compiler from pdfLaTeX to LuaLaTeX in the Overleaf settings
6
7 % 2. Add SciKGTEX package
8
9
10 \begin{document}
11
```

Add the command to the preamble of the LaTeX file:

`\usepackage{scikgtex}`

2. Add the SciKGT_EX Package to the LaTeX File: Result

```
1 \documentclass{article}
2 \usepackage{url}
3
4 % 0. Add scikgtex files (.lua and .sty) to the project: Already done
5 % 1. Change compiler from pdfLaTeX to LuaLaTeX in the Overleaf settings
6
7 % 2. Add SciKGTEX package
8 \usepackage{scikgtex}
9
10 \begin{document}
11
```

Add the command to the preamble of the LaTeX file:

`\usepackage{scikgtex}`

3. & 4. Annotate Metadata of the Paper

```
10 ▾ \begin{document}
11
12 % 3. Annotate the title of the paper using the command \metatitle{}
13 ▾ \title{Divide and Conquer the Empire:\\ A Community-Maintainable Knowledge Graph of
    Empirical Research in Requirements Engineering
14 }
15
16 % 4. Annotate authors of the paper using the command \metaauthor{}
17 \author{Oliver Karras \and Felix Wernlein \and Jil Klünder \and Sören Auer}
18
19 \maketitle
20
```

1. Annotate the title

`\metatitle{}`

2. Annotate the authors

`\metaauthor{}`

Remark:

The command `\metaauthor{}` must be used individually **for each author**. With four authors, you need the command four times.

3. & 4. Annotate Metadata of the Paper: Result

```
10 ▾ \begin{document}
11
12 % 3. Annotate the title of the paper using the command \metatitle{}
13 ▾ \title{\metatitle{Divide and Conquer the EmpiRE:\\ A Community-Maintainable Knowledge
    Graph of Empirical Research in Requirements Engineering}
14 }
15
16 % 4. Annotate authors of the paper using the command \metaauthor{}
17 \author{\metaauthor{Oliver Karras} \and \metaauthor{Felix Wernlein} \and
    \metaauthor{Jil Klünder} \and \metaauthor{Sören Auer}}
18
19 \maketitle
20
```

1. Annotate the title

`\metatitle{}`

2. Annotate the authors

`\metaauthor{}`

Remark:

The command `\metaauthor{}` must be used individually **for each author**. With four authors, you need the command four times.

5. Annotate Content of the Paper

```
21 % 5. Annotate the research problem, objective, method, result, and conclusion of the
    paper using the commands\researchproblem{}, \objective{}, \method{}, \result{},
    \conclusion{}
22 \begin{abstract}
23 [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.
24 \end{abstract}
```

1. Annotate the research problem
`\researchproblem{}`
2. Annotate the objective
`\objective{}`
3. Annotate the method
`\method{}`
4. Annotate the result
`\result{}`
5. Annotate the conclusion
`\conclusion{}`

5. Annotate Content of the Paper

21 % 5. Annotate the research problem, objective, method, result, and conclusion of the

Remark:

- All commands can be used **multiple** times.
- The annotated text elements should be as **short** as **possible** and as **long** as **necessary**.
- Definitions:
 - **Research problem:** Issue or gap in existing knowledge addressed by the paper.
 - **Objective:** Goal that the paper aims to achieve.
 - **Method:** Systematic approach, technique, or action plan used in the paper to achieve a goal and result.
 - **Result:** Outcome from a systematic approach, technique, or action plan used in the paper.
 - **Conclusion:** Findings from the analysis of the research results in the paper.

24 \end{abstract}

1. Annotate the research problem
`\researchproblem{}`
2. Annotate the objective
`\objective{}`
3. Annotate the method
`\method{}`
4. Annotate the result
`\result{}`
5. Annotate the conclusion
`\conclusion{}`

5. Annotate Content of the Paper: Result

```
21 % 5. Annotate the research problem, objective, method, result, and conclusion of the
    paper using the commands\researchproblem{}, \objective{}, \method{}, \result{},
    \conclusion{}
22 \begin{abstract}
23 [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
    \researchproblem{unavailability of data from earlier works}. Researchers need technical
    infrastructures to conduct sustainable literature reviews. [Aims.] We examine the
    \objective{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 \method{literature review using the ORKG} to build and publish
    KG-EmpIRE which we \method{evaluate against competency questions} derived from a
    published vision of empirical research in software (requirements) engineering for 2020
    -- 2025. [Results.] \result{From 570 papers of the IEEE International Requirements
    Engineering Conference (2000 -- 2022), we extract and analyze data on the reported
    empirical research} and \result{answer 16 out of 77 competency questions}. These
    answers show a positive development towards the vision, but also the need for future
    improvements. [Conclusions.] \conclusion{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.
24 \end{abstract}
```

1. Annotate the research problem
`\researchproblem{}`
2. Annotate the objective
`\objective{}`
3. Annotate the method
`\method{}`
4. Annotate the result
`\result{}`
5. Annotate the conclusion
`\conclusion{}`

6. Annotate Content of the Paper with Invisible Markup

```
21 % 5. Annotate the research problem, objective, method, result, and conclusion of the
    paper using the commands\researchproblem{}, \objective{}, \method{}, \result{},
    \conclusion{}
22 \begin{abstract}
23 [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
    \researchproblem{unavailability of data from earlier works}. Researchers need technical
    infrastructures to conduct sustainable literature reviews. [Aims.] We examine the
    \objective{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 \method{literature review using the ORKG} to build and publish
    KG-EmpIRE which we \method{evaluate against competency questions} derived from a
    published vision of empirical research in software (requirements) engineering for 2020
    -- 2025. [Results.] \result{From 570 papers of the IEEE International Requirements
    Engineering Conference (2000 -- 2022), we extract and analyze data on the reported
    empirical research} and \result{answer 16 out of 77 competency questions}. These
    answers show a positive development towards the vision, but also the need for future
    improvements. [Conclusions.] \conclusion{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.
24 \end{abstract}
```

What can I do, if the text is not suitable for annotation?

Example:

“... \method{evaluate against competency questions} ...”

Solution:

\method*{evaluation against competency questions}

Remark:

This text is added to the PDF metadata, but **not rendered** in the text of the PDF.

6. Annotate Content of the Paper with Invisible Markup: Result

```
21 % 5. Annotate the research problem, objective, method, result, and conclusion of the
    paper using the commands\researchproblem{}, \objective{}, \method{}, \result{},
    \conclusion{}
22 \begin{abstract}
23 [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 \objective{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
    \method{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.]
    \result{From 570 papers of the IEEE International Requirements Engineering Conference
    (2000 -- 2022), we extract and analyze data on the reported empirical research} and
    \result{answer 16 out of 77 competency questions}. These answers show a positive
    development towards the vision, but also the need for future improvements.
    [Conclusions.] \conclusion{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.
24 \end{abstract}
25
26 % 6. If written text is not suitable for annotation, we can also annotate invisible
    text using the *-notation
27 \researchproblem*{unavailability of the extracted and analyzed data from literature
    reviews}
28 \method*{evaluation against competency questions}
```

What can I do, if the text is not suitable for annotation?

Example:

“... \method{evaluate against competency questions} ...”

Solution:

\method*{evaluation against competency questions}

Remark:

This text is added to the PDF metadata, but **not rendered** in the text of the PDF.

Comparison of Annotated Paper Versions

```

21 % 5. Annotate the research problem, objective, method, result, and conclusion of the
    paper using the commands\researchproblem{}, \objective{}, \method{}, \result{},
    \conclusion{}
22 \begin{abstract}
23 [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
    \researchproblem{unavailability of data from earlier works}. Researchers need technical
    infrastructures to conduct sustainable literature reviews. [Aims.] We examine the
    \objective{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 \method{literature review using the ORKG} to build and publish
    KG-EmpIRE which we \method{evaluate against competency questions} derived from a
    published vision of empirical research in software (requirements) engineering for 2020
    -- 2025. [Results.] \result{From 570 papers of the IEEE International Requirements
    Engineering Conference (2000 -- 2022), we extract and analyze data on the reported
    empirical research} and \result{answer 16 out of 77 competency questions}. These
    answers show a positive development towards the vision, but also the need for future
    improvements. [Conclusions.] \conclusion{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.
24 \end{abstract}

```

```

21 % 5. Annotate the research problem, objective, method, result, and conclusion of the
    paper using the commands\researchproblem{}, \objective{}, \method{}, \result{},
    \conclusion{}
22 \begin{abstract}
23 [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 \objective{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
    \method{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.]
    \result{From 570 papers of the IEEE International Requirements Engineering Conference
    (2000 -- 2022), we extract and analyze data on the reported empirical research} and
    \result{answer 16 out of 77 competency questions}. These answers show a positive
    development towards the vision, but also the need for future improvements.
    [Conclusions.] \conclusion{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.
24 \end{abstract}
25
26 % 6. If written text is not suitable for annotation, we can also annotate invisible
    text using the *-notation
27 \researchproblem*{unavailability of the extracted and analyzed data from literature
    reviews}
28 \method*{evaluation against competency questions}

```

7. (Optional) Annotate Research Field

Remark:

1. A paper in the ORKG is assigned to a research field based on the DFG classification. All research fields: <https://orkg.org/fields>.
2. URI of the ORKG semantic web resource for Software Engineering: <https://orkg.org/resource/R140>.
3. Use the command `\uri{"URI"}{"Label"}` inside an annotation to refer to resources in the semantic web. The first argument is the **URI** to the semantic resource and the second is an optional **Label**.
4. The term "Software Engineering" does not appear in the abstract, so we need an **invisible** annotation and we also use the `\uri{}` command to create a reference to the semantic resource.

1. Annotate the research field

`\researchfield{}`

2. Refer to the ORKG semantic web resource [Software Engineering](https://orkg.org/resource/R140)

`\uri{"URI"}{"Label"}`

7. (Optional) Annotate Research Field: Result

```
29 % 7. Optional: Annotate research field of the paper
30 \researchfield*{\uri{https://orkg.org/resource/R140}{Software Engineering}}
```

Remark:

1. A paper in the ORKG is assigned to a research field based on the DFG classification. All research fields: <https://orkg.org/fields>.
2. URI of the ORKG semantic web resource for Software Engineering: <https://orkg.org/resource/R140>.
3. Use the command `\uri{"URI"}{"Label"}` inside an annotation to refer to resources in the semantic web. The first argument is the **URI** to the semantic resource and the second is an optional **Label**.
4. The term "Software Engineering" does not appear in the abstract, so we need an **invisible** annotation and we also use the `\uri{}` command to create a reference to the semantic resource.

1. Annotate the research field

`\researchfield{}`

2. Refer to the ORKG semantic web resource [Software Engineering](https://orkg.org/resource/R140)

`\uri{"URI"}{"Label"}`

8. Use Custom Annotations of REFSQ'24 and REFSQ'25

```
33 % 8. Optional: Using REFSQ'24 and REFSQ'25 annotations
34 \contribution*{code repository}{\url{https://github.com/okarras/EmpIRE-Analysis}}
35 \contribution*{dataset}{\url{https://orkg.org/api/rdf/dump}}
36
37 \end{document}
```

Remark:

1. Use the command `\contribution{"Property name"}{"Label"}` to add a custom annotation for your domain. The first argument is the **Property name** of the property from the ORKG you want to use, and the second is an optional **Label**. SciKGT_{EX} checks if a property with the provided exact **Property name** exists and replaces it with the internal property ID in the ORKG namespace.
2. All ORKG properties can be found here: <https://orkg.org/properties>.

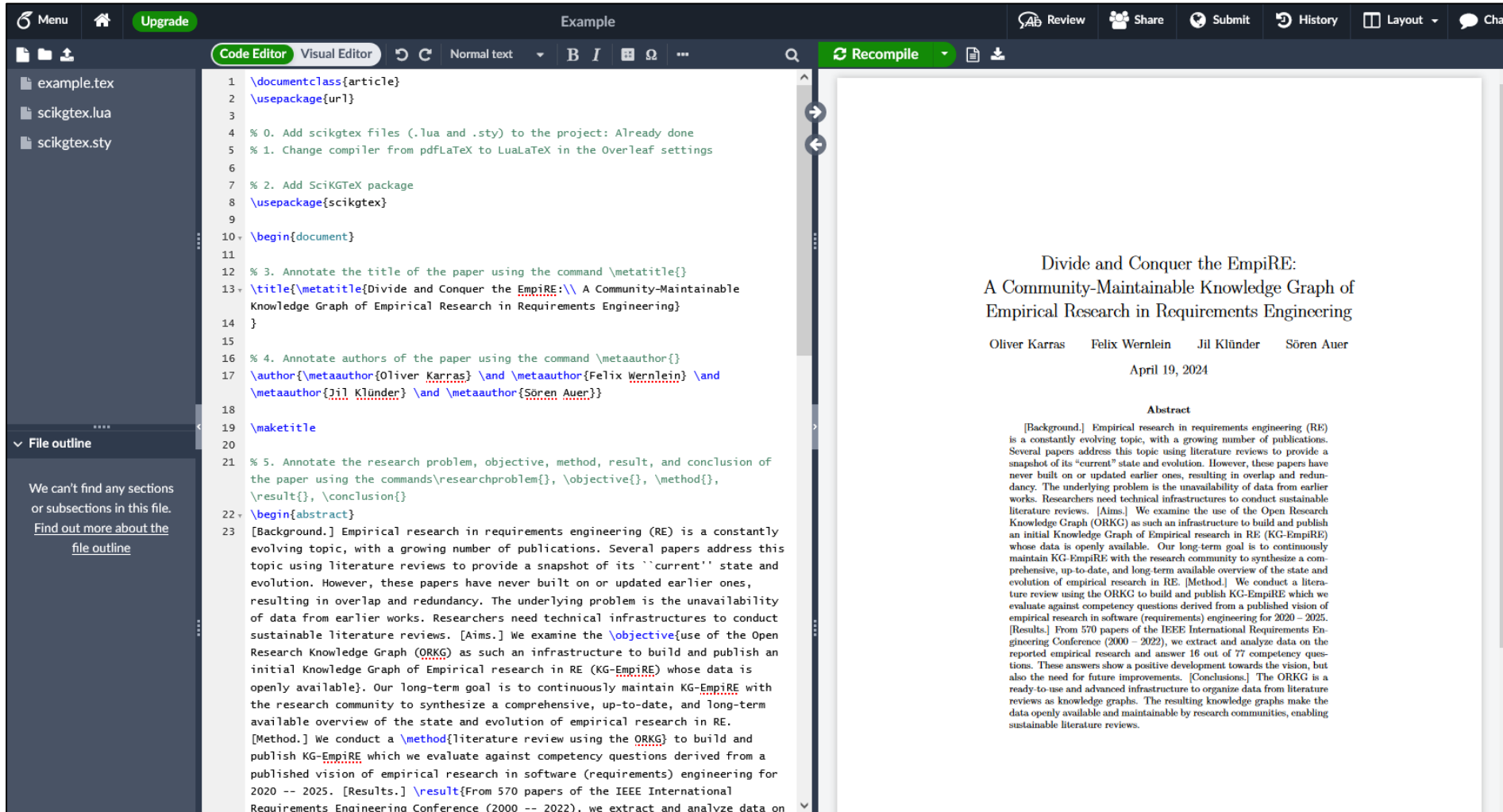
1. Annotate the code repository

`\contribution{code repository}{}`

2. Annotate the dataset

`\contribution{dataset}{}`

9. Generate FAIR-Annotated PDF of the Paper



The screenshot shows the Overleaf online LaTeX editor interface. The left sidebar displays the file explorer with files: `example.tex`, `scikgtex.lua`, and `scikgtex.sty`. The main editor area shows the LaTeX source code with line numbers and annotations. The right sidebar shows the compiled PDF output, which includes a title page and an abstract.

LaTeX Source Code (Left Panel):

```

1 \documentclass{article}
2 \usepackage{url}
3
4 % 0. Add scikgtex files (.lua and .sty) to the project: Already done
5 % 1. Change compiler from pdfLaTeX to LuaLaTeX in the Overleaf settings
6
7 % 2. Add ScikGTeX package
8 \usepackage{scikgtex}
9
10 \begin{document}
11
12 % 3. Annotate the title of the paper using the command \metatitle{}
13 \title{\metatitle{Divide and Conquer the EmpiRE: A Community-Maintainable
14 Knowledge Graph of Empirical Research in Requirements Engineering}
15 }
16
17 % 4. Annotate authors of the paper using the command \metaauthor{}
18 \author{\metaauthor{Oliver Karras} \and \metaauthor{Felix Wernlein} \and
19 \metaauthor{Jil Klünder} \and \metaauthor{Sören Auer}}
20
21 \maketitle
22
23 % 5. Annotate the research problem, objective, method, result, and conclusion of
24 the paper using the commands \researchproblem{}, \objective{}, \method{},
25 \result{}, \conclusion{}
26 \begin{abstract}
27 [Background.] Empirical research in requirements engineering (RE) is a constantly
28 evolving topic, with a growing number of publications. Several papers address this
29 topic using literature reviews to provide a snapshot of its "current" state and
30 evolution. However, these papers have never built on or updated earlier ones,
31 resulting in overlap and redundancy. The underlying problem is the unavailability
32 of data from earlier works. Researchers need technical infrastructures to conduct
33 sustainable literature reviews. [Aims.] We examine the \objective{use of the Open
34 Research Knowledge Graph (ORKG) as such an infrastructure to build and publish an
35 initial Knowledge Graph of Empirical research in RE (KG-EmpIRE) whose data is
36 openly available}. Our long-term goal is to continuously maintain KG-EmpIRE with
37 the research community to synthesize a comprehensive, up-to-date, and long-term
38 available overview of the state and evolution of empirical research in RE.
39 [Method.] We conduct a \method{literature review using the ORKG} to build and
40 publish KG-EmpIRE which we evaluate against competency questions derived from a
41 published vision of empirical research in software (requirements) engineering for
42 2020 -- 2025. [Results.] \result{From 570 papers of the IEEE International
43 Requirements Engineering Conference (2000 -- 2022), we extract and analyze data on
  
```

Compiled PDF Output (Right Panel):

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

Oliver Karras Felix Wernlein Jil Klünder Sören Auer

April 19, 2024

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.

Simply...
recompile

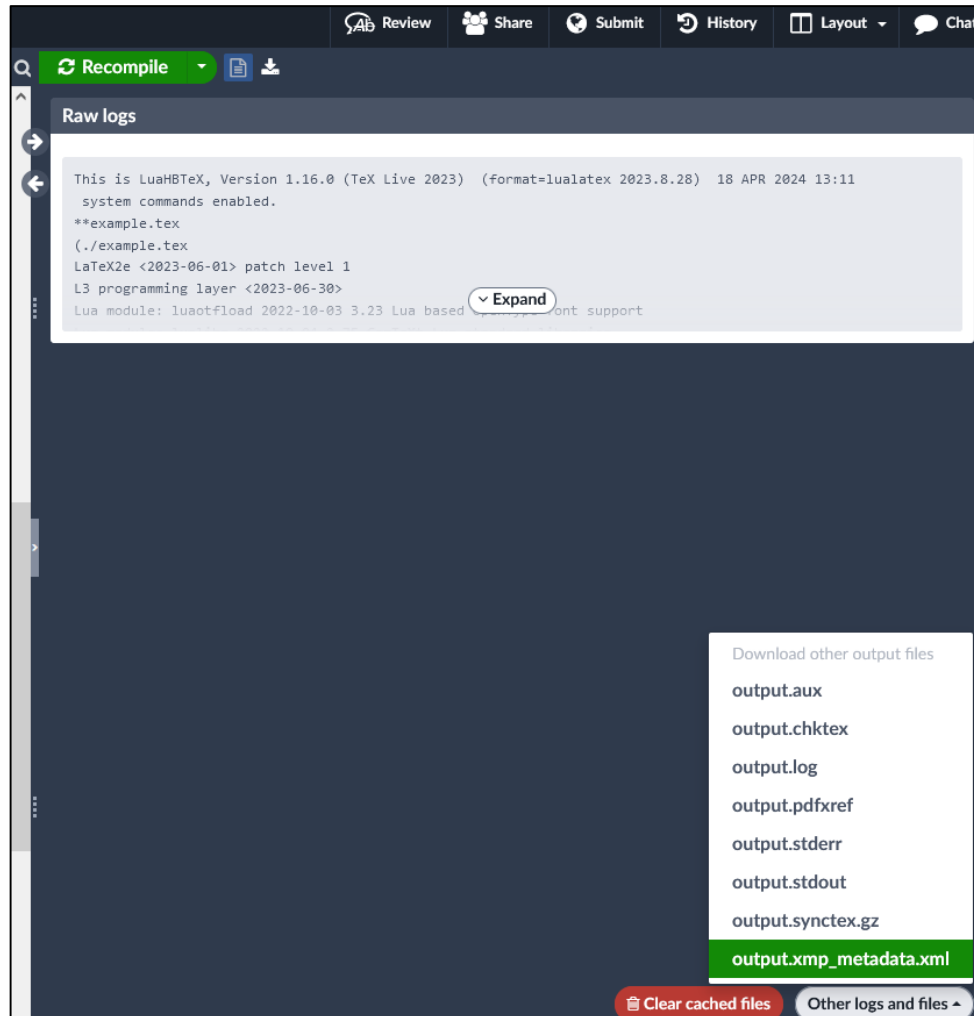
Remark:
Each recompile
adds the
annotations to the
metadata of the
generated PDF.

9. Generate FAIR-Annotated PDF of the Paper

The screenshot displays the SciKGTex web interface. On the left, a file explorer shows 'example.tex', 'scikgtex.lua', and 'scikgtex.sty'. Below it, a 'File outline' section states: 'We can't find any sections or subsections in this file. Find out more about the file outline'. The main area is a code editor showing LaTeX code for a document class, package loading, and annotations for a paper titled 'Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering'. The code includes annotations for authors (Oliver Karras, Felix Wernlein, Jiri Klunder, Soren Auer) and a detailed abstract. On the right, a 'Warnings' panel lists five SciKGTex warnings: 'No researchproblem annotation found!', 'No objective annotation found!', 'No method annotation found!', 'No result annotation found!', and 'No conclusion annotation found!'. Each warning has a link to the corresponding section in the document. Below the warnings is a 'Raw logs' section showing the LuaHBTeX version and system commands.

For the 5 predefined commands for content, SciKGTex also provides **warnings** if an annotation is missing.

10. Check the FAIR information embedded in the PDF



1. Open “Logs and outputs files”
2. Select “Other logs and files”
3. Download “output.xmp_metadata.xml”
4. Open “output.xmp_metadata.xml”

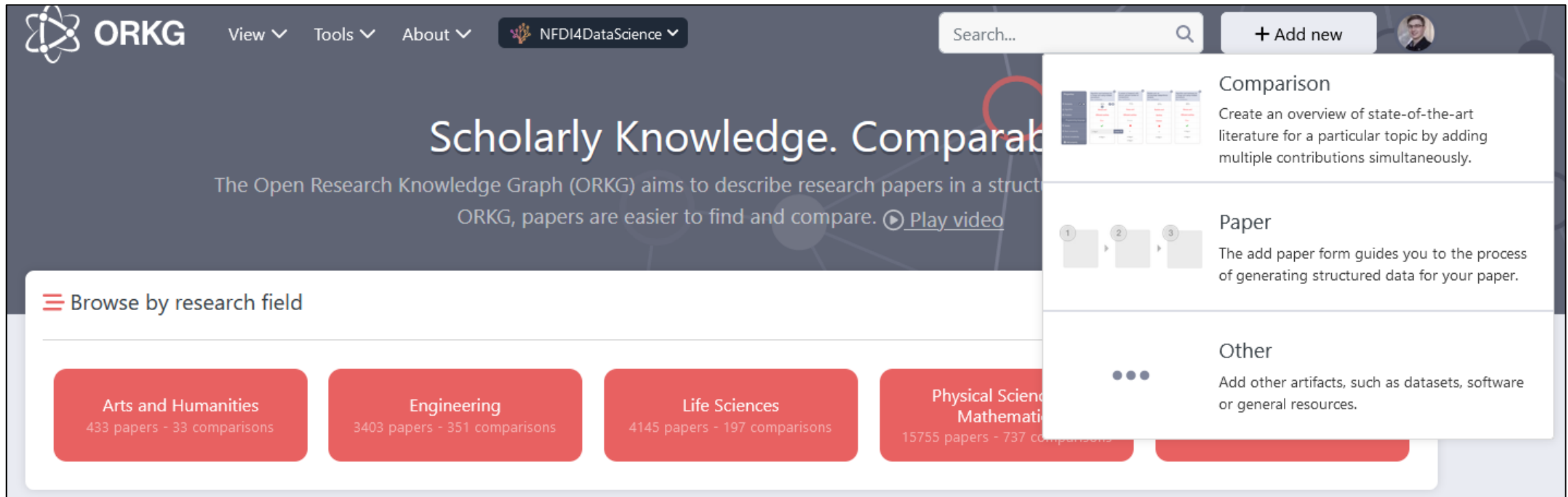
Remark:

We only use an abstract for the annotations as a simplified example. Annotations can be used **anywhere** (in one or more LaTeX files).

10. Check the FAIR information embedded in the PDF

