

GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER  
FAKULTÄT FÜR ELEKTROTECHNIK UND INFORMATIK

# SciKGDash - Ein modulares Dashboard zur Kuratierung von Forschungswissensgraphen

*Wissenschaftliche Arbeit zur Erlangung des  
Master of Science in Informatik*

von

**Lena Daniela John**

Matrikelnummer: 3217310

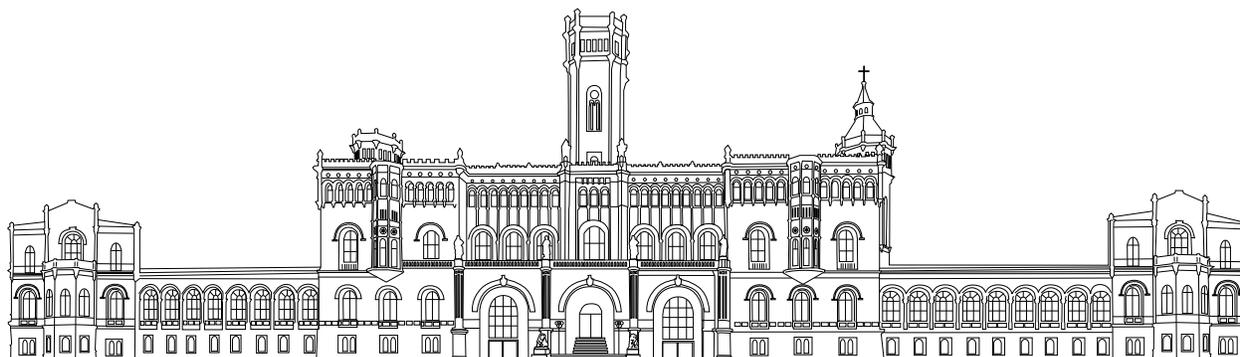
E-Mail: lena.john@stud.uni-hannover.de

Erstprüfer: Prof. Dr. Sören Auer

Zweitprüfer: Dr. rer. nat. Oliver Karras

Betreuer: Dr. rer. nat. Oliver Karras

Mai 2024





# Erklärung der Selbstständigkeit

Ich, Lena Daniela John, erkläre, dass die vorliegende Masterarbeit mit dem Titel „SciKGDash - Ein modulares Dashboard zur Kuratierung von Forschungswissensgraphen“ und die darin enthaltenen Arbeiten meine eigenen sind. Ich bestätige, dass:

- Diese Arbeit wurde im Rahmen einer Bewerbung für einen Forschungsabschluss an dieser Universität durchgeführt.
- Wenn ein Teil dieser Arbeit bereits für einen Abschluss oder eine andere Qualifikation an dieser Universität oder einer anderen Einrichtung eingereicht wurde, ist dies deutlich angegeben.
- An Stellen, bei denen ich veröffentlichte Arbeiten anderer konsultiert habe, ist dies immer deutlich angegeben.
- Wenn ich aus der Arbeit anderer zitiert habe, ist die Quelle immer angegeben. Mit Ausnahme solcher Zitate ist die vorliegende Arbeit vollständig mein eigenes Werk.
- Ich habe alle wichtigen Hilfsquellen angegeben.
- Ich habe ChatGPT zur Unterstützung bei der Inhaltserstellung und Implementierung verwendet, jedoch alle Inhalte auf ihre Korrektheit überprüft und stehe mit meinem Namen für den Inhalt ein.

Lena Daniela John

Unterschrift:  \_\_\_\_\_

Datum: 21.05.2024



*Für meinen Sohn, den besten Forscher, den man sich wünschen kann.*



## *Danksagung*

Ich möchte dem gesamten Team des Open Research Knowledge Graph (ORKG) meinen aufrichtigen Dank aussprechen für ihre unschätzbare Unterstützung während meiner Forschungsarbeit. Eure Bereitschaft, meine Fragen zu beantworten und mir in verschiedensten Belangen weiterzuhelfen, war entscheidend für den Erfolg dieser Masterarbeit.

Besonders danke ich meinem Betreuer Oliver Karras, der keinen Sonntag kennt und mich stets mit großem Engagement und Unterstützung begleitet hat.