```
<?xpacket begin="?" id="a7c48312-233b-400b-c0f3-0c296941c6"?>
▼<x:xmpmeta xmlns:x="adobe:ns:meta/">
▼<rdf:RDF xmlns:orkg="http://orkg.org/core#" xmlns:orkg_property="http://orkg.org/property/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
▼<rdf:Description rdf:about="https://www.orkg.org/orkg/paper/a7c48312-233b-400b-c0f3-0c296941c6">
  <rdf:type rdf:resource="http://orkg.org/core#Paper"/>
  <orkg:hasTitle>Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering</orkg:hasTitle>
  <orkg:hasAuthor>Oliver Karras</orkg:hasAuthor>
  <orkg:hasAuthor>Felix Wernlein</orkg:hasAuthor>
  <orkg:hasAuthor>Jil Klünder</orkg:hasAuthor>
  <orkg:hasAuthor>Sören Auer</orkg:hasAuthor>
▼<orkg_property:P30>
▼<rdf:Description rdf:about="https://orkg.org/resource/R140">
  <rdfs:label>Software Engineering</rdfs:label>
  </rdf:Description>
</orkg_property:P30>
▼<orkg:hasResearchContribution>
▼<orkg:ResearchContribution rdf:about="https://www.orkg.org/orkg/paper/a7c48312-233b-400b-c0f3-0c296941c6/contribution_ORKG_default">
  <orkg_property:P15051>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</orkg_property:P15051>
  <orkg_property:P1005>literature review using the ORKG</orkg_property:P1005>
  <orkg_property:P1006>From 570 papers of the IEEE International Requirements Engineering Conference (2000 -- 2022), we extract and analyze data on the reported empirical research</orkg_property:P1006>
  <orkg_property:P1006>answer 16 out of 77 competency questions</orkg_property:P1006>
  <orkg_property:P15419>The ORKG is a ready-to-use and advanced infrastructure to organize data from literature reviews as knowledge graphs</orkg_property:P15419>
  <orkg_property:P32>unavailability of the extracted and analyzed data from literature reviews</orkg_property:P32>
  <orkg_property:P1005>evaluation against competency questions</orkg_property:P1005>
  <orkg_property:P49000>https://github.com/okarras/EmpiRE-Analysis</orkg_property:P49000>
  <orkg_property:P2005>https://orkg.org/api/rdf/dump</orkg_property:P2005>
  </orkg:ResearchContribution>
  </orkg:hasResearchContribution>
</rdf:Description>
</rdf:RDF>
</x:xmpmeta>
<?xpacket end="r"?>
```


11. Import the FAIR information into the ORKG

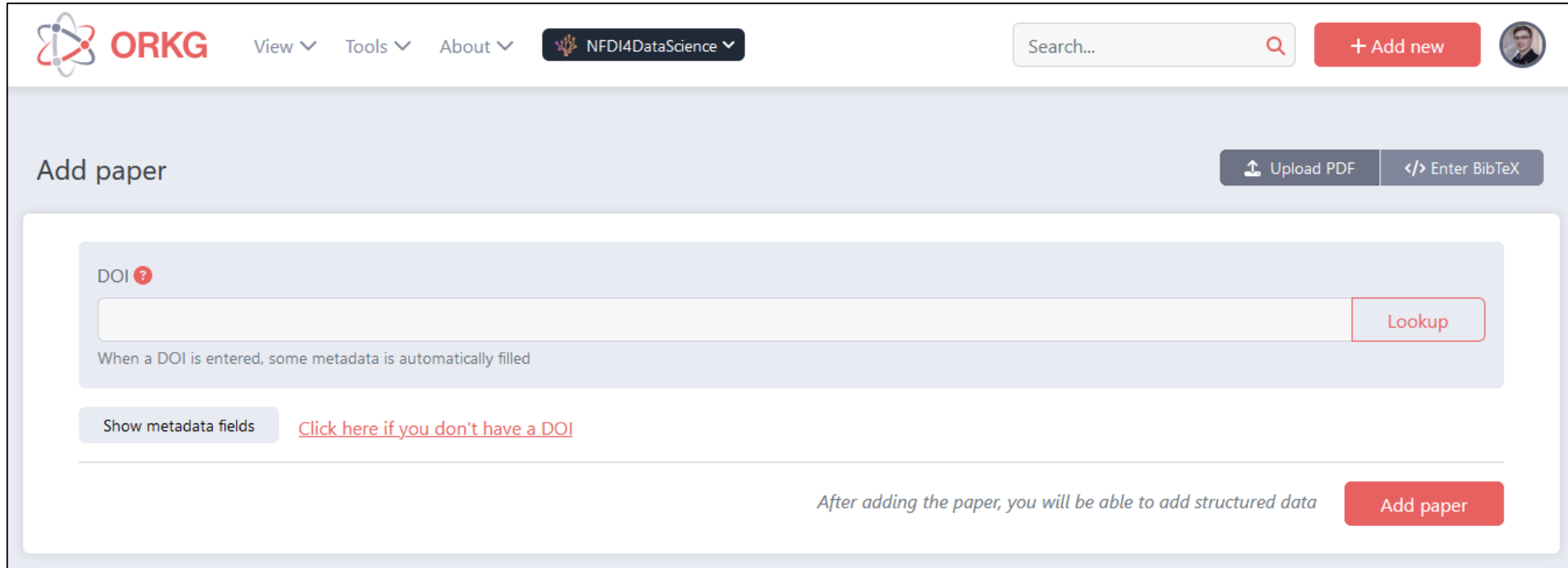


1. Log in with your credentials
2. Click “+ Add new”
3. Select “Paper”

Remark:

ORKG does **not save** the uploaded PDF. It **only extracts** the FAIR annotations embedded in the PDF metadata.

11. Import the FAIR information into the ORKG



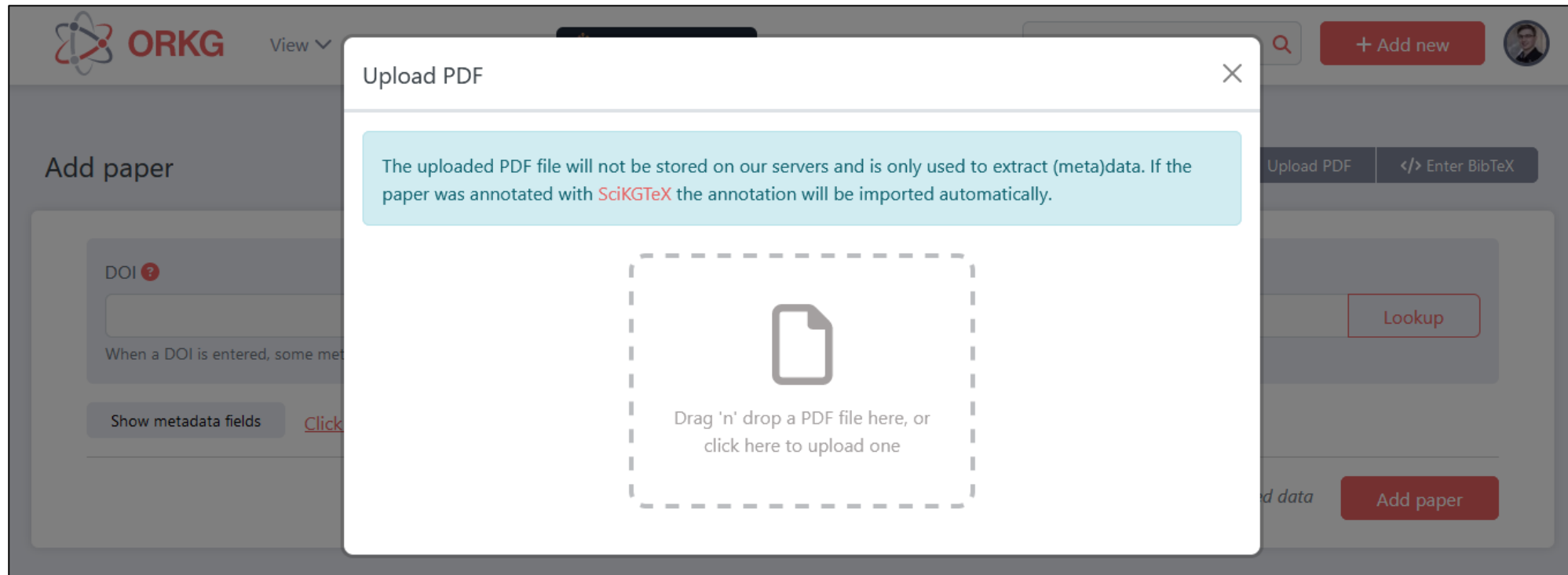
The screenshot shows the ORKG (Open Research Knowledge Graph) interface for adding a new paper. The header includes the ORKG logo, navigation links (View, Tools, About), a dropdown menu for 'NFDI4DataScience', a search bar, and a '+ Add new' button. The main section is titled 'Add paper' and contains two buttons: 'Upload PDF' and 'Enter BibTeX'. Below these is a form for entering a DOI, with a 'Lookup' button. A note states: 'When a DOI is entered, some metadata is automatically filled'. There is also a link 'Click here if you don't have a DOI' and a button 'Show metadata fields'. At the bottom, a message says 'After adding the paper, you will be able to add structured data' next to an 'Add paper' button.

1. Select “Upload PDF”

Remark:

ORKG does **not save** the uploaded PDF. It **only extracts** the FAIR annotations embedded in the PDF metadata.

11. Import the FAIR information into the ORKG



1. Select or drag your PDF file

Remark:

ORKG does **not save** the uploaded PDF. It **only extracts** the FAIR annotations embedded in the PDF metadata.

11. Import the FAIR information into the ORKG

ORKG View Tools About NFDI4DataScience Search... + Add new

Add paper Upload PDF Enter BibTeX

DOI ? Look up
When a DOI is entered, some metadata is automatically filled

Hide metadata fields

Paper title *(required)* ?
Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering

Research field *(required)* ?
Software Engineering Choose

Paper authors ?
Oliver Karras ✎ ✕
Felix Wernlein ✎ ✕
Jil Klünder ✎ ✕
Sören Auer ✎ ✕
+ Add author

Publication month ? Month ▼ Publication year ? Year ▼

Published in ? x ▼

Paper URL ?

After adding the paper, you will be able to add structured data Add paper

1. ORKG shows imported metadata
2. Optional: Edit the metadata
3. Select “Add paper” at the bottom

Remark:

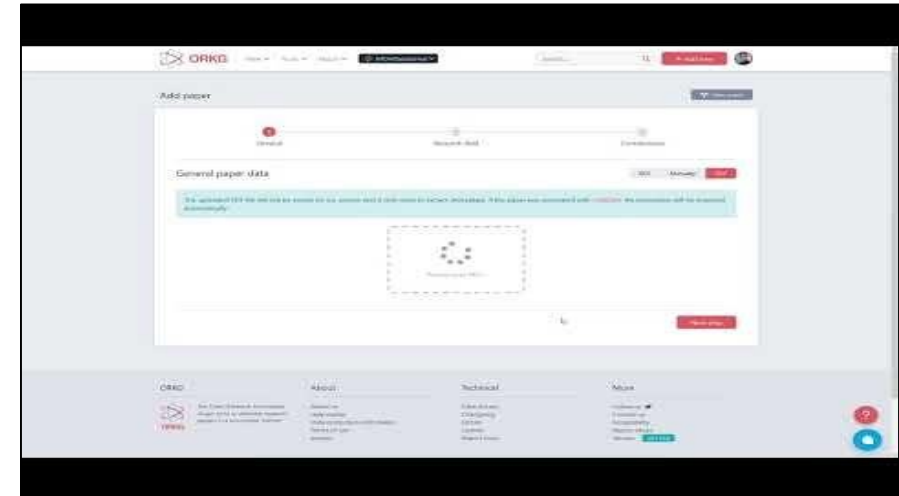
ORKG does **not save** the uploaded PDF. It **only extracts** the FAIR annotations embedded in the PDF metadata.

11. Import the FAIR information into the ORKG: Result

The screenshot shows the ORKG interface with a paper titled "Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering". The paper is categorized under "Software Engineering" and lists authors: Oliver Karras, Felix Wernlein, Jill Klünder, and Sören Auer. The main content area displays "Contribution 1" with a table of metadata:

Field	Value
code repository	https://github.com/okarras/EmpiRE-Analysis
Conclusion	The ORKG is a ready-to-use and advanced infrastructure to organize data from literature reviews as knowledge graphs
dataset	https://orkg.org/api/rdf/dump
method	evaluation against competency questions literature review using the ORKG
Objective	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
research problem	unavailability of the extracted and analyzed data from literature reviews
result	answer 16 out of 77 competency questions From 570 papers of the IEEE International Requirements Engineering Conference (2000 -- 2022), we extract and analyze data on the reported empirical research

On the right side, there is a "Provenance" section showing the paper was added on 27 May 2024 by Oliver Karras. A button "Assign to observatory" is visible at the bottom of the provenance section.



https://www.youtube.com/watch?v=ZzrQ_YCKVsYa

Remark:

ORKG does **not save** the uploaded PDF. It **only extracts** the FAIR annotations embedded in the PDF metadata.

ORKG: Describing Papers Manually and Creating an ORKG Comparison

Oliver Karras, Alessio Ferrari, Davide Fucci, and Davide Dell'Anna

oliver.karras@tib.eu, alessio.ferrari@isti.cnr.it, davide.fucci@bth.se, d.dellanna@uu.nl

32nd IEEE International Requirements Engineering 2024 Conference – Exploring New Horizons: Expanding the Frontiers of Requirements Engineering

June 24th, 2024, Reykjavik, Iceland

ORKG Comparisons

Acknowledgement
of creators

DOI

Visualizations

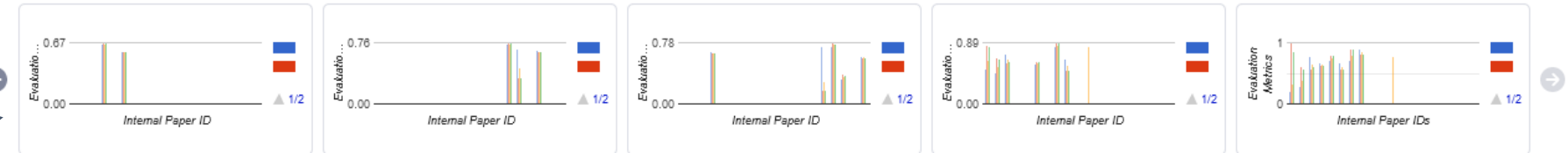
Interactive filtering

Overview of Approaches that Classify User Feedback as Feature Request ☆👁

📅 June 2021 👤 Oliver Karras 👤 Eduard C. Groen

This overview shows the classification results of approaches that use the machine learning algorithms Naïve Bayes, Support Vector Machines, and Decision Trees C4.5 in combination with the machine learning features Bag of Words or Term Frequency - Inverse Document Frequency to classify user feedback as feature request.

DOI: <https://doi.org/10.48366/r112387>



Properties

[has dataset](#)

Software Feature Request
Detection in Issue Tracking
Systems

User Feedback Classification - 2016

https://zenodo.org/record/56907#.YKT_NudCRPY

Mining User Requirements from
Application Store Reviews Using
Frame Semantics

User Feedback Classification - 2017

https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip

<https://sites.google.com/site/appuserreviews/>
seel.cse.lsu.edu/data/refsq17.zip

Mining Twitter Feeds for Software
User Requirements

User Feedback Classification - 2017

seel.cse.lsu.edu/data/re17.zip

Automatic Classification of Non-
Functional Requirements from
Augmented App User Reviews

User Feedback Classification - 2017

Not available

Bug reports
simply p

classifying

User Feedback Classification - 2017

https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip

Example: ORKG Comparison of Related Work – State of the Art^[5]

A Comparison of Scientific Publications on the State of Empirical Research in Requirements Engineering and Software Engineering

November 2023

Oliver Karras

Felix Wernlein

Jil Ann-Christin Klünder

Sören Auer

This comparison provides an overview of scientific publications that have investigated primary studies in requirements engineering and software engineering to give a snapshot of the "current" state of empirical research in requirements engineering and software engineering. In particular, the comparison shows for each publication (1) which research fields and topics were investigated, (2) whether and where the extracted and analyzed data is available, and (3) which method was used to determine the state, including further details about the respective method.

DOI: <https://doi.org/10.48366/R650023>

Properties	Empirical research in requirements engineering: trends and opportunities <i>Empirical research - 2016</i>	Empirical research methodologies and studies in Requirements Engineering: How far did we come? <i>Empirical research - 2014</i>	A Survey on Empirical Requirements Engineering Research Practices <i>Empirical research - 2012</i>	Evidence-Based Structuring and Evaluation of Empirical Research in Requirements Engineering: Fundamentals, Framework, Research Map <i>Empirical research - 2010</i>	An Analysis of Requirements Engineering Data <i>Empirical research - 2008</i>
research problem	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering
research field investigated	Requirements Engineering	Requirements Engineering	Requirements Engineering	Requirements Engineering	Requirements Engineering
topic investigated	bibliographic metadata context data collection	bibliographic metadata research topic theory	context	context	context

<https://doi.org/10.48366/R650023>


Properties	State of Empirical Research in Software Engineering Empirical research - 2007	The type of evidence produced by empirical software engineers Empirical research - 2005	Research in software engineering: an analysis of the literature Empirical research - 2002
method	Literature review	Literature review	Literature review
data availability	✗	✗	✗
time interval/interval	1996-01-01	1997-01-01	1999-01-01
"has been investigated"	2006-06-30	2003-12-31	1999-12-31
number of papers	133	119	360

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 State	Venues (Frequency > 2)	Themes (Frequency > 2)
[8]	2005	RE	LR	1968 – 2002	35 papers Unavailable	1) Empirical Software Engineering Journal (8)	1) Data collection (12)
[15]	2016	RE	SLR	1977 – 2015	42 respondents Unavailable	2) IEEE Software (4)	2) Research method (11)
[11]	2012	RE	Survey	1970.01.2012 – 2010.01.2012	2237 papers Unavailable	3) Requirements Engineering Journal (4)	3) Bibliographic metadata (10)
[2]	2016	SE	SLR	Open – 2012	290 papers Unavailable	4) ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (4)	4) Data analysis (8)
[1]	2005	RE	LR	1997 – 1999	700 papers Unavailable	5) IEEE Transactions on Software Engineering (3)	5) Research paradigm (7)
[3]	2002	RE	LR	1996 – 2002	68 papers Unavailable	6) Information and Software Technology Journal (3)	6) Research design (7)
[10]	2006	RE	LR	1997 – 2005	65 papers Unavailable	7) IEEE International Requirements Engineering Conference (3)	7) Research topic (5)
[11]	2007	SE	LR	1996 – 2003	133 papers Unavailable	8) Journal of Systems and Software (3)	8) Research context (4)
[14]	2015	SE	SLR	1996 – 2013	991 papers Broken link	9) International Conference on Software Engineering (3)	9) Theory (5)
[13]	2018	SE	SLR	2017 – 2017	535 papers Broken link		
[14]	2019	SE	SLR	1997 – 2014	941 papers Unavailable		
[15]	2021	SE	SMS	Open – 2019	20 papers Available		

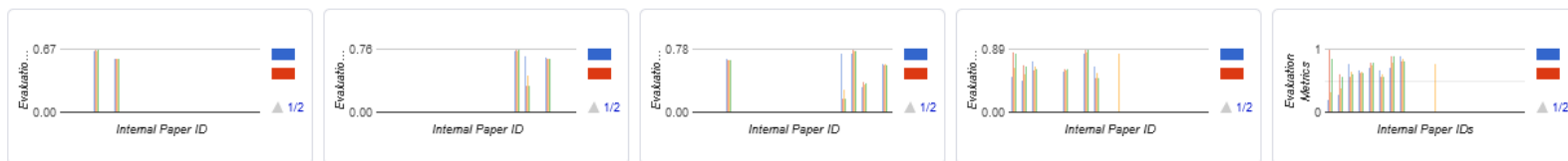
Example: ORKG Comparison of Literature – ML Approaches^[9]

Overview of Approaches that Classify User Feedback as Feature Request ☆🔍

June 2021 Oliver Karras Eduard C. Groen

This overview shows the classification results of approaches that use the machine learning algorithms Naïve Bayes, Support Vector Machines, and Decision Trees C4.5 in combination with the machine learning features Bag of Words or Term Frequency - Inverse Document Frequency to classify user feedback as feature request.

DOI: <https://doi.org/10.48366/r112387>



Properties	Software Feature Request Detection in Issue Tracking Systems <i>User Feedback Classification - 2016</i>	Mining User Requirements from Application Store Reviews Using Frame Semantics <i>User Feedback Classification - 2017</i>	Mining Twitter Feeds for Software User Requirements <i>User Feedback Classification - 2017</i>	Automatic Classification of Non-Functional Requirements from Augmented App User Reviews <i>User Feedback Classification - 2017</i>	Bug reports simply classified <i>User Feedback Classification - 2017</i>
has dataset	https://zenodo.org/record/56907#.YKT_NudCRPY	https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip https://sites.google.com/site/appsuserreviews/ seel.cse.lsu.edu/data/refsq17.zip	seel.cse.lsu.edu/data/re17.zip	Not available	https://seel.cse.lsu.edu/data/re17.zip
https://doi.org/10.48366/r112387					



Fig. 2: Excerpt from our comparison for Case II [35].

77 papers from the SLR by Khan et al. [18], describing the relation of the papers to five phases of RE and the CrowdRE utilities applied [36]. We are still in the process of adding the contributions from the remaining 50 papers, which is more time-consuming than for the quantitative data from Case I because of the expert judgments needed for classifying the papers' contributions. The comparison of the 27 papers makes it easy to identify, for example, the four papers that address the runtime purpose of monitoring for requirements evolution.

With the created comparisons [35], [36], we achieved our goal of acquiring and curating the detailed results of both SLRs. The knowledge-based representation in the form of comparisons has several advantages over a purely document-based representation. The comparisons are interactive and allow filtering of views by different scholarly knowledge contained in each row, even by specific value ranges of qualitative and quantitative content. The ORKG also provides a service for generating several graphical visualizations based on data in the comparisons, helping the reader understand information faster than through the large comparison table. The most important feature of the ORKG is that the added contributions and created comparisons are available to anyone. In this way, every crowd member can use the curated scholarly knowledge and created comparisons as a basis for new comparisons. Moreover, the existing comparisons can be expanded with additional scholarly knowledge from papers already included, and even with new contributions from papers added later to the ORKG. We already expanded several contributions, e.g., the results of other classifications reported in Dhinakaran et al.'s paper [40]¹⁹. For Case I, we added the details of the three crowd properties *scale*, *level of knowledge*, *skills & expertise*, and *roles*, which are only briefly and superficially described in the SLR [18]. For Case II, we added links to the datasets used and performance values to classification categories other than "Feature Request". This expansion is relevant to enable long-term curation. For

¹⁹<https://www.orkg.org/orkg/paper/RC76818/RC76825>

example, a development succeeding the SLR by Santos et al. [17] are reports of Deep Learning algorithms showing promising results in classifying user feedback [41], [42], which should be successively added to the comparison.

Despite all these advantages, the ORKG also has limitations. Most of the limitations we experienced can be attributed to the development status of the platform, which is currently in beta. Further development of the ORKG must improve interactions for the expert crowd by enabling better workflows for entering data and creating visualizations. Nevertheless, we also experienced that the project team has always responded directly to our reported issues, which we could see getting added to the GitHub issue tracker²¹ and addressed shortly thereafter.

V. DISCUSSION

The ORKG aroused our interest as a crowdsourcing platform for applying and communicating CrowdRE research. In this experience report, we explored whether the ORKG can promote the potential of CrowdRE in open source and open research settings, taking two perspectives: that of CrowdRE researchers and that of crowd members.

Our first contribution is that we provide a comprehensive overview of the ORKG's features as a crowdsourcing platform for acquiring and curating scholarly knowledge [37], mapped to the four key activities of CrowdRE. Our findings show that the ORKG is a crowdsourcing platform offering several features that can facilitate successful CrowdRE. Although the ORKG project team has not yet consciously applied CrowdRE, they already address crucial parts of the CrowdRE cycle by motivating crowd members to participate, eliciting feedback, and monitoring context & usage data, which they analyze to derive and implement the needs and requirements of the crowd.

To motivate crowd members, the project team uses established mechanisms and incentives to boost intrinsic and extrinsic motivation (see Finding 1). Feedback is elicited

²¹<https://github.com/TIBHannover/orkg-blocks-frontendl-fissues/634>

Example: ORKG Comparison of Literature – Simulation Parameters^[10]

Comparison of Studies on Germany's Energy Supply in 2050 ★🔍

November 2021 Felix Kullmann Jan Göpfert Oliver Karras Patrick Kuckertz Sören Auer Markus Stocker Peter Markewitz
 Leander Kotzur Detlef Stolten

This comparison compiles the results from various studies analyzing a future low-carbon energy system for Germany. The focus of this study comparison is electricity generation. In the future, however, other essential characteristics of the respective energy system designs in the individual studies will be listed. Installed capacity is given in GW and electricity generation is given in TWh. The authors would like to thank the German Federal Government, the German State Governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) - project number: 442146713. This work was also supported by the Helmholtz Association under the program "Energy System Design".

DOI: <https://doi.org/10.48366/r153801>



Properties	Klimaneutrales Deutschland Contribution - 2020	Wasserstoff-Roadmap Nordrhein- Westfalen Contribution - 2020	Wege zu einem klimaneutralen Energiesystem Contribution - 2020	Wege für die Energiewende Contribution - 2019	Den Weg treibhaus Deutsch gestalten Contribu
has energy sources	all sources	all sources	all sources	all sources	
	bioenergy	bioenergy	bioenergy	bioenergy	
	geothermics	geothermics	geothermics	geothermics	
	hydropower	hydropower	hydropower	hydropower	
	import	import			
	net import	net import			

<https://doi.org/10.48366/r153801>



Figure 1. ORKG comparison of 25 scenarios from GHG studies for Germany [14].

In addition, we established an ORKG observatory on Energy System Research¹². The ORKG observatory serves as a central access point to all related curated publications, comparisons, and visualizations so that other researchers can easily explore the content. For example, Auer et al. [15] already reused the curated scientific knowledge from our two ORKG comparisons by identifying and answering further natural language competency questions from domain experts beyond the previous consideration. For this purpose, they specified the competency question as SPARQL query (see Listing 1). We executed this query on the SPARQL endpoint and visualized the results in Figure 2. In particular, these results show that average energy supply from photovoltaics and onshore wind power increased approximately fourfold from the 2006 – 2010 interval to the 2016 – 2020 interval.

¹²https://orkg.org/observatory/Energy_System_Research

[10] Karras et al.: *Organizing Scientific Knowledge From Energy System Research Using the Open Research Knowledge Graph*. 1st NFDI4Energy Conference, DOI:

[10.48550/arXiv.2401.13365](https://doi.org/10.48550/arXiv.2401.13365), 2024.

Example: ORKG Comparison of Literature – Process Variants^[11]

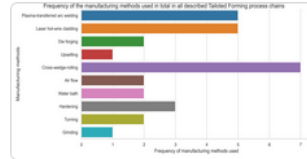
Tailored Forming Process Chain for the Manufacturing of Hybrid Components with Bearing Raceways Using Different Material Combinations ★

May 2022 Oliver Karras Laura Budde Paulina Merkel

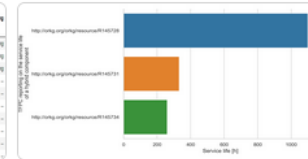
This comparison provides an overview of Tailored Forming process chains carried out in the CRC 1153 "Tailored Forming" for the manufacturing of hybrid components with bearing raceways using different material combinations. A variety of materials combinations is used for the Tailored Forming of such hybrid components. The comparison shows the combined materials, the entire Tailored Forming process chain with its individual steps, and the resulting hybrid component with its qualities. For each step, the comparison shows the measurement methods performed and their results for the individual qualities of the hybrid components. In this way, the comparison shows how different qualities of the hybrid components change during the process due to the manufacturing methods used.

DOI: <https://doi.org/10.48366/r187049>

Data analysis with Jupyter notebook



Step name	Depositor welding	Hot forming	Cooling	Heat treatment	Blowforming
Depositor welding	+	-	-	-	-
Hot forming	-	+	-	-	-
Cooling	-	-	+	-	-
Heat treatment	-	-	-	+	-
Blowforming	-	-	-	-	+



Properties	Investigation of the material combination 20MnCr5 and X45CrSi9-3 in the Tailored Forming of shafts with bearing seats	Investigation of the material combination 20MnCr5 and X45CrSi9-3 in the Tailored Forming of shafts with bearing seats	Investigation of the material combination 20MnCr5 and X45CrSi9-3 in the Tailored Forming of shafts with bearing seats	Investigation of the material combination 20MnCr5 and X45CrSi9-3 in the Tailored Forming of shafts with bearing seats	Cross-welded rolling material 1 Cladding (100Cr6)
has material/material	Air-cooled 2 Cladding Layer Component (x45CrSi9-3) - 2022				
belongs to material group*	Steel				
has type*	20MnCr5				
	X45CrSi9-3				

<https://doi.org/10.48366/r187049>



Figure 7: Comparison of TFFCs for the Manufacturing of Hybrid Components [38].

Example: ORKG Comparison of Literature: Software Features^[12]

Comparison of Hyperparameter Optimization Tools ★

January 2023 Oliver Karras Difan Deng Marius Lindauer

This comparison looks at the capabilities of various optimization tools for hyperparameter optimization, focusing on the following: Complex hyperparameter space, multi-objective, multi-fidelity, instances, command-line interface, and parallel computing. In addition, this comparison provides further information about the respective optimization tool, such as the locations of the code repository, README, software documentation and issue tracker, license and programming language.

DOI: <https://doi.org/10.48366/r281265>