# Zusammenfassung

**[Hintergrund]** Forschungswissensgraphen, als strukturierte Repräsentationen wissenschaftlicher Domänen, stehen vor der Herausforderung, ihre Qualität angesichts ihrer wachsenden Größe zu erhalten. Unzureichende Wartung kann zu minderwertigen Anwendungen führen, weshalb die Kuratierung eine zentrale Rolle spielt. Im spezifischen Kontext des Open Research Knowledge Graph (ORKG) hat sich ein Team für Kuratierung und Gemeinschaftsbildung formiert. Der ORKG wird durch Crowdsourcing erweitert, wodurch die Qualität des Graphen stark von den Beiträgen der Nutzer:innen abhängt. Ein zentrales Problem besteht darin, dass die manuelle Kuratierung sehr zeit- und kostenintensiv ist, jedoch notwendig bleibt, da der semantische domänenspezifische Kontext am besten von Menschen erfasst werden kann. Mit dem fortschreitenden Wachstum des ORKG wird jedoch die Implementierung maschineller Unterstützung erwogen. Diese soll die menschliche Kuratierung erleichtern, ohne selbst direkte Änderungen am Wissensgraph vorzunehmen.

**[Ziel]** Zur Unterstützung des Teams wurde ein Dashboard entwickelt, das gezielt auf deren Bedürfnisse zugeschnitten ist. In diesem Kontext werden ausgewählte User Stories in Form eines Minimum Viable Product (MVP) umgesetzt. Diese User Stories umfassen Metriken, die gemäß der Expertise des Teams Aufschluss über die Qualität des ORKGs geben. Der MVP soll bereits zum Abschluss dieser Arbeit einen signifikanten Mehrwert für das Team darstellen. Darüber hinaus soll das Dashboard Modularisierungsoptionen bieten, um die Qualität verschiedener Wissensgraphen zu überwachen und zu vergleichen.

**[Methode]** Der Ansatz beruht auf einer engen Zusammenarbeit im Sinne der Aktionsforschung, wodurch das Team kontinuierlich Einfluss auf die Gestaltung nehmen kann. Dabei werden zwei Zyklen angesetzt, in denen der Prototyp iterativ entsprechend den Anforderungen des Teams weiterentwickelt wird. Ein Zyklus umfasst die Planung und Umsetzung der Erkenntnisse aus dem vorherigen Zyklus, eine Evaluierung sowie die Reflexion über gewonnene Einsichten. Die Evaluierung erfolgt in Form eines Pluralistic Walkthroughs, wobei das Feedback der zukünftigen Nutzer:innen von zentraler Bedeutung ist. Abschließend wird das Dashboard im Rahmen eines

Experiments mit 15 Teilnehmenden, bestehend aus zukünftigen Nutzer:innen und Expert:innen im Kontext des ORKG, evaluiert. Hierbei werden sowohl objektive als auch subjektive Metriken zur Bedienbarkeit des Dashboards und zur Umsetzung der Bedürfnisse erhoben.

**[Ergebnisse]** Die Ergebnisse zeigen, dass die Bedienbarkeit, im Form von Effizienz, Effektivität und Zufriedenheit, als positiv zu betrachten ist. Die Teilnehmenden waren in der Lage, vier von fünf Aufgaben in der vorgegebenen Zeit zu erledigen, mehr als die Hälfte der Aufgaben erfolgreich abzuschließen und gaben eine positive Nutzererfahrung an. Die Bedürfnisse wurden laut den Teilnehmenden in großen Teilen erfolgreich umgesetzt.

**[Schlussfolgerung]** Dadurch entstand eine maßgeschneiderte Anwendung, die die Arbeit des Teams effektiv unterstützt. Diese spezialisierte Lösung stellt einen Kompromiss zwischen einem generischen und einem spezifischen Ansatz dar. Bei der Modularität offenbarte sich jedoch eine Grenze, da jeder Wissensgraph einzigartig ist und Qualität eine multidimensionale Metrik darstellt, sodass die Graphen nicht pauschal verglichen werden können.

# Abstract

[**Context**] Research knowledge graphs, as structured representations of scientific domains, face the challenge of maintaining their quality as they grow. Insufficient maintenance can lead to subpar applications, making curation a critical task. In the specific context of the Open Research Knowledge Graph (ORKG), a team dedicated to curation and community building has formed. The ORKG is expanded through crowdsourcing, making the quality of the graph heavily dependent on user contributions. A central problem is that manual curation is very time- and cost-intensive, but remains necessary because the semantic, domain-specific context is best understood by humans. However, with the continuous growth of the ORKG, the implementation of machine support is being considered. This support aims to ease human curation without making direct changes to the knowledge graph itself.

[**Objective**] To support the team, a dashboard was developed specifically tailored to their needs. In this context, selected user stories are implemented as a Minimum Viable Product (MVP). These user stories encompass metrics that, according to the team's expertise, provide insights into the quality of the ORKG. The MVP is intended to provide significant value to the team by the conclusion of this work. Additionally, the dashboard is designed to offer modularization options to monitor and compare the quality of various knowledge graphs.

[**Method**] The approach is based on close collaboration in the spirit of action research, allowing the team to continuously influence the design. Two cycles are implemented, in which the prototype is iteratively refined according to the team's requirements. Each cycle includes planning and implementing the lessons learned from the previous cycle, evaluation, and reflection on the insights gained. Evaluation is conducted in the form of a pluralistic walkthrough, where feedback from future users is crucial. Finally, the dashboard is evaluated through an experiment involving 15 participants, consisting of future users and experts in the context of the ORKG. Both objective and subjective metrics are collected to assess the usability of the dashboard and the extent to which the needs have been met.

[**Results**] The results indicate that the usability, in terms of efficiency, effectiveness,

ess, and satisfaction, is considered positive. Participants were able to complete four out of five tasks within the allotted time, successfully complete more than half of the tasks, and reported a positive user experience. According to the participants, the needs were largely successfully implemented.

**[Conclusion]** This resulted in a tailored application that effectively supports the team's work. This specialized solution represents a trade-off between a generic and a specific approach. However, a limitation was revealed in terms of modularity, as each knowledge graph is unique and quality is a multidimensional metric, making it impossible to compare the graphs uniformly.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
1.1	Motivation . . . . .	2
1.2	Zielsetzung . . . . .	4
1.3	Struktur . . . . .	5
1.4	Mehrwert . . . . .	6
<b>2</b>	<b>Grundlagen</b>	<b>8</b>
2.1	(Forschungs-)wissensgraphen . . . . .	8
2.2	Kuratierung von Wissensgraphen . . . . .	10
2.3	ORKG Kuratierung und Gemeinschaftsbildung . . . . .	11
2.4	Dashboardspezifische Methoden . . . . .	13
2.5	Aktionsforschung . . . . .	15
<b>3</b>	<b>Verwandte Arbeiten</b>	<b>18</b>
<b>4</b>	<b>Methodik</b>	<b>22</b>
<b>5</b>	<b>Technische Basis</b>	<b>28</b>
5.1	Technischer Aufbau des ORKG . . . . .	28
5.2	Technische Umsetzung des Dashboards . . . . .	31
5.3	Ausführung . . . . .	34
<b>6</b>	<b>Durchführung der Aktionsforschung</b>	<b>36</b>
6.1	User Stories . . . . .	37
6.2	Erster Zyklus . . . . .	42
6.2.1	Diagnose . . . . .	43
6.2.2	Planung . . . . .	43
6.2.3	Maßnahmengreifung . . . . .	44
6.2.4	Evaluierung . . . . .	45