Properties	Tool	BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization <i>BoTorch</i> - 2020	OpenBox: A Generalized Black-box Optimization Service <i>OpenBox</i> - 2021	BOHB: Robust and Efficient Hyperparameter Optimization at Scale <i>HpBandSter</i> - 2018	SMAC3: A Versatile Bayesian Optimization Package for Hyperparameter Optimization <i>SMAC3</i> - 2021
result/software					
↳ name*		BoTorch	OpenBox	HpBandSter	SMAC3
↳ software features*		Multi-Fidelity	Complex Hyperparameter Space	Complex Hyperparameter Space	command-line interface
		Multi-Objective	Multi-Objective	Multi-Fidelity	Complex Hyperparameter Space
		parallel computing	parallel computing	parallel computing	Instances
					Multi-Fidelity
					Multi-Objective
					parallel computing
↳ complex hyperparameter space*		×	✓	✓	✓
↳ multi-objective*		✓	✓	×	✓
↳ multi-fidelity*		✓	×		
↳ instance*		×	×		

<https://doi.org/10.48366/r281265>

Command-Line Interface

SMAC can not only be executed within a python file but also from the commandline. Consequently, not only algorithms in python can be optimized, but implementations in other languages as well.

Note

Command-line interface has been temporarily disabled in v2.0. Please fall back to v1.4 if you need it.

Comparison

The following table provides an overview of SMAC's capabilities in comparison with other optimization tools.

Package	Complex Hyperparameter Space	Multi-Objective	Multi-Fidelity	Instances	Comm
HyperMapper	✓	✓	×	×	×
Optuna	✓	✓	✓	×	✓
Hyperopt	✓	×	×	×	✓
BoTorch	×	✓	✓	×	×
OpenBox	✓	✓	×	×	×
HpBandSter	✓	×	✓	×	×
SMAC	✓	✓	✓	✓	✓

[12] Lindauer et al.: *SMAC3 Documentation*, URL: https://automl.github.io/SMAC3/v1.4.0/getting_started/package_overview.html#comparison, 2024.

Goal: ORKG Comparison on Reported Empirical Research

Comparison 4 contributions				
Overview of Reported Empirical Research in Requirements Engineering Publications from 2021				
This comparison shows an overview of reported empirical research in publications of the IEEE International Requirements Engineering Conference from the year 2021				
Properties	On the impact of using different templates on creating and understanding user stories Contribution 1 - 2021	Ambiguity and Generality in Natural Language Privacy Policies Contribution 1 - 2021	Environment-Driven Abstraction Identification for Requirements-Based Testing Contribution 1 - 2021	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
research question/research question				
↳ question*	Do different user story templates have an impact on the creation of user stories?	To what extent does the classification model reduce the manual ontology construction effort?	No question.	How much is the difference in terms of documented requirements and roles with respect to initial ideas?
	Do different user story templates have an impact on the understanding of user stories?	What is the effect of missing transitive hypernymy on classification performance?		What is the relevance given to the different categories of requirements and roles with respect to initial ideas?
↳ highlighted in text*	✓	✓	✗	✓
	✓	✓		✓
↳ hidden in text*	✗	✗	✗	✗
	✗	✗		✗
				✗
				✗
data collection method/data collection method				
↳ method*	experiment	experiment	experiment	experiment
↳ number of participants*	41	0	0	30
data analysis method/data analysis method/method*	descriptive statistics	descriptive statistics	comparative analysis	descriptive statistics
	inferential statistics	machine learning	descriptive statistics	inferential statistics
			machine learning	thematic analysis
threats to validity/threats to validity				
↳ conclusion validity*	✓	✗	✗	✗
↳ construct validity*	✓	✓	✓	✓
↳ external validity*	✓			
↳ internal validity*	✓			

<https://sandbox.orkg.org/comparison/R369109>

Topic: Reported Empirical Research

- Empirical Research
 - *Research question*
 - Question
 - Hidden in text
 - Highlighted in text
 - *Data Collection Method*
 - Method
 - Number of participants
 - *Data Analysis Method*
 - Method
 - *Threats to Validity*
 - Construct Validity
 - Internal Validity
 - External Validity
 - Conclusion Validity

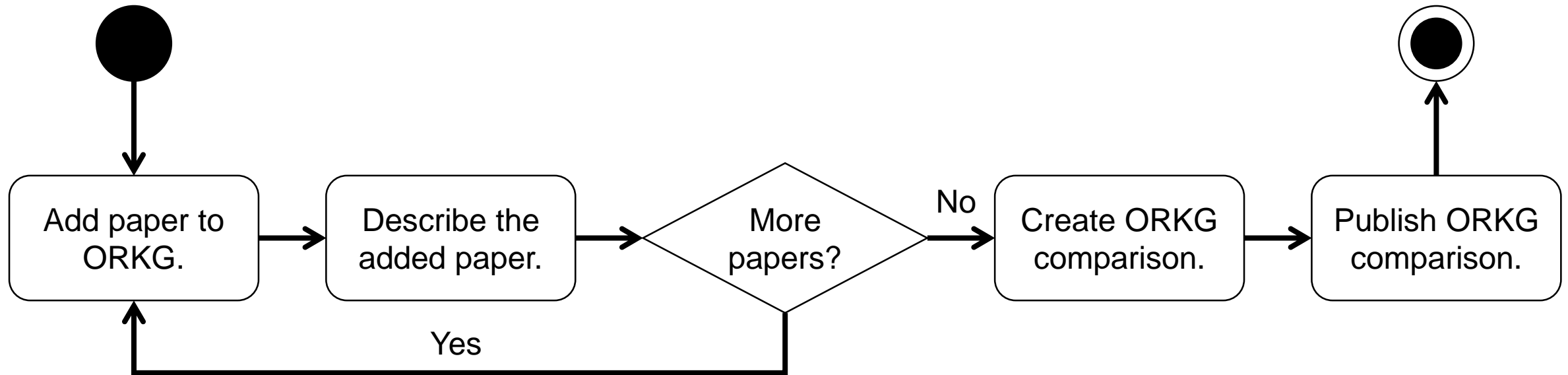
The screenshot displays the ORKG (Open Research Knowledge Graph) interface. At the top, there's a navigation bar with 'View', 'Tools', and 'About' dropdowns, a search bar, and a '+ Add new' button. Below this, the paper title 'Environment-Driven Abstraction Identification for Requirements-Based Testing' is shown, along with its publication date (September 2021), citation count (5 citations), and associated tags like 'Software Engineering'. The paper is attributed to several authors: Zedong Peng, Prachi Rathod, Nan Niu, Tanmay Bhowmik, Hui Liu, Lin Shi, and Zhi Jin. It was published in the '2021 IEEE 29th International Requirements Engineering Conference (RE)' with a DOI of <https://doi.org/10.1109/re51729.2021.00029>.

The main content area shows 'Contribution 1' with a table of data analysis methods and their corresponding results:

Method	Result
data analysis method	comparative analysis machine learning descriptive statistics
data collection method	experiment
research question	No question.
threats to validity	Construct validity: true, Internal validity: true, External validity: true Conclusion validity: false

On the right side, there's a sidebar with 'Add to comparison' and 'Assign to observatory' buttons. Below these, the 'Provenance' section shows the paper was added on 08 May 2024 by Oliver Karras. The 'Contributors' section lists Oliver Karras.

Overall Process for Creating an ORKG Comparison

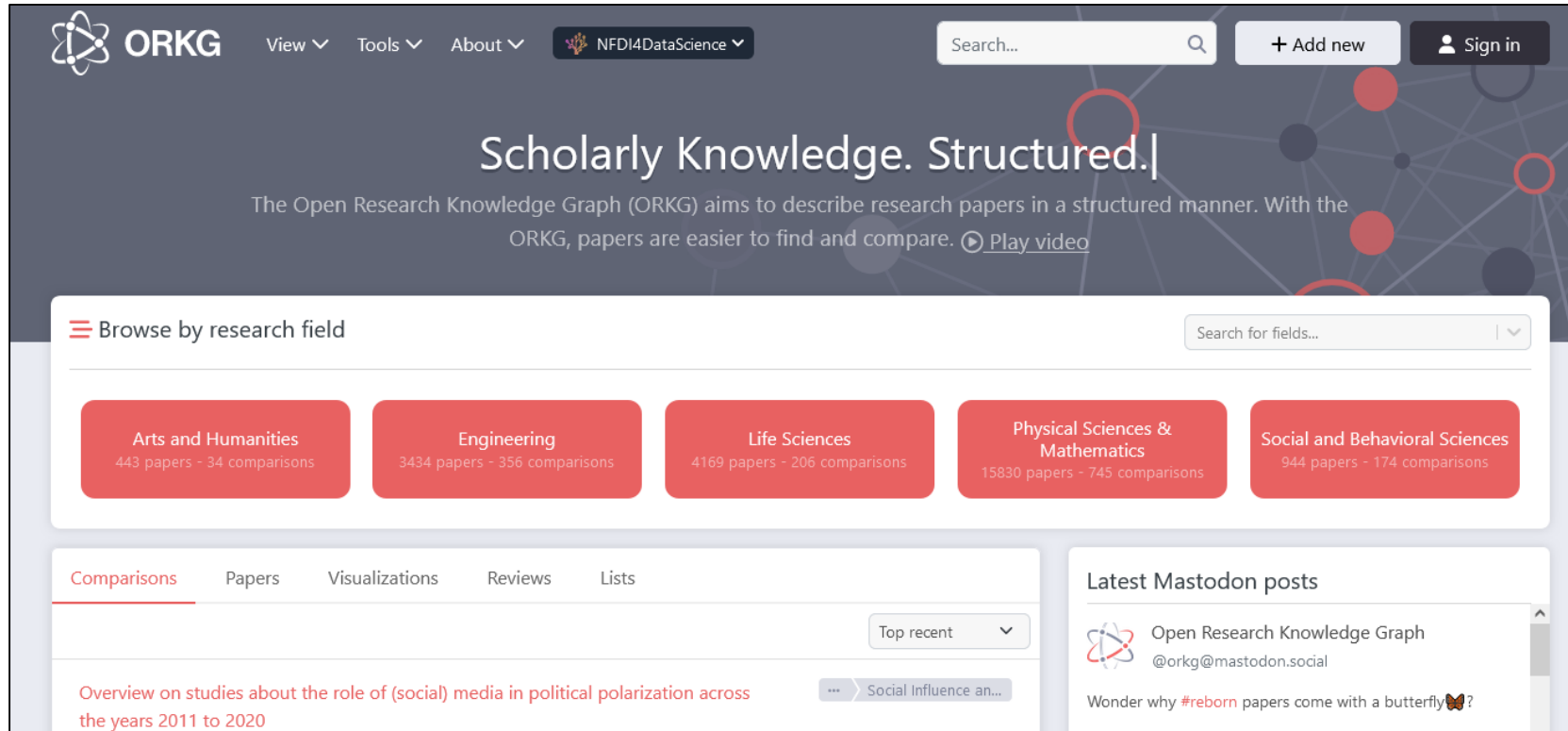


Remark:

Instead of going through the loop, we will work **collaboratively** in the tutorial as conceived by ORKG. We will take all papers from all participants and compare them with each other in an ORKG comparison.

An ORKG comparison requires **at least two** publications.

1. Open ORKG Website



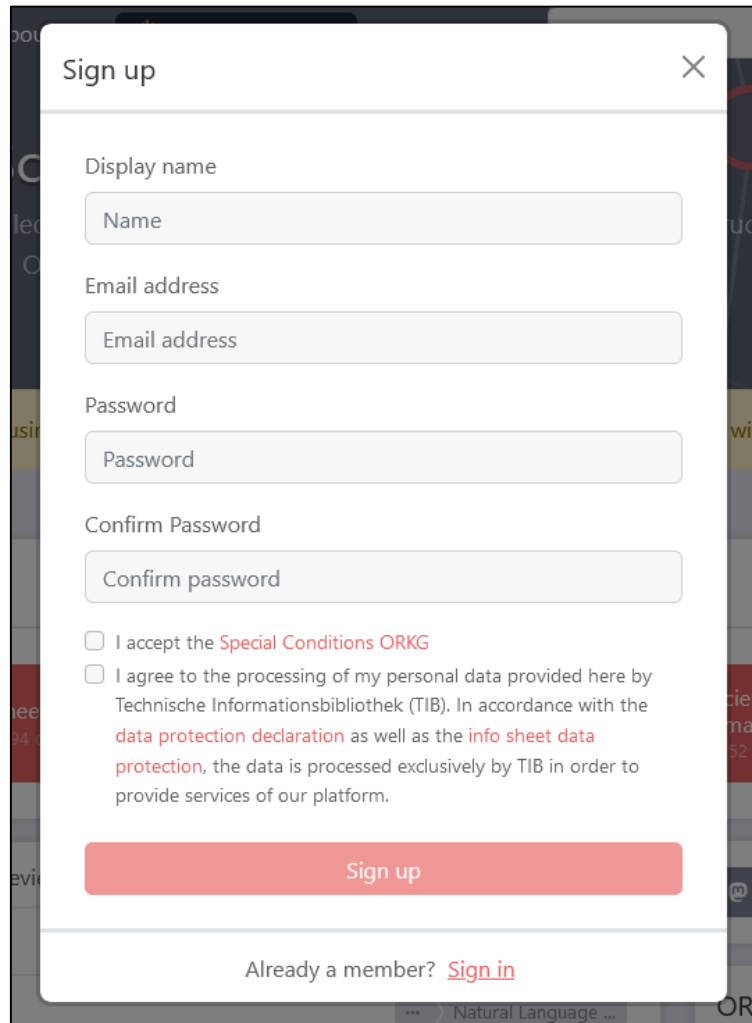
Remark:

For the tutorial, we use the **ORKG test environment** (<https://sandbox.orkg.org/>) to try everything without hesitation. You can find the official ORKG at <https://orkg.org/>.

Get your exemplary paper: <https://bit.ly/3UGFwhj>

Use <https://sandbox.orkg.org/>!
NOT <https://orkg.org/>!

2. Sign Up & Sign In



A screenshot of a 'Sign up' form. The form has a title bar with 'Sign up' and a close button. It contains four input fields: 'Display name' (with a placeholder 'Name'), 'Email address', 'Password', and 'Confirm Password' (with a placeholder 'Confirm password'). Below the fields are two checkboxes with text: 'I accept the Special Conditions ORKG' and 'I agree to the processing of my personal data provided here by Technische Informationsbibliothek (TIB). In accordance with the data protection declaration as well as the info sheet data protection, the data is processed exclusively by TIB in order to provide services of our platform.' At the bottom is a red 'Sign up' button and a link 'Already a member? Sign in'.

Sign up

Display name

Name

Email address

Email address

Password

Password

Confirm Password

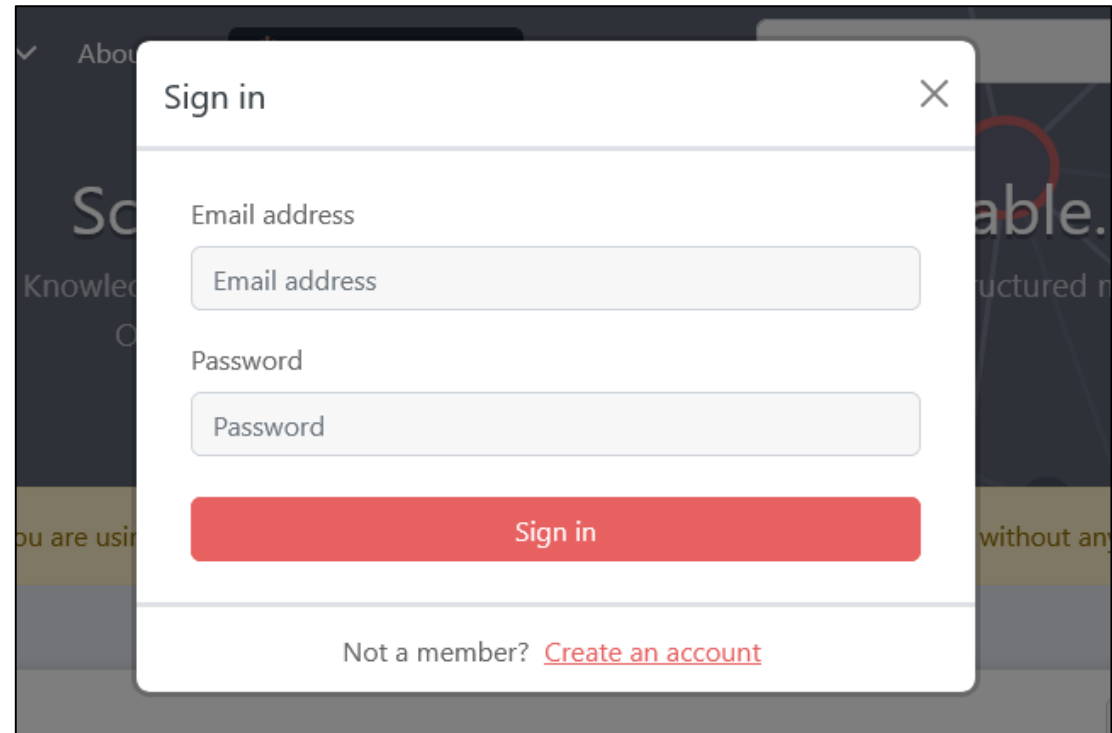
Confirm password

☐ I accept the [Special Conditions ORKG](#)

☐ I agree to the processing of my personal data provided here by Technische Informationsbibliothek (TIB). In accordance with the [data protection declaration](#) as well as the [info sheet data protection](#), the data is processed exclusively by TIB in order to provide services of our platform.

Sign up

Already a member? [Sign in](#)



A screenshot of a 'Sign in' form. The form has a title bar with 'Sign in' and a close button. It contains two input fields: 'Email address' and 'Password'. Below the fields is a red 'Sign in' button. At the bottom is a link 'Not a member? Create an account'.

Sign in

Email address

Email address

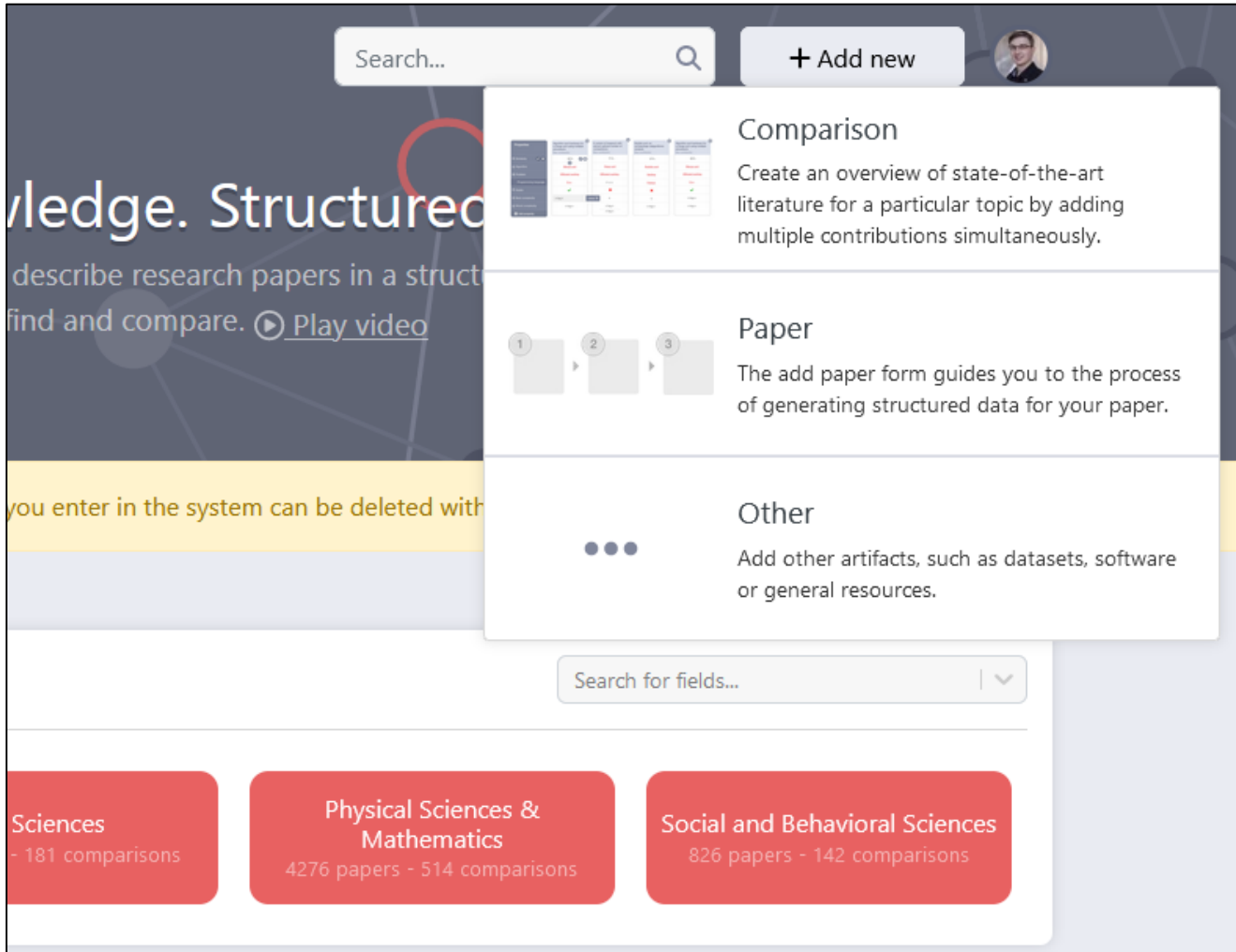
Password

Password

Sign in

Not a member? [Create an account](#)

3. Add a New Comparison to the ORKG



1. Click on “+Add new”
2. Select “Comparison”

Remark:

We could also start by adding individual papers, but since we want to create a comparison at the end, it is easy to start immediately.

4. Go to the Contribution Editor

ORKG View Tools About NFDI4DataScience Search... + Add new

Add comparison

How to make an ORKG comparison - An Example from Virology

Comparisons in ORKG provide an overview of state-of-the-art literature for a particular topic. Comparisons are dynamic and FAIR. A comparison is created from contributions, [view example of comparison](#). To create your own comparisons in ORKG, you can either import existing data (via CSV import) or start from scratch by adding your own contributions. This page guides you in creating new comparisons.

- 1. Existing data**
In case you have existing data, you can import this via the CSV import tool. This is especially helpful if you already have a large file in which related work is compared.
[Go to CSV import tool](#)
- 2. Contribution editor**
If you don't have existing data, go to the contribution editor to add contributions that will be used in the comparison. After creating contributions, you can create a comparisons.
[Go to contribution editor](#)
- 3. Publish comparison**
Once you are done editing contributions, you can create and publish a comparison. Published comparisons are persistent so they are perfectly suitable for publications.
[Publish comparison](#)

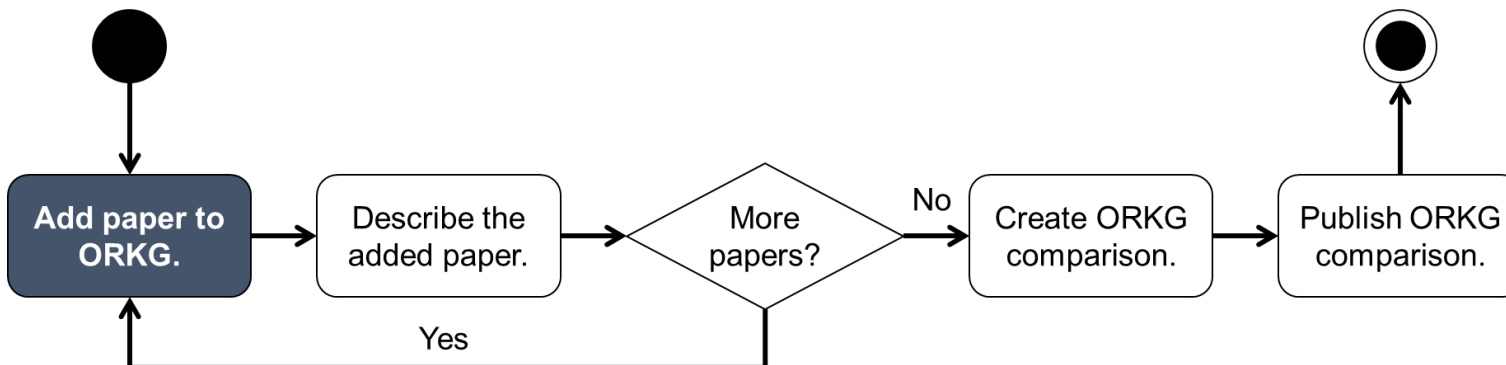
1. Click on “Go to contribution editor”

Remark:

When the data is already extracted, e.g., in a spreadsheet, we could use CSV import tool. However, the tool is a prototype and still requires a lot of manual work to map the data.

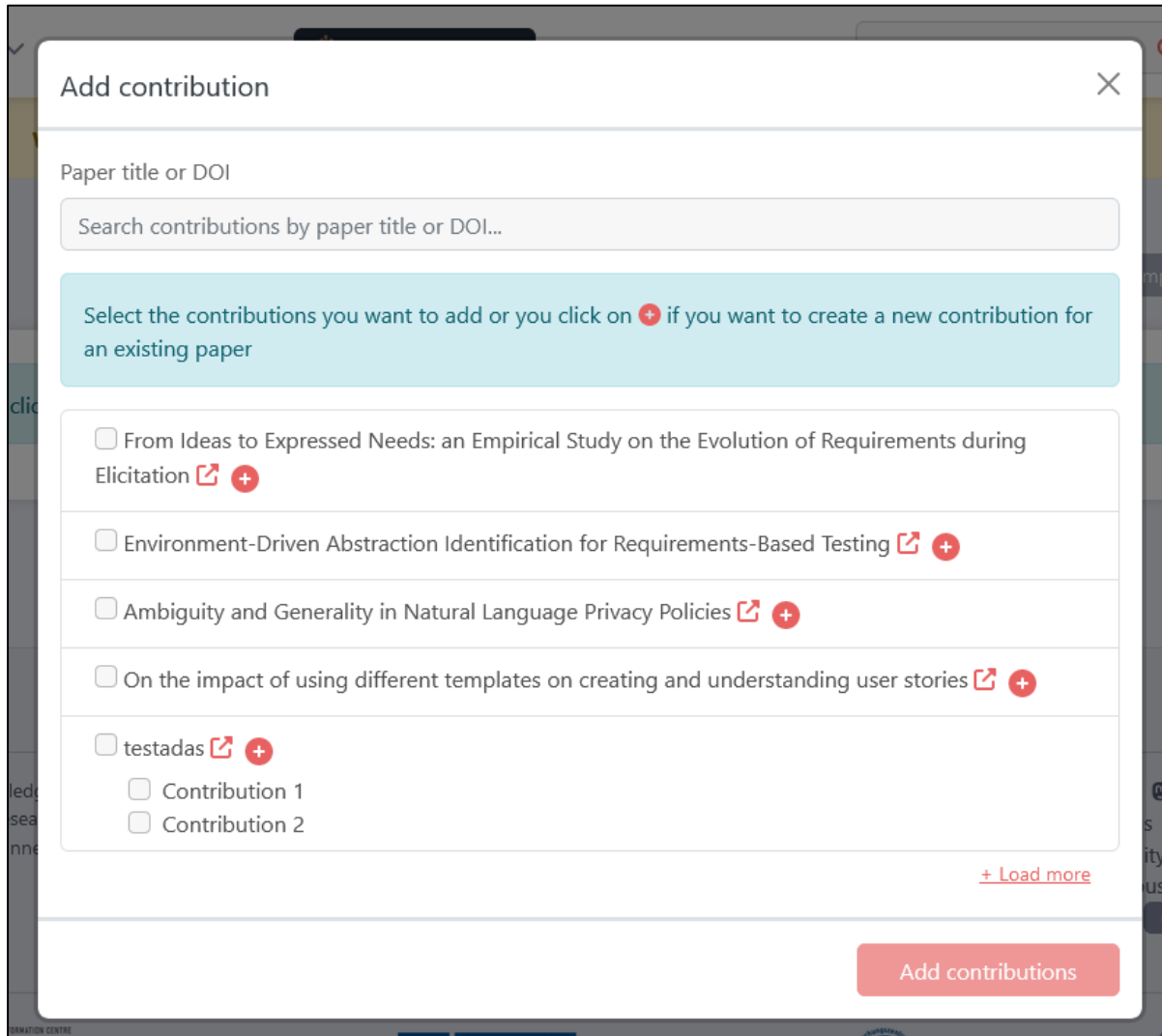
5. Add a Paper

The screenshot shows the ORKG (Open Research Knowledge Graph) Contribution editor interface. At the top, there is a navigation bar with the ORKG logo, links for View, Tools, and About, a dropdown menu for NFDI4DataScience, a search bar, and a red '+ Add new' button. A yellow warning banner states: 'Warning: You are using a testing environment. Data you enter in the system can be deleted without any notice.' Below the banner, the main area is titled 'Contribution editor' and contains a light blue instruction box: 'Start adding contributions by clicking the button *Add contribution* on the right'. On the right side of the editor, there are two buttons: 'View comparison' and '+ Add contribution'.



1. Click on “Add contribution”

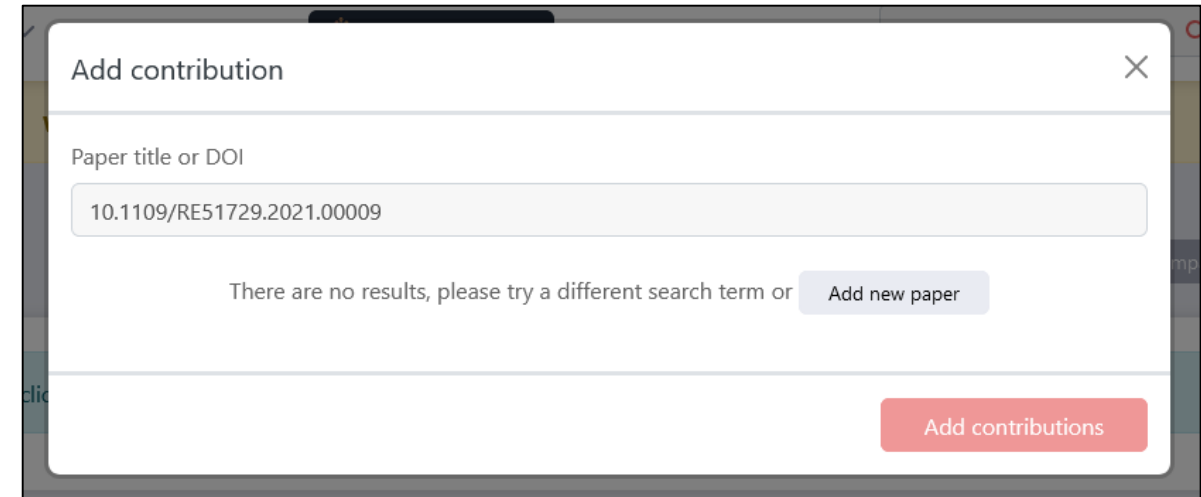
5. Add a Paper: Enter DOI (or Paper Title)



The 'Add contribution' dialog box features a search bar with the placeholder text 'Search contributions by paper title or DOI...'. Below the search bar is a light blue instruction box: 'Select the contributions you want to add or you click on + if you want to create a new contribution for an existing paper'. A list of search results follows, each with a checkbox, a document icon, and a red '+' button. The results are:

- ☐ From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation
- ☐ Environment-Driven Abstraction Identification for Requirements-Based Testing
- ☐ Ambiguity and Generality in Natural Language Privacy Policies
- ☐ On the impact of using different templates on creating and understanding user stories
- ☐ testadas
 - ☐ Contribution 1
 - ☐ Contribution 2

A '+ Load more' link is located at the bottom right of the list. An 'Add contributions' button is at the bottom right of the dialog.



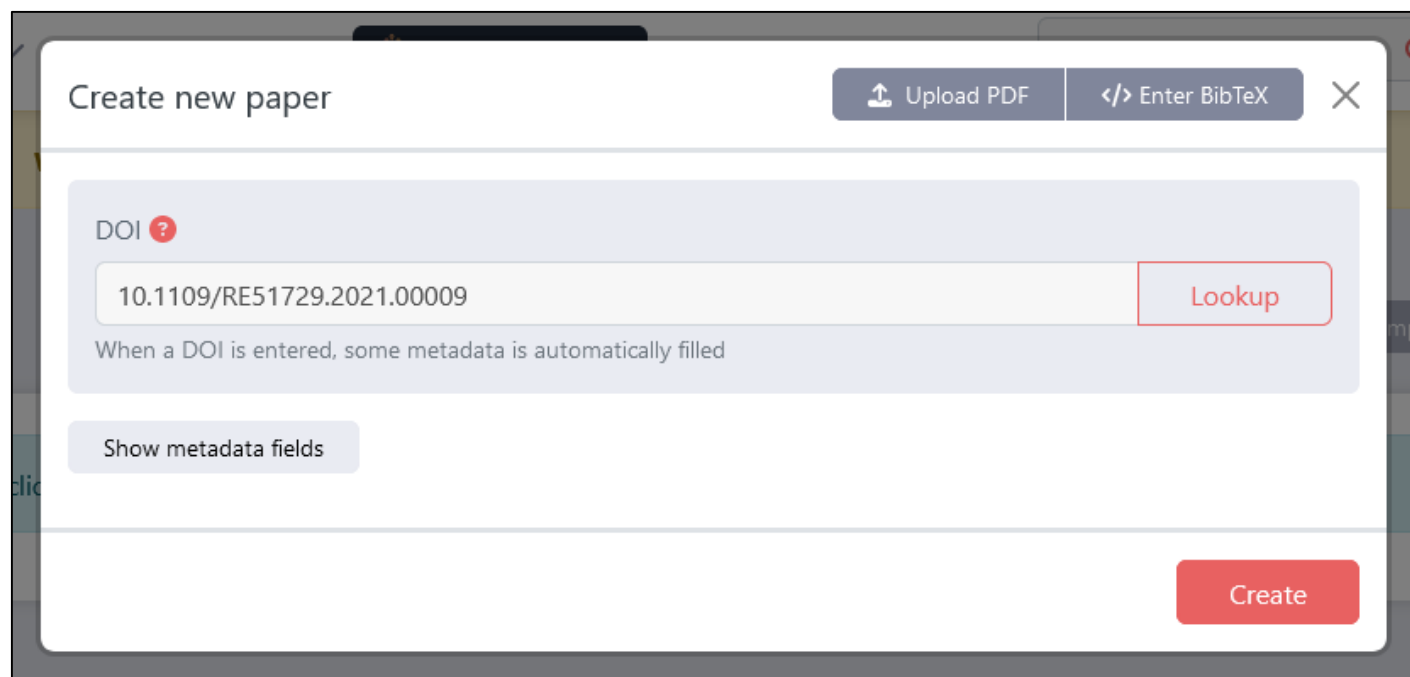
The 'Add contribution' dialog box shows the search bar with the text '10.1109/RE51729.2021.00009'. Below the search bar, it displays the message 'There are no results, please try a different search term or' followed by an 'Add new paper' button. An 'Add contributions' button is located at the bottom right of the dialog.

1. Click on “Add new paper”

Remark:

If the paper is already in the ORKG, it is shown and we can either reuse an existing contribution or create a new one.

5. Add a Paper: Lookup Paper by DOI

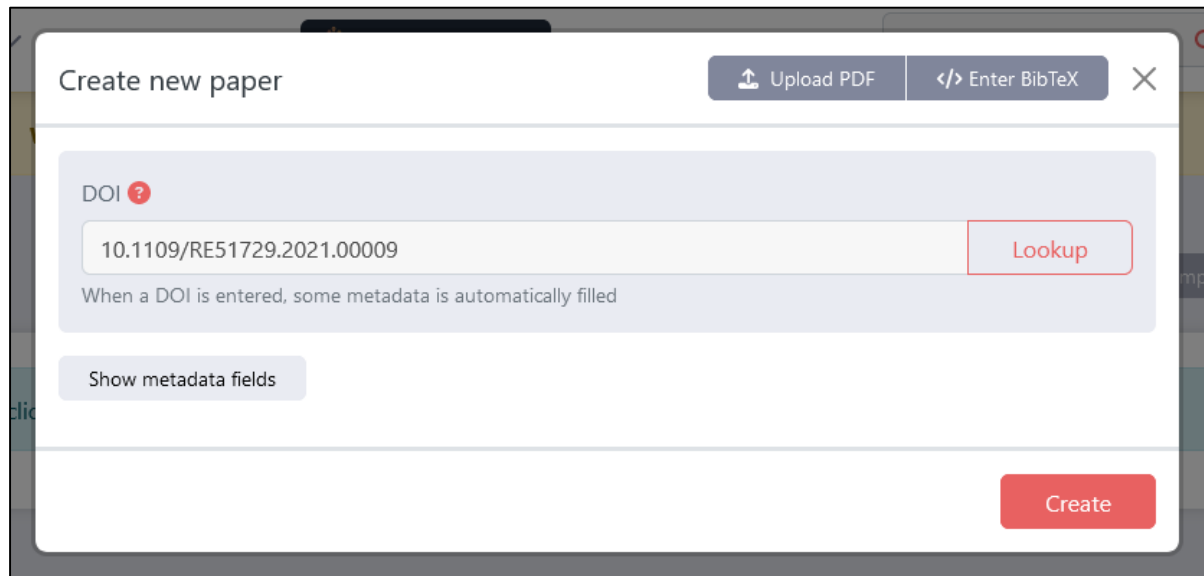


The screenshot shows a web interface for creating a new paper. At the top, there's a header bar with the title 'Create new paper' and two buttons: 'Upload PDF' and 'Enter BibTeX'. Below this, there's a section for entering a DOI. It includes a label 'DOI' with a red question mark icon, a text input field containing '10.1109/RE51729.2021.00009', and a red 'Lookup' button. Below the input field, a note states: 'When a DOI is entered, some metadata is automatically filled'. Underneath this note is a button labeled 'Show metadata fields'. At the bottom right of the dialog, there is a red 'Create' button.

Remark:

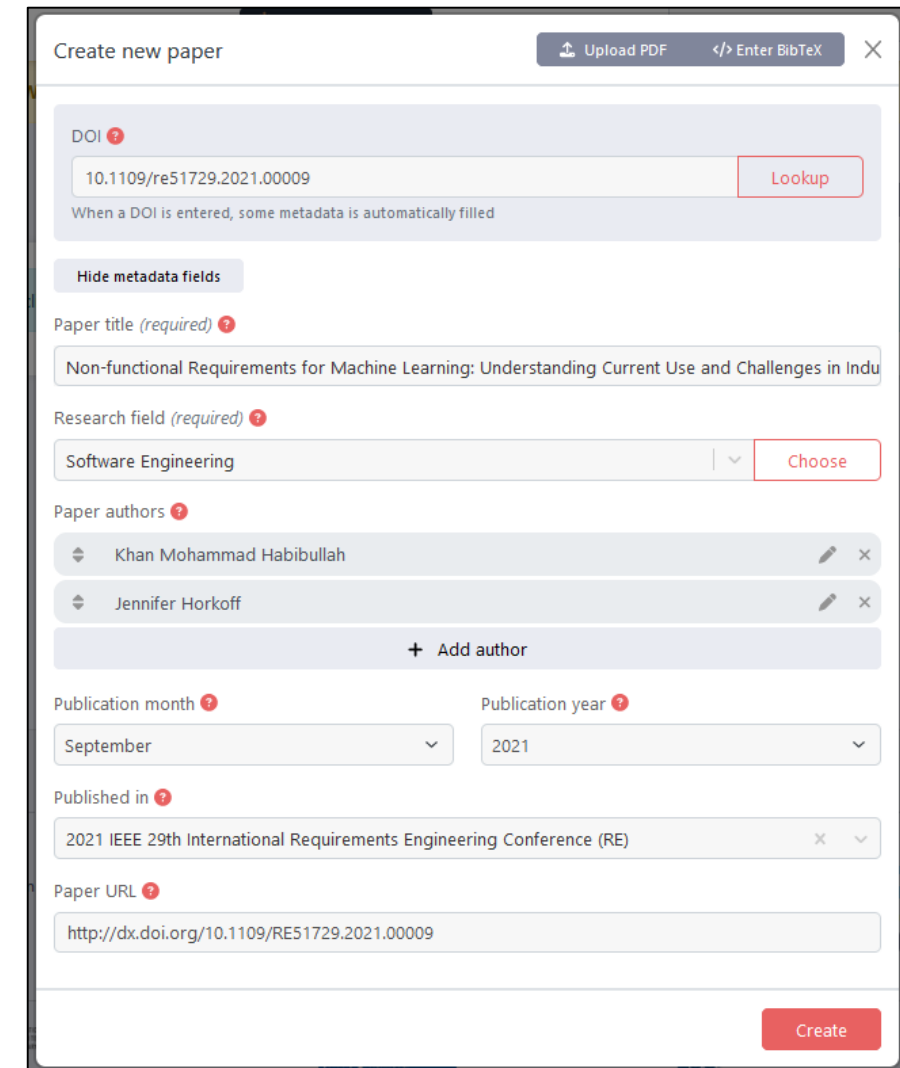
ORKG attempts to retrieve the metadata automatically. If the retrieval is not successful, you can also manually enter or correct the metadata.

5. Add a Paper: Lookup Paper by DOI



The screenshot shows a 'Create new paper' dialog box. At the top, there are two buttons: 'Upload PDF' and 'Enter BibTeX'. Below these, there is a section for 'DOI' with a red question mark icon. A text input field contains the DOI '10.1109/RE51729.2021.00009'. To the right of the input field is a red 'Lookup' button. Below the input field, a message states: 'When a DOI is entered, some metadata is automatically filled'. At the bottom left, there is a 'Show metadata fields' button. At the bottom right, there is a red 'Create' button.

1. Click on “Lookup”
2. Check metadata fetched
3. Add a “Research field”. We use “Software Engineering”.
4. Click on “Create”



The screenshot shows the 'Create new paper' dialog box after the DOI lookup. The 'DOI' field is filled with '10.1109/re51729.2021.00009' and a red 'Lookup' button is visible. Below this, a 'Hide metadata fields' button is present. The 'Paper title (required)' field is filled with 'Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Indu'. The 'Research field (required)' dropdown menu is set to 'Software Engineering' with a red 'Choose' button. The 'Paper authors' section lists 'Khan Mohammad Habibullah' and 'Jennifer Horkoff', with an '+ Add author' button below. The 'Publication month' dropdown is set to 'September' and the 'Publication year' dropdown is set to '2021'. The 'Published in' dropdown is set to '2021 IEEE 29th International Requirements Engineering Conference (RE)'. The 'Paper URL' field is filled with 'http://dx.doi.org/10.1109/RE51729.2021.00009'. A red 'Create' button is at the bottom right.

6. Describe a Paper

Contribution editor

View comparison

+ Add contribution

Properties

✓ Saved

Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry

Contribution 1

Start adding properties or use templates by using the buttons below

+ Add property

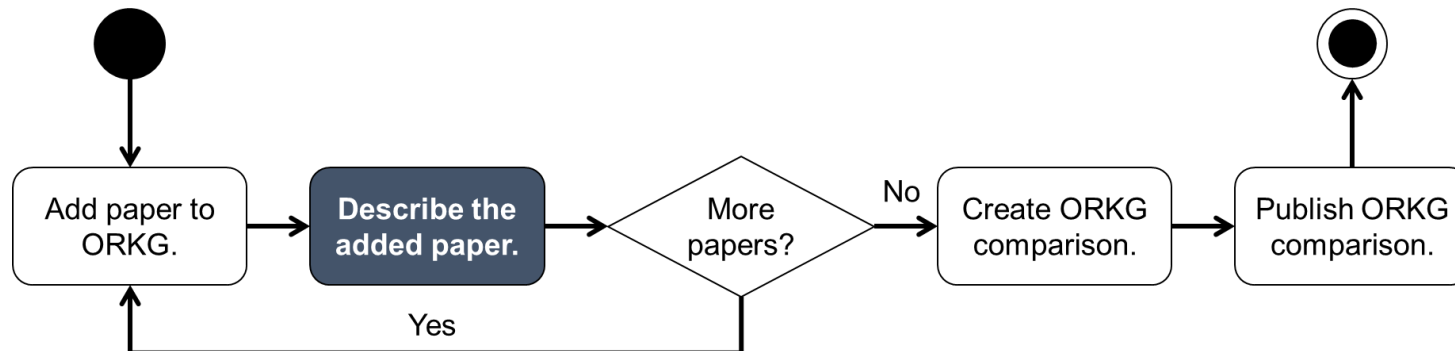
💡

+ Templates

Suggested properties

+ research problem

Problem



Remark:

Instead of adding all properties individually, we use an ORKG template.

6. Describe a Paper: ORKG Templates

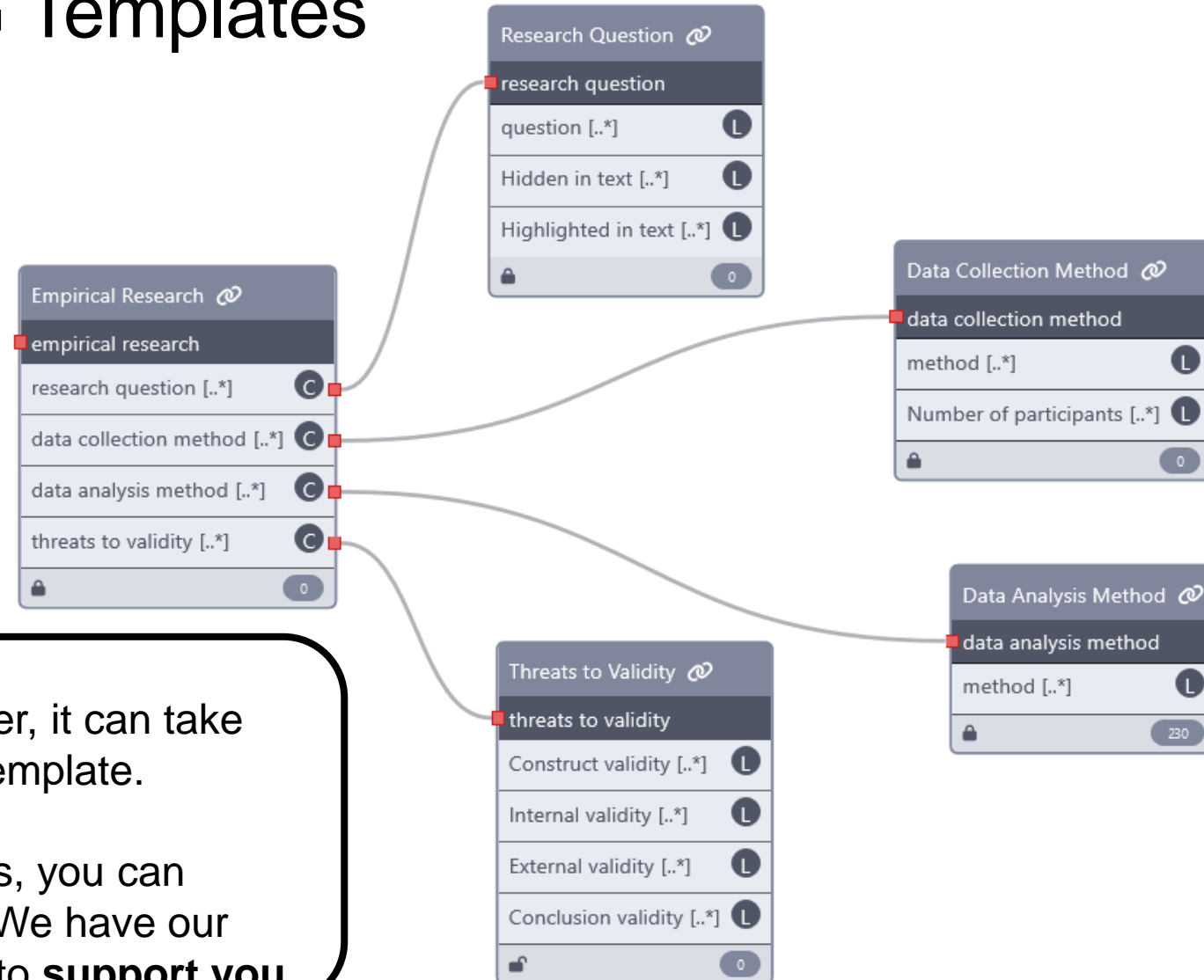
An **ORKG template** specifies the **structure** and **terminology** used to describe a **publication** in the ORKG. Its use **ensures FAIR data** that is **consistent** and **comparable** across publications.

<https://sandbox.orkg.org/template/R369028>

Remark:

Every user can create ORKG templates. However, it can take some time and practice to build a good ORKG template.

If you want to use the feature and have problems, you can always contact the ORKG team, especially me. We have our own **Curation and Community Building team** to **support you**.



6. Describe a Paper: Add Template “Empirical Research”

Contribution editor

View comparison

+ Add contribution

Properties

✓ Saved

Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry

Contribution 1

Start adding properties or use templates by using the buttons below

+ Add property

+ Templates

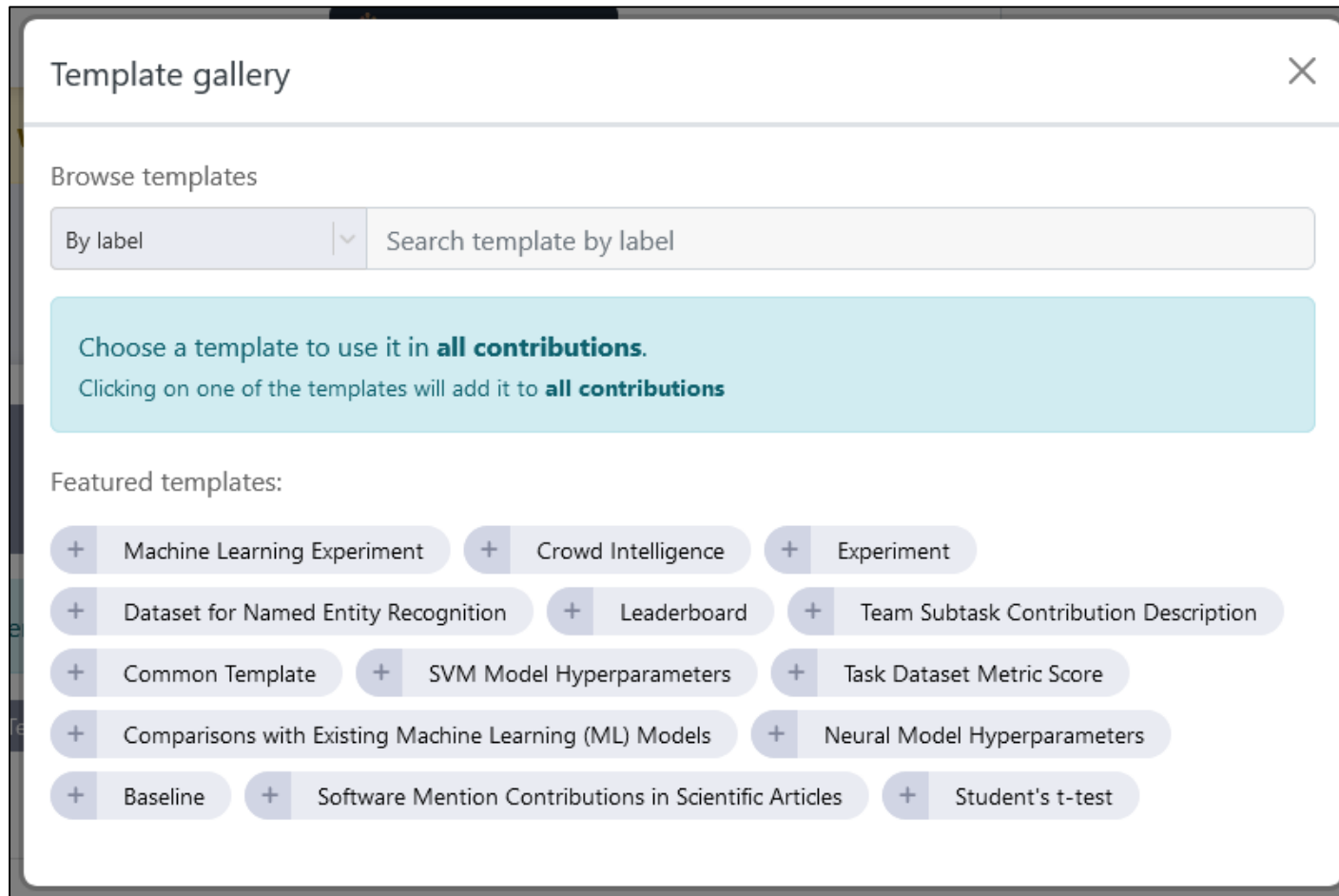
Suggested properties

+ research problem

Problem

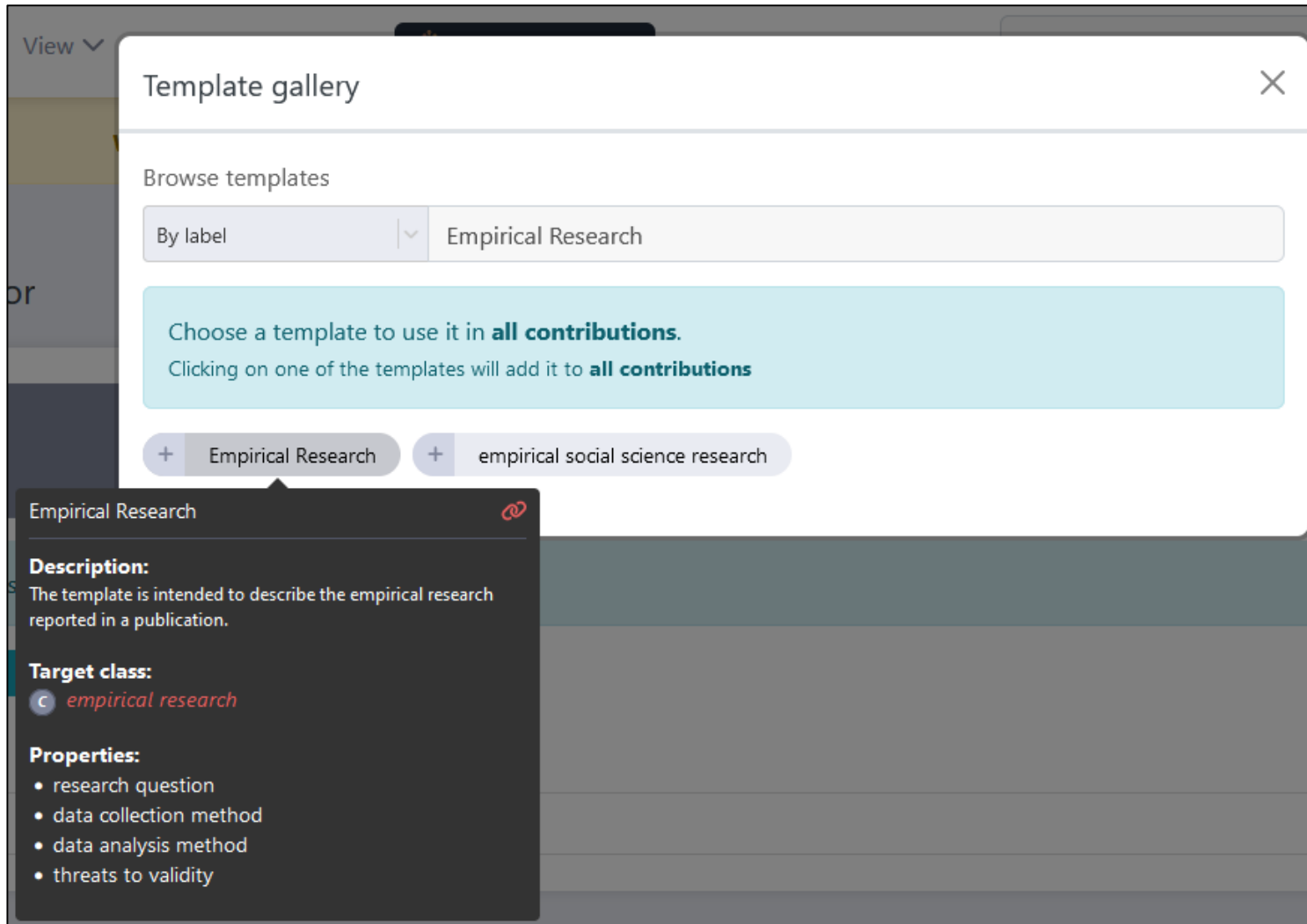
1. Click on “Templates”

6. Describe a Paper: Add Template “Empirical Research”



1. Search for “Empirical Research”

6. Describe a Paper: Add Template “Empirical Research”



1. Select “Empirical Research”

6. Describe a Paper: Fill out the Table

Contribution editor

View comparison

+ Add contribution

Properties

✓ Saved

data analysis method

data collection method

research question

threats to validity

+ Add property

💡

+ Templates

Suggested properties

+ research problem

Problem

Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry

Contribution 1

Add value

+

1. Click “Add value” for adding the data analysis method

6. Describe a Paper: Enter the Data Analysis Method(s)

The screenshot shows the ORKG Contribution editor interface. A modal dialog box titled "View existing resource: Data Analysis Method" is open. The dialog has a close button (X) and a link "Open resource" with an external link icon. Inside the dialog, there are tabs for "Templates", "Help", and "Preferences". Below the tabs, it says "Applied template: Data Analysis Method". There is a text input field containing the word "method" and a plus sign button (+) to its right. At the bottom of the dialog is a button labeled "+ Add property". In the background, the "Contribution editor" is visible, showing a list of properties: "data analysis method", "data collection method", "research question", and "threats to validity". There is also a "Suggested properties" section with a button for "+ research problem".

1. Click the +
2. Enter method

6. Describe a Paper: Enter the Data Analysis Method(s)

The screenshot shows the ORKG Contribution editor interface. A modal dialog titled "View existing resource: Data Analysis Method" is open. The dialog has a header with "Open resource" and a close button. Below the header are tabs for "Templates", "Help", and "Preferences". The main content area shows "Applied template: Data Analysis Method". There is a text input field labeled "method" containing the text "thematic analysis". To the right of the input field are buttons for "Text" (a dropdown), a lightbulb icon, "Cancel", and "Create". At the bottom of the dialog is a button labeled "+ Add property". The background shows the "Contribution editor" with a "Properties" panel on the left containing links like "data analysis method", "data collection method", "research question", and "threats to validity". There is also a "Suggested properties" section at the bottom with a "+ research problem" button.

1. Click “Create”
2. Close dialog

6. Describe a Paper: Enter the Data Analysis Method(s)

The screenshot shows the ORKG (Open Research Knowledge Graph) Contribution editor. At the top, there's a navigation bar with the ORKG logo, links for View, Tools, and About, a user profile icon, and a search bar. A yellow warning banner states: "Warning: You are using a testing environment. Data you enter in the system can be deleted without any notice." Below this, the "Contribution editor" section has buttons for "View comparison" and "Add contribution". The main area is divided into a left sidebar with a "Properties" list (including "data analysis method", "data collection method", "research question", and "threats to validity") and a right pane for editing the selected property. The "data analysis method" property is currently selected and shows the title "Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry" and the value "thematic analysis". At the bottom, there are buttons for "Add property", "Templates", and a "Suggested properties" section with a "research problem" suggestion.

ORKG View Tools About NFDI4DataScience Search... + Add new

Warning: You are using a testing environment. Data you enter in the system can be deleted without any notice.

Contribution editor View comparison + Add contribution

Properties ✓ Saved

data analysis method

data collection method

research question

threats to validity

+ Add property Templates

Suggested properties

+ research problem Problem

Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry

Contribution 1

thematic analysis

If you have more analysis methods, **repeat** the process.

Otherwise, continue with the next property.

6. Describe a Paper: Enter the Data Collection Method(s)

The screenshot shows the ORKG Contribution editor interface. A modal dialog box titled "View existing resource: Data Collection Method" is open. The dialog has a header with "Open resource" and a close button. Below the header are tabs for "Templates", "Help", and "Preferences". The main content area shows the "Applied template: Data Collection Method" with two input fields: "method" and "Number of participants", each with a plus icon to its right. At the bottom of the dialog is a "+ Add property" button. The background shows the "Contribution editor" with a "Properties" panel on the left containing links like "data analysis method", "data collection method", "research question", and "threats to validity". At the bottom of the background interface are buttons for "+ Add property", "Templates", and a "Suggested properties" section with a "+ research problem" button.

1. Click the +
2. Enter method
3. Click the +
4. Enter number of participants

6. Describe a Paper: Enter the Data Collection Method(s)

The screenshot shows the ORKG Contribution editor interface. A modal dialog box titled "View existing resource: Data Collection Method" is open. The dialog has a header with "Open resource" and a close button. Below the header are tabs for "Templates", "Help", and "Preferences". The main content area shows the "Applied template: Data Collection Method". There are two input fields: "method" with a dropdown set to "Text" and a value of "interview", and "Number of participants" with a dropdown set to "Integer" and a value of "10". Each input field has a lightbulb icon, a "Cancel" button, and a "Create" button. At the bottom of the dialog is a "+ Add property" button. The background shows the "Contribution editor" with a "Properties" panel on the left containing links for "data analysis method", "data collection method", "research question", and "threats to validity". At the bottom of the background, there is a "Suggested properties" section with a "+ research problem" button and a "Problem" button.

1. Click “Create”
2. Click “Create”
3. Close dialog

6. Describe a Paper: Enter the Data Collection Method(s)

The screenshot shows the ORKG (Open Research Knowledge Graph) Contribution editor. The top navigation bar includes the ORKG logo, menu items (View, Tools, About), a user profile icon, and a search bar. A yellow warning banner states: "Warning: You are using a testing environment. Data you enter in the system can be deleted without any notice." The main area is titled "Contribution editor" and contains a table with properties and their values. The "data collection method" property is highlighted in blue. Below the table, there are buttons for "Add property", "Templates", and "Suggested properties".