---

6.2.5	Lernen . . . . .	46
6.3	Zweiter Zyklus . . . . .	47
6.3.1	Diagnose . . . . .	48
6.3.2	Planung . . . . .	48
6.3.3	Maßnahmengreifung . . . . .	48
6.3.4	Evaluierung . . . . .	52
6.3.5	Lernen . . . . .	53
6.4	MVP . . . . .	53
<b>7</b>	<b>Experiment</b>	<b>56</b>
7.1	Hypothesen . . . . .	57
7.2	Abhängige und unabhängige Variablen . . . . .	59
7.3	Material . . . . .	60
7.4	Stichprobe . . . . .	63
7.5	Ablauf des Experiments . . . . .	65
7.6	Datenanalyse . . . . .	66
7.6.1	Korrelation . . . . .	67
7.6.2	Effizienz . . . . .	68
7.6.3	Effektivität . . . . .	73
7.6.4	Zufriedenheit . . . . .	76
7.6.5	Bedarfe des Kuratierungs- und Gemeinschaftsbildungsteams . . . . .	80
7.7	Einschränkungen der Validität . . . . .	81
7.7.1	Konstruktvalidität . . . . .	81
7.7.2	Externe Validität . . . . .	83
7.7.3	Interne Validität . . . . .	83
7.7.4	Schlussfolgerungvalidität . . . . .	85
<b>8</b>	<b>Diskussion</b>	<b>86</b>
<b>9</b>	<b>Zusammenfassung</b>	<b>90</b>
9.1	Fazit . . . . .	90
9.2	Ausblick . . . . .	91
<b>A</b>	<b>Screenshots des MVP - SciKG Dashboard</b>	<b>93</b>
<b>B</b>	<b>Begleitdokument Experiment</b>	<b>101</b>
	<b>Literatur</b>	<b>106</b>

# Kapitel 1

## Einleitung

### 1.1 Motivation

In einer Ära des exponentiellen Wissenswachstums durch Forschung steht die Wissenschaftsgemeinschaft vor großen Herausforderungen. Die hohe Anzahl der jährlich veröffentlichten Forschungsbeiträge (im Jahr 2014: über 7 Millionen [18]) macht es für Forschende unmöglich den aktuellen Stand der Forschung in Gänze zu erfassen [52]. Die traditionelle Aufzeichnung von wissenschaftlichen Erkenntnissen erfolgte in gedruckten Formaten [30]. Trotz der Digitalisierung bleiben Artikel zumeist nur digitale Repräsentationen ihrer gedruckten Gegenstücke, was die wissenschaftliche Kommunikation durch ihre dokumentbasierte Form weiterhin behindert [30]. So sind Forschungsberichte und andere Publikationen zumeist in menschenlesbarer und in nicht-maschinenlesbarer Form vorhanden, sodass die Forschungsbeiträge nicht effektiv durchsucht, gefiltert oder visualisiert werden können [4]. Diese Limitierung zeigt, dass die digitale Transformation im wissenschaftlichen Kontext bisher nur unzureichend umgesetzt wurde.

Die fehlende Strukturierung von Daten für wissensbasierte System behindert die effektive Nutzung dieses reichhaltigen Wissensschatzes erheblich. Um die Inhalte somit ebenfalls für Maschinen zugänglich zu machen, benötigt es eine kontextsensitive, semantische Interpretation des Geschriebenen [29]. Dies bildet die Grundlage des Semantic Webs, in dem Daten nicht nur zur Anzeige, sondern auch zur Automatisierung, Integration und Wiederverwendung über verschiedene Anwendungen hinweg miteinander verknüpft sind [55].

In diesem Zusammenhang spielen Wissensgraphen eine zentrale Rolle. Wissensgraphen ermöglichen die strukturierte Repräsentation von Informationen und deren semantische Verknüpfung, wodurch sie die Interoperabilität und Nutzbarkeit von

Daten erheblich verbessern [29]. Sie dienen als Grundlage für die effiziente Organisation und Nutzung von Wissen im Semantic Web, indem sie Beziehungen und Kontexte zwischen verschiedenen Datenpunkten explizit darstellen und somit eine umfassendere und tiefere Nutzung des vorhandenen Wissens ermöglichen [38].

Ein herausragendes Beispiel für die Anwendung dieser Prinzipien im wissenschaftlichen Bereich ist der Open Research Knowledge Graph (ORKG)<sup>1</sup>. Der ORKG nutzt die Vorteile von Wissensgraphen, um wissenschaftliches Wissen strukturiert und semantisch verknüpft darzustellen. Dies unterstützt die wissenschaftliche Gemeinschaft dabei, effizienter zusammenzuarbeiten und neues Wissen zu generieren. Ein wesentlicher Aspekt für den Erfolg des ORKG ist die sorgfältige Kuratierung der Inhalte.