ORKG View Tools About NFDI4DataScience Search... + Add new

Warning: You are using a testing environment. Data you enter in the system can be deleted without any notice.

Contribution editor View comparison + Add contribution

Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry
data analysis method	thematic analysis
data collection method	interview
research question	
threats to validity	

+ Add property Templates

Suggested properties

+ research problem Problem

If you have more collection methods, **repeat** the process.

Otherwise, continue with the next property.

6. Describe a Paper: Enter the Research Question(s)

The screenshot shows the ORKG Contribution editor interface. A modal dialog titled "View existing resource: Research Question" is open. The dialog has a header with "Open resource" and a close button. Below the header are tabs for "Templates", "Help", and "Preferences". The main content area shows the applied template "Research Question" and a list of properties: "question", "Hidden in text", and "Highlighted in text", each with a "+" button to its right. At the bottom of the dialog is a "+ Add property" button. The background shows the "Contribution editor" with a "Properties" sidebar containing links like "data analysis method", "data collection method", "research question", and "threats to validity". At the bottom of the editor are buttons for "+ Add property", "Templates", and "Suggested properties" with a search bar containing "+ research problem".

1. Click the +
2. Enter question
3. Click the +
4. Enter, if the question is hidden in text
5. Click the +
6. Enter, if the question is highlighted in text

6. Describe a Paper: Enter the Research Question(s)

The screenshot shows the ORKG Contribution editor interface. A modal dialog box titled "View existing resource: Research Question" is open. The dialog has a header bar with "Open resource" and a close button. Below the header, there are tabs for "Templates", "Help", and "Preferences". The main content area shows the "Applied template: Research Question". There are three input fields with "Cancel" and "Create" buttons:

- question**: Text input with the value "nt treatment of NFRs in ML in industry?".
- Hidden in text**: Boolean input with the value "False".
- Highlighted in text**: Boolean input with the value "True".

At the bottom of the dialog is a "+ Add property" button. The background shows the "Contribution editor" with a "Properties" sidebar containing links like "data analysis method", "data collection method", "research question", and "threats to validity". There is also a "Suggested properties" section with a "+ research problem" button.

1. Click “Create”
2. Click “Create”
3. Click “Create”
4. Close dialog

6. Describe a Paper: Enter the Research Question(s)

The screenshot shows the ORKG (Open Research Knowledge Graph) Contribution editor. The top navigation bar includes the ORKG logo, menu items (View, Tools, About), a user profile dropdown (NFDI4DataScience), a search bar, and a '+ Add new' button. A yellow warning banner states: 'Warning: You are using a testing environment. Data you enter in the system can be deleted without any notice.'

The main area is titled 'Contribution editor' and contains a 'View comparison' button and an '+ Add contribution' button. The central form is divided into two panels. The left panel, titled 'Properties', lists several properties: 'data analysis method', 'data collection method', 'research question', and 'threats to validity'. The 'research question' property is currently selected. The right panel shows the details for 'Contribution 1', titled 'Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry'. It displays the selected property 'research question' with the value 'What is the perception and current treatment of NFRs in ML in industry?'. Below the main form, there are buttons for '+ Add property', a lightbulb icon, and '+ Templates'. At the bottom, a 'Suggested properties' section shows a '+ research problem' button and a 'Problem' button.

If you have more research questions, **repeat** the process.

Otherwise, continue with the next property.

6. Describe a Paper: Enter the Threats to Validity

The screenshot shows the ORKG Contribution editor interface. A modal window titled 'View existing resource: Threats to Validity' is open, displaying a template for entering threats to validity. The template includes a list of four categories, each with a plus icon for adding a new entry: Construct validity, Internal validity, External validity, and Conclusion validity. The background shows the 'Contribution editor' with a 'Properties' panel on the left containing links like 'data analysis method', 'data collection method', 'research question', and 'threats to validity'. At the bottom, there are buttons for '+ Add property' and '+ Add template', and a 'Suggested properties' section with a '+ research problem' button.

View existing resource: Threats to Validity [Open resource](#) ✕

Templates Help Preferences

Applied template: Threats to Validity

Construct validity	+
Internal validity	+
External validity	+
Conclusion validity	+

+ Add property

Suggested properties

+ research problem




1. Click the +
2. Enter construct validity
3. Click the +
4. Enter internal validity
5. Click the +
6. Enter external validity
7. Click the +
8. Enter conclusion validity

6. Describe a Paper: Enter the Threats to Validity

The screenshot shows the ORKG Contribution editor interface. A modal dialog titled "View existing resource: Threats to Validity" is open. The dialog has a header with "Open resource" and a close button. Below the header are tabs for "Templates", "Help", and "Preferences". The main content area shows the "Applied template: Threats to Validity". There are four rows of form fields, each with a label, a "Boolean" dropdown, a "True" value, a dropdown arrow, a lightbulb icon, and "Cancel" and "Create" buttons. The rows are: "Construct validity", "Internal validity", "External validity", and "Conclusion validity". At the bottom of the dialog is a "+ Add property" button with a lightbulb icon. The background shows the "Contribution editor" with a "Properties" panel on the left containing links like "data analysis method", "data collection method", "research question", and "threats to validity". There is also a "Suggested properties" section at the bottom with a "+ research problem" button.

1. Click “Create”
2. Click “Create”
3. Click “Create”
4. Click “Create”
5. Close dialog

6. Describe a Paper: Enter the Threats to Validity

 View ▾ Tools ▾ About ▾ NFDI4DataScience ▾ Search...  + Add new 

Warning: You are using a testing environment. Data you enter in the system can be deleted without any notice.

Contribution editor View comparison + Add contribution


Properties ✓ Saved

[data analysis method](#)

[data collection method](#)

[research question](#)

[threats to validity](#)

+ Add property  + Templates

Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry

Contribution 1

[thematic analysis](#)

[interview](#)

[What is the perception and current treatment of NFRs in ML in industry?](#)

[Construct validity: true, Internal validity: true, External validity: true, Conclusion validity: true](#)

Suggested properties

+ research problem

Problem

Now, we have described our first paper regarding its reported empirical research.

We can repeat the entire process for the next paper or work collaboratively!

7. Add the Next Paper

Contribution editor

View comparison Add contribution

Properties ✓ Saved

[data analysis method](#)

[data collection method](#)

[research question](#)

[threats to validity](#)

+ Add property

Suggested properties

+ research problem

Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry

Contribution 1

[thematic analysis](#)

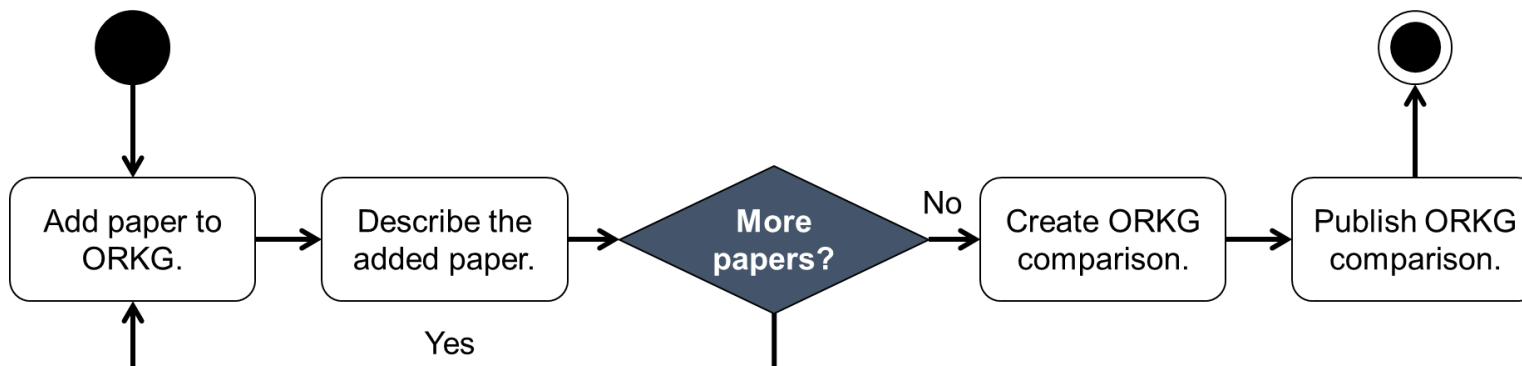
[interview](#)

[What is the perception and current treatment of NFRs in ML in industry?](#)

[Construct validity: true, Internal validity: true, External validity: true, Conclusion validity: true](#)

Problem

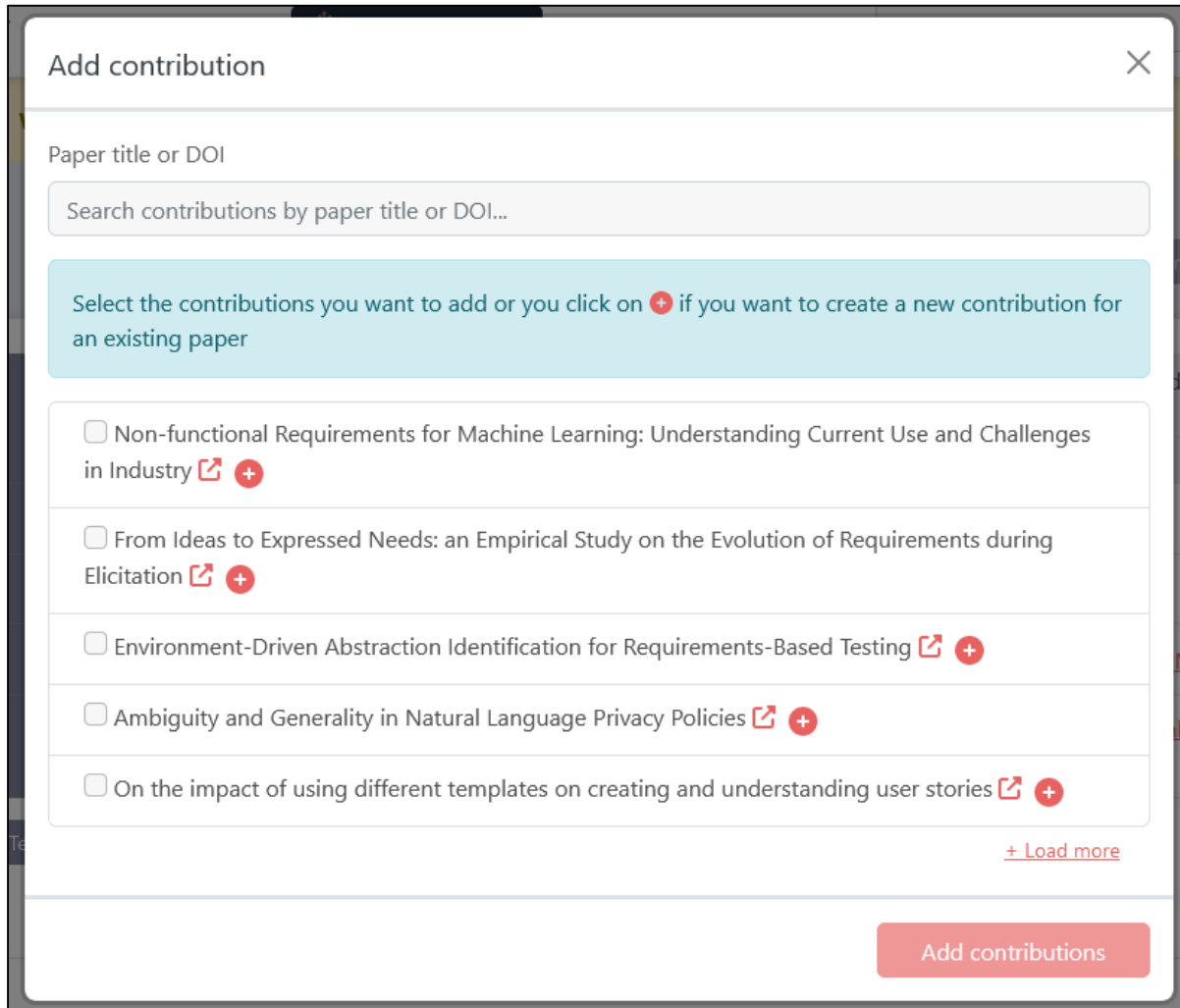
1. Click on “Add contribution”



Remark:

Now, we use a paper called **“From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation”** already described with the ORKG template.

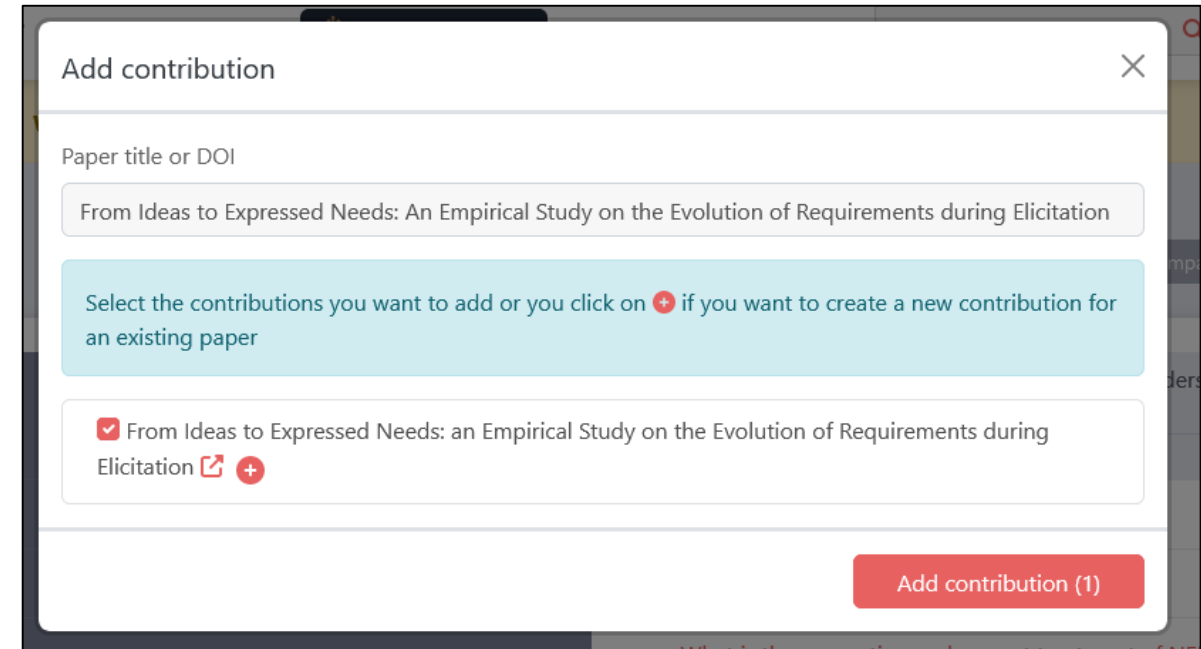
7. Add the Next Paper: Enter Paper Title



The screenshot shows a modal window titled "Add contribution" with a close button (X) in the top right corner. Below the title is a section labeled "Paper title or DOI" containing a search bar with the placeholder text "Search contributions by paper title or DOI...". Below the search bar is a light blue instruction box: "Select the contributions you want to add or you click on + if you want to create a new contribution for an existing paper". Below this is a list of five papers, each with an unchecked checkbox, a paper title, a link icon, and a red plus icon:

- ☐ Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry
- ☐ From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation
- ☐ Environment-Driven Abstraction Identification for Requirements-Based Testing
- ☐ Ambiguity and Generality in Natural Language Privacy Policies
- ☐ On the impact of using different templates on creating and understanding user stories

At the bottom right of the list is a red link: "+ Load more". At the bottom center of the modal is a red button labeled "Add contributions".



This screenshot shows the same "Add contribution" modal window, but with the second paper selected. The checkbox for "From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation" is now checked. The red button at the bottom right now says "Add contribution (1)".

1. Enter paper title
2. Select the checkbox
3. Click "Add contribution (1)"

7. Add the Next Paper: Result

Contribution editor
View comparison
Add contribution

Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation
	Contribution 1	Contribution 1
data analysis method	thematic analysis	thematic analysis inferential statistics descriptive statistics
data collection method	interview	experiment
research question	What is the perception and current treatment of NFRs in ML in industry?	What is the relevance given to the different categories of requirements and roles with respect to initial ideas? What is the relevance given to the different categories of requirements and roles with respect to initial ideas? How much is the difference in terms of documented requirements and roles with respect to initial ideas?
threats to validity	Construct validity: true, Internal validity: true, External validity: true, Conclusion validity: true	Construct validity: true, Internal validity: true, External validity: true, Conclusion validity: false

Now, we have two described papers.

If we want, we can add further ones or we can create an ORKG comparison.

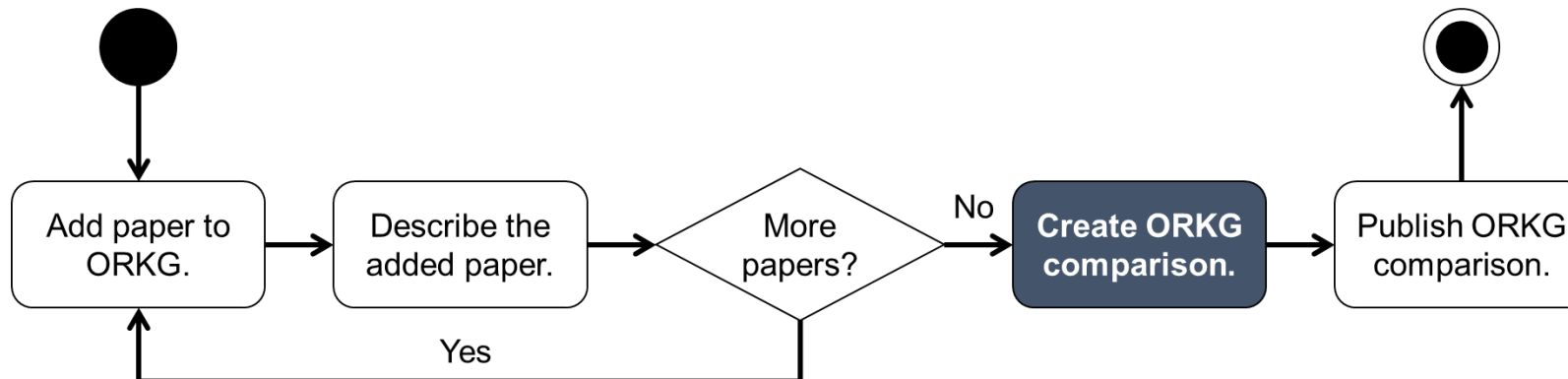
8. Create ORKG Comparison

Contribution editor

View comparison Add contribution

Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation
	Contribution 1	Contribution 1
data analysis method	thematic analysis	thematic analysis
		inferential statistics
		descriptive statistics
data collection method	interview	experiment
research question	What is the perception and current treatment of NFRs in ML in industry?	What is the relevance given to the different categories of requirements and roles with respect to initial Ideas?
		What is the relevance given to the different categories of requirements and roles with respect to initial Ideas?
		How much is the difference in terms of documented requirements and roles with respect to initial ideas?
threats to validity	Construct validity: true, Internal validity: true, External validity: true, Conclusion validity: true	Construct validity: true, Internal validity: true, External validity: true, Conclusion validity: false

1. Click on “View comparison”



8. Create ORKG Comparison: Result and Improvements

Comparison 2 contributions		
Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry Contribution 1 - 2021	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
data.analysis.method	Data Analysis Method	Data Analysis Method
		Data Analysis Method
		Data Analysis Method
data.analysis.method/data.analysis.method/method*	thematic analysis	descriptive statistics
		inferential statistics
		thematic analysis
data.collection.method	Data Collection Method	Data Collection Method
data.collection.method/data.collection.method		
method*	interview	experiment
number of participants*	10	30
research question	Research Question	Research Question
		Research Question
		Research Question
research question/research question		
hidden in text*	✗	✗
		✗
		✗
highlighted in text*	✓	✓
		✓
		✓
question*	What is the perception and current treatment of NFRs in ML in industry?	How much is the difference in terms of documented requirements and roles with respect to initial ideas?
		What is the relevance given to the different categories of requirements and roles with respect to initial ideas?
		What is the relevance given to the different categories of requirements and roles with respect to initial ideas?
threats to validity	Threats to Validity	Threats to Validity
threats to validity/threats to validity		
conclusion validity*	✓	✗
construct validity*	✓	✓
external validity*	✓	✓
internal validity*	✓	✓

Now, we **created our ORKG comparison**.

Options before publishing the ORKG comparison:

1. Edit the ORKG comparison by ordering the rows

Edit → Drag & Drop the property cells as required

2. Select properties we want to show.

Actions → Select properties → Disable checkboxes of properties to hide

3. Save ORKG comparison as a draft for later

Actions → Save as draft → Enter title → Save → Draft in “My account”

9. Improve ORKG Comparison: Order Rows

Comparison | 2 contributions

Edit + Add contribution Visualize Actions

Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry Contribution 1 - 2021	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
data.analysis.method	Data Analysis Method	Data Analysis Method
data.analysis.method/data.analysis.method/method*	thematic analysis	descriptive statistics
		inferential statistics
		thematic analysis
data.collection.method	Data Collection Method	Data Collection Method
data.collection.method/data.collection.method		
method*	interview	experiment
number of participants*	10	30
research question	Research Question	Research Question
		Research Question
		Research Question
research question/research question		
hidden in text*	✗	✗
		✗
		✗
highlighted in text*	✓	✓
		✓
		✓
question*	What is the perception and current treatment of NFRs in ML in industry?	How much is the difference in terms of documented requirements and roles with respect to initial ideas?
		What is the relevance given to the different categories of requirements and roles with respect to initial ideas?
		What is the relevance given to the different categories of requirements and roles with respect to initial ideas?
threats to validity	Threats to Validity	Threats to Validity
threats to validity/threats to validity		
conclusion validity*	✓	✗
construct validity*	✓	✓
external validity*	✓	✓
internal validity*	✓	✓

ORKG View Tools About NFDI4DataScience

Search... + Add new

Comparison | 2 contributions

Stop editing + Add contribution Visualize Actions

Edit mode

Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry Contribution 1 - 2021	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
research question		
data analysis	Data Analysis Method	Data Analysis Method
		Data Analysis Method
		Data Analysis Method
data analysis method/data analysis method/method*	thematic analysis	descriptive statistics
		inferential statistics
		thematic analysis

1. Click on “Edit”
2. Click on the grey property cell you want to move
3. Drag & Drop the property cell where you want

9. Improve ORKG Comparison: Order Rows – Result

Comparison 2 contributions	
Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry <i>Contribution 1 - 2021</i>
data analysis.method	Data Analysis Method
data analysis.method/data analysis.method/method*	thematic analysis
data collection.method	Data Collection Method
data collection.method/data collection.method	
↳ method*	interview
↳ number of participants*	10
research question	Research Question
research question/research question	
↳ hidden in text*	✗
↳ highlighted in text*	✓
↳ question*	What is the perception and current treatment of NFRs in ML in industry?
threats to validity	Threats to Validity
threats to validity/threats to validity	
↳ conclusion validity*	✓
↳ construct validity*	✓
↳ external validity*	✓
↳ internal validity*	✓

Comparison 2 contributions	
Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry <i>Contribution 1 - 2021</i>
research question	Research Question
research question/research question	
↳ question*	What is the perception and current treatment of NFRs in ML in industry?
↳ highlighted in text*	✓
↳ hidden in text*	✗
data collection.method	Data Collection Method
data collection.method/data collection.method	
↳ method*	interview
↳ number of participants*	10
data analysis.method	Data Analysis Method
data analysis.method/data analysis.method/method*	thematic analysis
threats to validity	Threats to Validity
threats to validity/threats to validity	
↳ conclusion validity*	✓
↳ construct validity*	✓
↳ external validity*	✓
↳ internal validity*	✓

Data Analysis

Data Collection

Research Question

Threats to Validity

Data Analysis

Data Collection

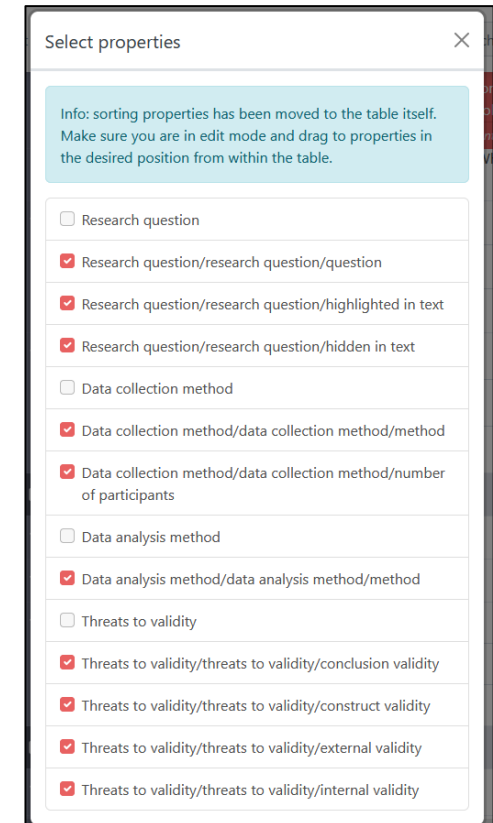
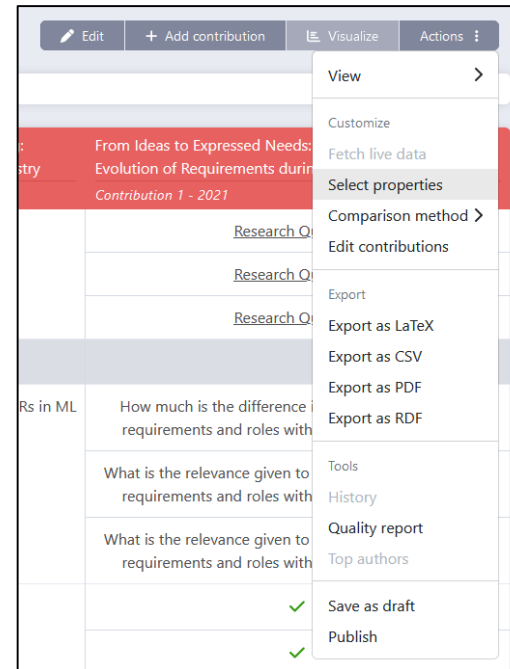
Research Question

Threats to Validity

9. Improve ORKG Comparison: Select Properties

Comparison | 2 contributions

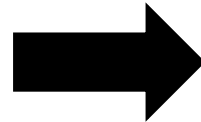
Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry Contribution 1 - 2021	From Ideas to Expressed Needs: An Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
research question	Research Question	Research Question
research question/research question		Research Question
question*	What is the perception and current treatment of NFRs in ML in industry?	How much is the difference in terms of documented requirements and roles with respect to initial Ideas?
highlighted in text*	✓	✓
hidden in text*	✗	✗
data collection method	Data Collection Method	Data Collection Method
data collection method/data collection method		
method*	interview	experiment
number of participants*	10	30
data analysis method	Data Analysis Method	Data Analysis Method
data analysis method/data analysis method/method*	thematic analysis	descriptive statistics
threats to validity	Threats to Validity	Threats to Validity
threats to validity/threats to validity		
conclusion validity*	✓	✗
construct validity*	✓	✓
external validity*	✓	✓
internal validity*	✓	✓



1. Click on “Actions”
2. Click on “Select properties”
3. Disable all checkboxes of properties you want to hide
4. Close dialog

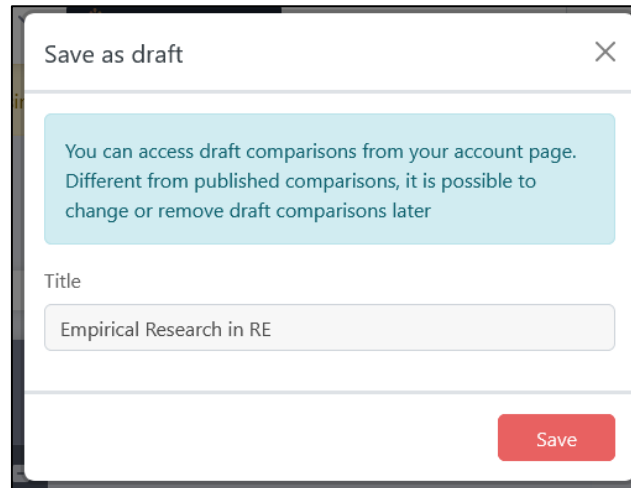
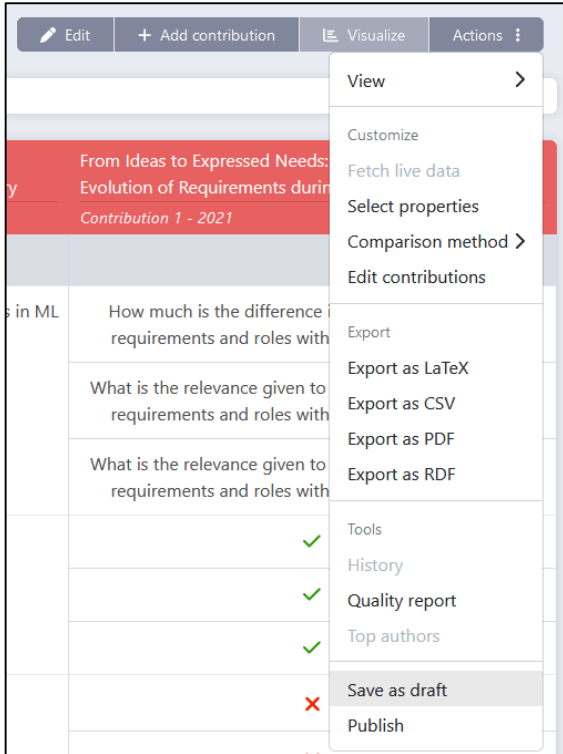
9. Improve ORKG Comparison: Select Properties – Result

Comparison 2 contributions		
Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry Contribution 1 - 2021	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
research question	Research Question	Research Question
research question/research question		Research Question
↳ question*	What is the perception and current treatment of NFRs in ML in industry?	How much is the difference in terms of documented requirements and roles with respect to initial ideas? What is the relevance given to the different categories of requirements and roles with respect to initial ideas? What is the relevance given to the different categories of requirements and roles with respect to initial ideas?
↳ highlighted in text*	✓	✓ ✓ ✓
↳ hidden in text*	✗	✗ ✗ ✗
data collection method	Data Collection Method	Data Collection Method
data collection method/data collection method		
↳ method*	interview	experiment
↳ number of participants*	10	30
data analysis method	Data Analysis Method	Data Analysis Method
data analysis method/data analysis method/method*	thematic analysis	descriptive statistics inferential statistics thematic analysis
threats to validity	Threats to Validity	Threats to Validity
threats to validity/threats to validity		
↳ conclusion validity*	✓	✗
↳ construct validity*	✓	✓
↳ external validity*	✓	✓
↳ internal validity*	✓	✓



Comparison 2 contributions		
Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry Contribution 1 - 2021	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
research question/research question		
↳ question*	What is the perception and current treatment of NFRs in ML in industry?	How much is the difference in terms of documented requirements and roles with respect to initial ideas? What is the relevance given to the different categories of requirements and roles with respect to initial ideas? What is the relevance given to the different categories of requirements and roles with respect to initial ideas?
↳ highlighted in text*	✓	✓ ✓ ✓
↳ hidden in text*	✗	✗ ✗ ✗
data collection method/data collection method		
↳ number of participants*	10	30
↳ method*	interview	experiment
data analysis method/data analysis method/method*	thematic analysis	descriptive statistics inferential statistics thematic analysis
threats to validity/threats to validity		
↳ conclusion validity*	✓	✗
↳ construct validity*	✓	✓
↳ external validity*	✓	✓
↳ internal validity*	✓	✓

9. Improve ORKG Comparison: Save as Draft

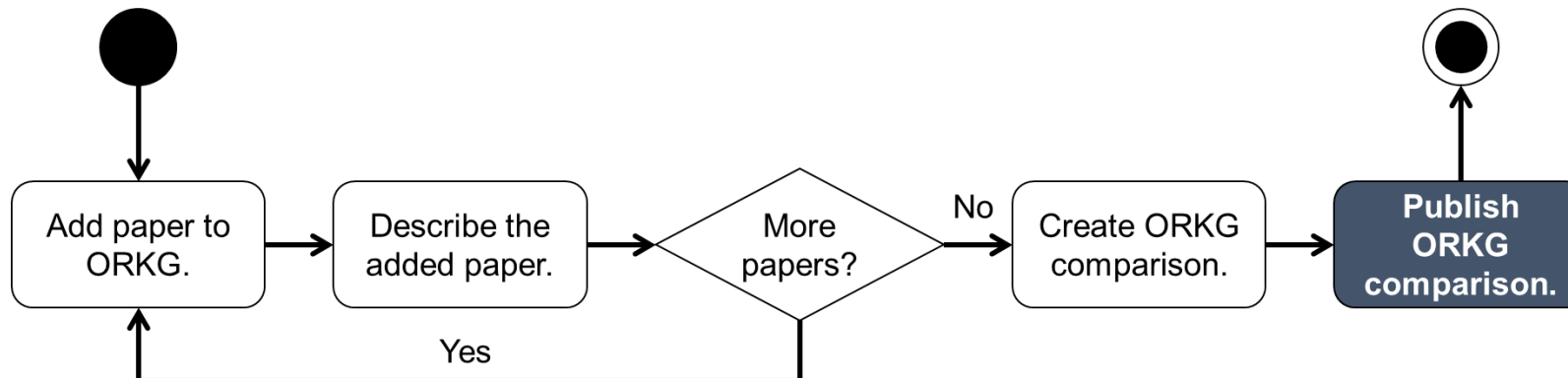
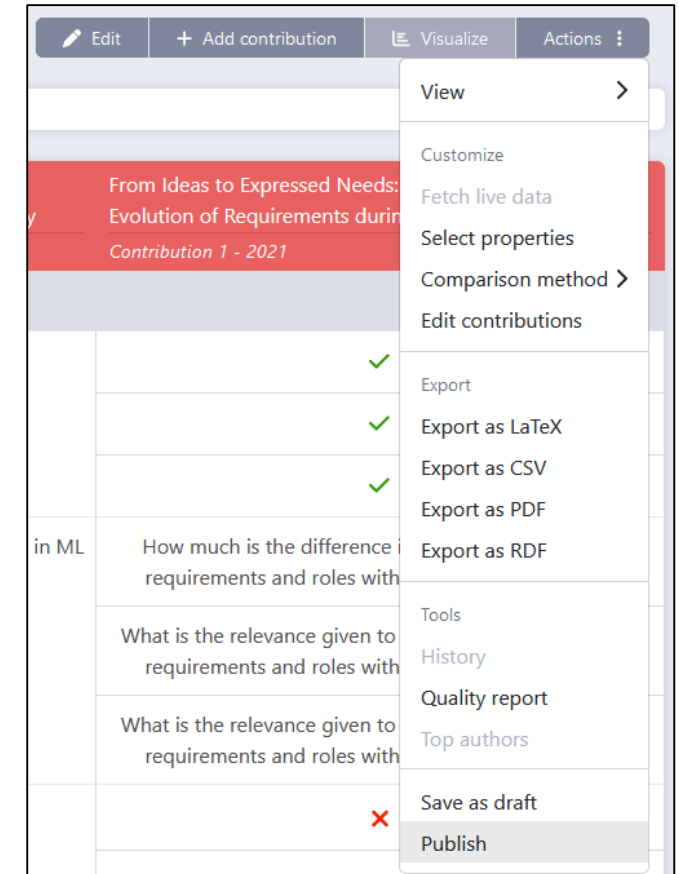


1. Click on “Actions”
2. Click on “Save as draft”
3. Enter a title and click on “Save”
4. The draft is saved in your account

10. Publish ORKG Comparison

Comparison | 2 contributions

Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry <i>Contribution 1 - 2021</i>	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation <i>Contribution 1 - 2021</i>
research question/research question		
↳ highlighted in text*	✓	✓
↳ question*	What is the perception and current treatment of NFRs in ML in industry?	How much is the difference in terms of documented requirements and roles with respect to initial ideas? What is the relevance given to the different categories of requirements and roles with respect to initial Ideas? What is the relevance given to the different categories of requirements and roles with respect to initial Ideas?
↳ hidden in text*	✗	✗



1. Click on “Actions”
2. Click on “Publish”

10. Publish ORKG Comparison

Publish comparison ✕

A published comparison is made public to other users. The state of the comparison is saved and a persistent link is created.

Title ?

Description ?

Research field ?

Creators ?

☐ Assign a DOI to the comparison ?

Reference (optional) ?

Sustainable development goals (optional) ?

Conference (optional) ?

Publish

1. Fill out the dialog
2. Click on “Publish”

Remark:

You can add

- A **DOI** for citation (can also be done later)
- Additional **References**
- Related **Sustainable development goals**
- ORKG partner **Conferences**, such as **REFSQ'25**, that award the **Best ORKG Comparison Award**

Publish comparison ✕

A published comparison is made public to other users. The state of the comparison is saved and a persistent link is created.

Title ?

Description ?

Research field ?

Creators ?

☐ Assign a DOI to the comparison ?

Reference (optional) ?

Sustainable development goals (optional) ?

Conference (optional) ?

a conference
asdadfa
31st International Working Conference on Requirement Engineering: Foundation for Software Quality (REFSQ'25)

10. Publish ORKG Comparison: Result

Comparison | 2 contributions

An Overview of Empirical Research in Requirements Engineering ☆

May 2024 | Oliver Karras

This comparison shows an overview of empirical research reported in publication of the IEEE International Requirements Engineering Conference 2021 regarding the topics research question, data collection, data analysis, and threats to validity.

Properties	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry Contribution 1 - 2021	From Ideas to Expressed Needs: an Empirical Study on the Evolution of Requirements during Elicitation Contribution 1 - 2021
research question/research question		
highlighted in text	✓	✓
question	What is the perception and current treatment of NFRs in ML in industry?	How much is the difference in terms of documented requirements and roles with respect to initial ideas?
hidden in text	✗	✗
data collection method/data collection method		
method	interview	experiment
number of participants	10	30
data analysis method/data analysis method/method	thematic analysis	descriptive statistics inferential statistics thematic analysis
threats to validity/threats to validity		
conclusion validity	✓	✗
construct validity	✓	✓
external validity	✓	✓
internal validity	✓	✓

Added by: Oliver Karras

Assign to observatory

Now, we **published our ORKG comparison**, a stable version that can be maintained, extended, updated, and published as new versions.

Options after publishing the ORKG comparison:

1. Add visualizations
2. Add DOI later and export citation
3. Use the quality report to get feedback from other researchers
4. Fetch live data for a new draft or published version
5. Fetch data for later analysis with different interfaces

Remark:

These options are only demonstrated live in the tutorial. If you need help, do not hesitate to contact the ORKG team, especially **Oliver Karras** (oliver.karras@tib.eu).

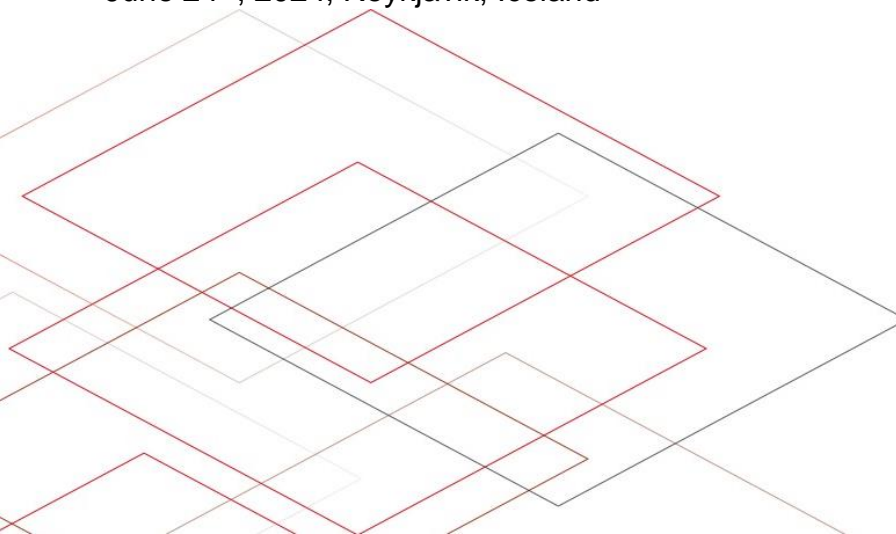
Summary, Reflection, and Closing

Oliver Karras, Alessio Ferrari, Davide Fucci, and Davide Dell'Anna

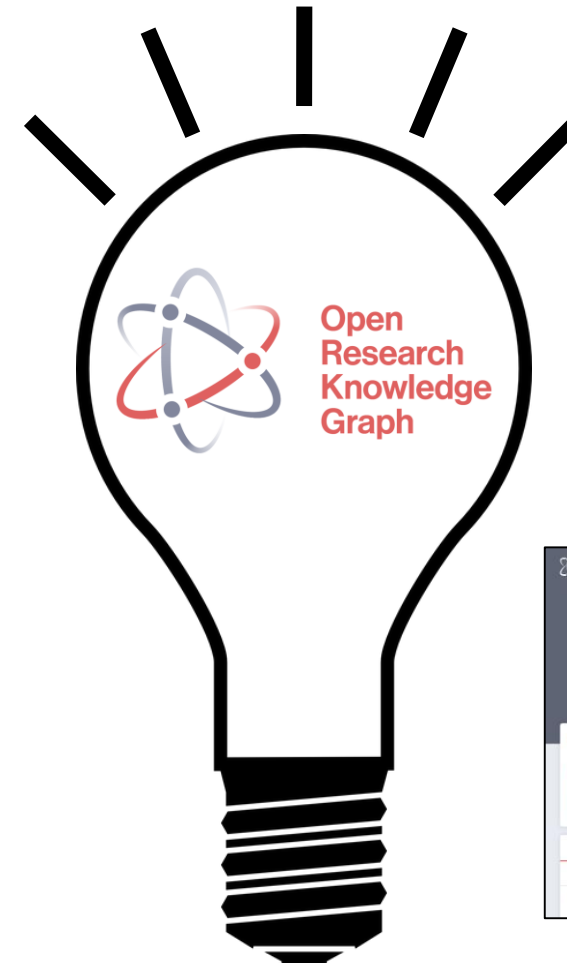
oliver.karras@tib.eu, alessio.ferrari@isti.cnr.it, davide.fucci@bth.se, d.dellanna@uu.nl

32nd IEEE International Requirements Engineering 2024 Conference – Exploring New Horizons: Expanding the Frontiers of Requirements Engineering

June 24th, 2024, Reykjavik, Iceland

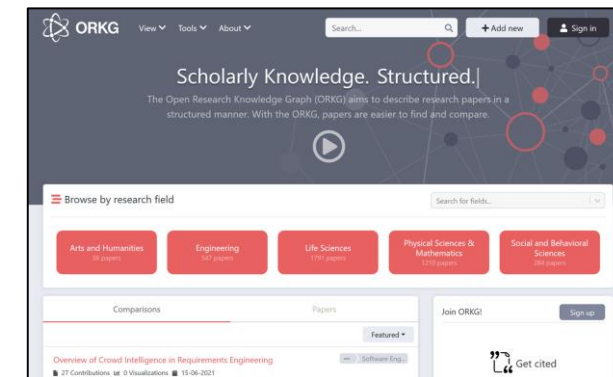


Summary



Open
Research
Knowledge
Graph

<https://orkg.org>



Let's bring scholarly communication and open science in Requirements Engineering to the 21st century!

Recap of the Tutorial

Session	Time	Table of Content	Style	Speaker
Theoretical	09:00 - 09:25	1. Welcome (5 min) 2. Introduction to open science in RE (10 min) 3. Introduction to the ORKG (10 min)	Presentation Presentation Presentation	All organizers Alessio Ferrari Oliver Karras
Practical	09:25 - 10:15	4. Create a FAIR-annotated publication for the ORKG (50 min) 4.1 Set up an Overleaf project for an exemplary publication 4.2 Use the LaTeX package SciKGT _E X to annotate the publication 4.3 Generate PDF with embedded FAIR scientific information 4.4 Optional: Upload the FAIR-annotated publication to the ORKG	Exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise	Oliver Karras All organizers All organizers All organizers All organizers
Break	10:15 - 10:45	Coffee break		
Practical	10:45 - 11:45	5. Use the ORKG based on a RE use case (60 min) 5.1 Add an exemplary publication to the ORKG 5.2 Describe the scientific information of the publication in the ORKG 5.3 Create an ORKG comparison of the publications added by participants 5.4 Publish the created ORKG comparison as a citable digital artifact 5.5 Optional: Create visualizations for the created ORKG comparison 5.6 Optional: Retrieve the information with the SPARQL endpoint	Exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise Sub-exercise	Oliver Karras All organizers All organizers All organizers All organizers All organizers All organizers
Feedback	11:45 - 12:15	6. Reflection of the tutorial with the participants (25 min) 7. Farewell and closing (5 min)	Discussion Presentation	All organizers All organizers

Reflection and Feedback

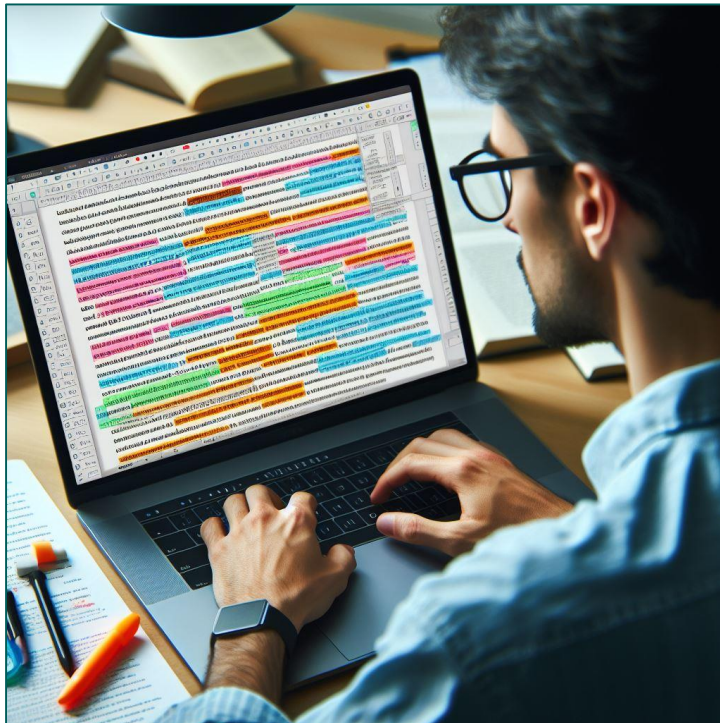
Your thoughts, please!

Teaser: REFSQ'25 Open Science Competition

Challenge 1:

Annotate your REFSQ'25 paper with SciKGTeX.

The accepted paper, best annotated with SciKGTeX, will be awarded the **Best ORKG Annotation Award** (prize: 100€).



SciKGTeX

Challenge 2:

Enrich your RESFQ'25 paper with an ORKG comparison.

The accepted paper, enriched with the best ORKG comparison, will be awarded the **Best ORKG Comparison Award** (prize: 200€).



Open
Research
Knowledge
Graph

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@t3s.de

Abstract—(Background.) Empirical research in requirements engineering (RE) is a constantly evolving topic, with a growing

(systematic) literature reviews, and even examine overlapping periods, venues, and themes (cf. Table 1) [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], [29] 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 *Dispedia* [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,

A Comparison of Scientific Publications on the State of Empirical Research in Requirements Engineering and Software Engineering ★

November 2023 Oliver Karras Felix Wernlein Jil Ann-Christin Klünder Sören Auer

This comparison provides an overview of scientific publications that have investigated primary studies in requirements engineering and software engineering to give a snapshot of the “current” state of empirical research in requirements engineering and software engineering. In particular, the comparison shows for each publication (1) which research fields and topics were investigated, (2) whether and where the extracted and analyzed data is available, and (3) which method was used to determine the state, including further details about the respective method.

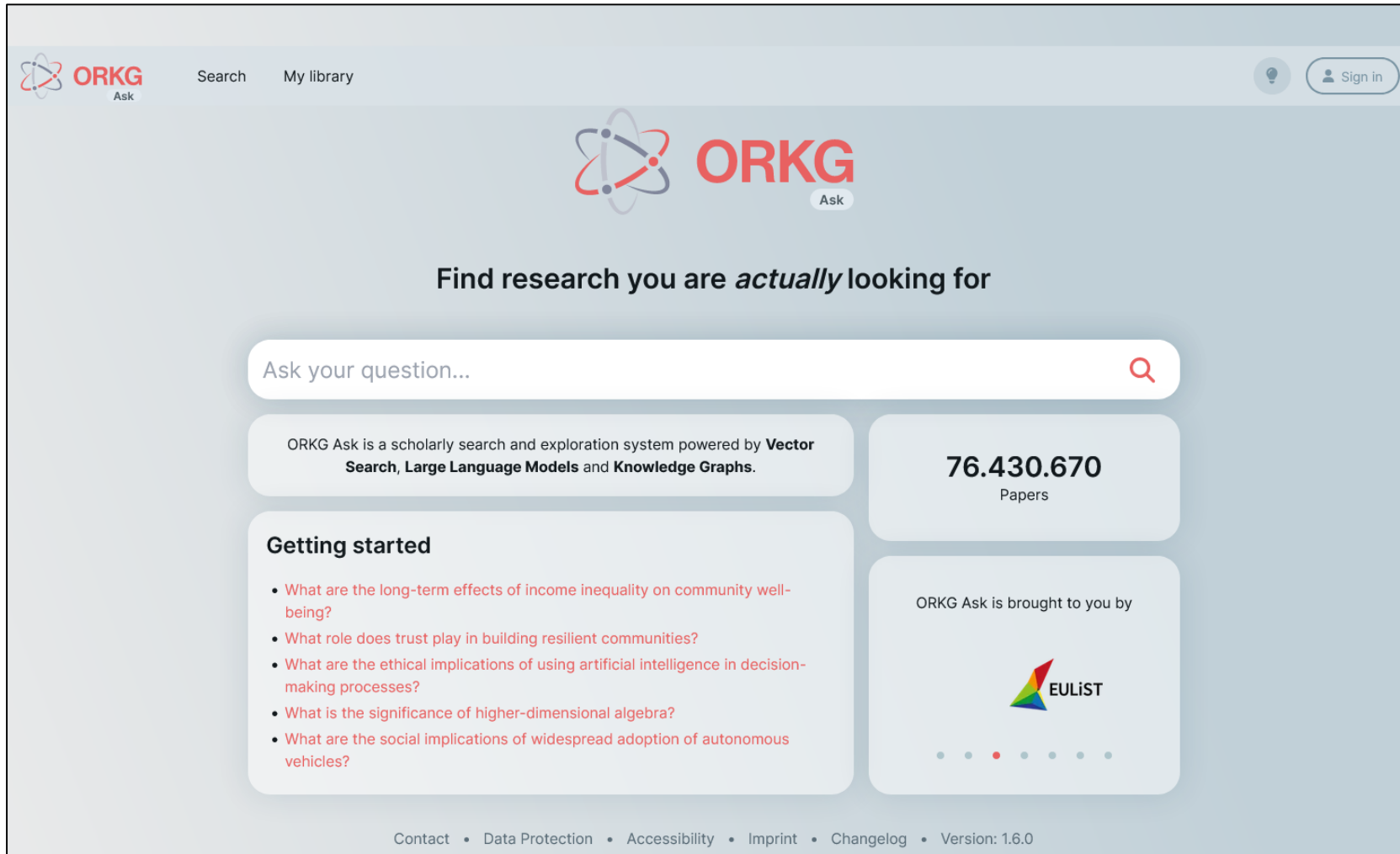
DOI: <https://doi.org/10.48366/RES0023>



Properties	Empirical research in requirements engineering: trends and opportunities <i>Empirical research - 2016</i>	Empirical research methodologies and studies in Requirements Engineering: How far did we come? <i>Empirical research - 2014</i>	A Survey on Empirical Requirements Engineering Research Practices <i>Empirical research - 2012</i>	Evidence-Based Structuring and Evaluation of Empirical Research in Requirements Engineering: Fundamentals, Framework, Research Map <i>Empirical research - 2010</i>	An Anal. Requires Data <i>Empirical research - 2010</i>
research problem	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering	empirical research in requirements engineering
research field investigated	Requirements Engineering	Requirements Engineering	Requirements Engineering	Requirements Engineering	Requirements Engineering
topic investigated	bibliographic metadata context data collection	bibliographic metadata research topic theory	context data analysis data collection	context research method result	context research method result

your personal use. Not for redistribution. The definitive version or record was published in the proceedings of 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), <https://doi.org/10.1109/ESEM56168.2023.10304795>.

Looking to the Future: ORKG Ask



The ORKG Ask is a scientific **search and exploration system** that helps researchers **find** the **research articles** they are really looking for.

Ask a **question**.

Get an **answer** with **references**.

<https://ask.orkg.org/>
(Released: May16th, 2024)

Looking to the Future: ORKG Ask

The screenshot displays the ORKG Ask web application. At the top, there's a navigation bar with the ORKG logo, 'Search', 'My library', and a 'Sign in' button. Below this is a toolbar with 'Save', 'Share', 'Hidden', 'CSV', 'ORKG', and 'Edit columns' buttons. A search query box on the left contains the text 'How does exposure to nature effect overall health and well-being?' with a 'Search' button. Below the search box are filter sections for 'Year', 'Language', and 'Add filter...'. A sidebar on the right has a 'Feedback' button. The main content area shows a warning message about the use of a large language model, followed by an 'Answer (based on top 5 papers)' section. Below this is a table with four columns: 'Answer', 'Insights', 'TL;DR', and 'Conclusions'. The table contains two rows of results, each with a title, authors, date, and DOI link.

Search query
How does exposure to nature effect overall health and well-being?

Filters
Year
Language
Add filter...

Answer (based on top 5 papers)
Exposure to nature, whether in actual settings or through virtual means, has been shown to increase positive emotions, attentional capacity, and ability to reflect on life problems [1], while contact with natural environments also improves air quality and reduces stress [2]. Moreover, a psychological characteristic called nature relatedness, which reflects the degree of connection one feels towards nature, is linked to better overall health and reduced symptoms of depression, anxiety, and stress [3,4]. [5] adds that cultural connections to nature contribute significantly to human well-being, affecting mental, physical, and spiritual health, as well as inspiration and identity.

Answer	Insights	TL;DR	Conclusions
Why is nature beneficial? The role of connectedness to nature Mayer, F. Stephan • Frantz, Cynthia M. • Bruehlman-Senecal, E. • Dolliver, Kyffin September 2009 • 10.1177/0013916508319745	Increased connectedness to nature, Positive emotions, Ability to reflect on a life problem	The study discusses the mechanisms underlying the exposure to nature/well-being effects and the ecological validity of laboratory studies.	Exposure to nature improves connectedness to nature, enhances cognitive abilities, and boosts positive emotions.
Health benefits of nature Berg, van den, A.E. • Joye, Y. • Vries, de, S. January 2018 • 10.1002/9781119241072.ch6	Improvement in air quality, stimulation of physical activity, facilitation of social cohesion, stress reduction	The scientific research and insights discussed in the chapter	Exposure to nature leads to improvements in air quality, stimulation of physical activity, facilitation of social cohesion, and stress reduction.

1. Enter a natural language **question** in the **UI**
2. **Semantic search** finds the most relevant **papers**
3. **LLMs** extract the required information and create the **answer**
4. **KGs** are used to **support** content extraction, synthesis, and enhancement