„Das Ziel der Wissenskuratierung ist es, kostensensitive Methoden bereitzustellen, um die Qualität großer Wissensgraphen zu verbessern und ihre Nützlichkeit für vorgesehene Anwendungen sicherzustellen.“ [16]

Die Kuratierung von Wissen ist eine Herausforderung, die eine Balance zwischen Genauigkeit und Effizienz erfordert. Aktuell erfolgt die meiste Kuratierung noch durch menschliche Hand oder halbautomatisierte Verfahren [35]. Im Fall des ORKG kümmert sich eine kleine Gruppe sowohl um die Kuratierung als auch um die Gemeinschaftsbildung. Sie leiten die Datengewinnung durch Crowdsourcing und überwachen die Qualität der eingehenden Daten. Dabei müssen sie sich mit der Herausforderung von Duplikaten und fehlerhaften Aussagen auseinandersetzen, die im Rahmen der Erstellung von Wissensgraphen und Crowdsourcing auftreten können. Dabei unterstützen sie die Nutzenden bei der korrekten Dateneingabe und identifizieren potenzielle Verbesserungen, um die Qualität der Daten zu steigern.

Hier setzt die vorliegende Arbeit an. Es wird ein Dashboard entwickelt, welches das Kuratierungs- und Gemeinschaftsbildungsteam in ihren Arbeitsprozessen unterstützen soll. Dabei wird eine maßgeschneiderte Anwendung erstellt, die eng mit dem Team zusammenarbeitet und speziell auf ihre Bedürfnisse zugeschnitten ist. Bei der Bewertung der Qualität von Wissensgraphen wird stets ein *fitness for use*-Ansatz verfolgt [24]. Langfristig birgt das Dashboard das Potenzial, neue Workflows zu integrieren und die Möglichkeit zur Erforschung automatisierter Kuratierung zu bieten. Dadurch könnte der ORKG zukünftig optimiert und weiterentwickelt werden.

---

<sup>1</sup><https://orkg.org/>

## 1.2 Zielsetzung

In der Welt der Wissenskuration und Gemeinschaftsbildung, insbesondere im Kontext des Open Research Knowledge Graphs (ORKG), ist der Zugang zu umfassenden Metriken und Statistiken von grundlegender Bedeutung. Vor diesem Hintergrund hat das Team vorab priorisierte Issues identifiziert, die für ihre Arbeit von großer Relevanz sind. Diese Issues umfassen Informationen, die bisher nicht auf einen Blick verfügbar sind. Um an diese Informationen zu gelangen, müssten sie entweder manuell Listen durchsuchen oder das Entwicklerteam nach spezifischen Daten fragen, da das Frontend nicht für die Kuration, sondern primär für die Benutzerfreundlichkeit beim Hinzufügen von Inhalten zum ORKG ausgelegt ist. Weiterhin bietet der externe Dienst Matomo<sup>2</sup> Besucherstatistiken an. Die Integration dieser Statistiken würde es ermöglichen, alle relevanten Informationen auf einen Blick zu haben. Diese Daten liefern wertvolle Einblicke in das Navigationsverhalten der Nutzer:innen auf der Plattform, was wiederum Erkenntnisse für mögliche Verbesserungen ermöglicht um die Nutzer:innen besser bei ihren Aufgaben zu unterstützen.

Die Mitarbeitenden des ORKG legen großen Wert auf einen regen Austausch zwischen allen Teams, was mir die Möglichkeit bietet, schnell sowohl mit dem Kuratierungs- und Gemeinschaftsbildungsteam als auch mit dem Entwicklerteam des ORKG in Kontakt zu treten, um die benötigten Daten zu erhalten und eventuelle Probleme zügig zu lösen.

Das zu entwickelnde Dashboard wird Metriken in visualisierter Form bereitstellen, wobei die priorisierten Bedürfnisse des Teams in Abstimmung mit den technischen Möglichkeiten umgesetzt werden, um die Daten vom Backend zu erhalten. Zusätzlich werden die Matomo-Besucherstatistiken integriert. Dabei wird ein modularer Ansatz verfolgt, der die Vergleichbarkeit von Metriken mit anderen Wissensgraphen ermöglicht und die Erweiterung für andere Wissensgraphen anbietet. Die Herausforderung besteht darin, eine maßgeschneiderte und flexible Lösung zu schaffen, die nicht nur anpassbar ist, sondern auch für andere Wissensgraphen wiederverwendbar sein soll.

Dabei wird der flexible Ansatz der Aktionsforschung (vgl. Kapitel 2.5) angewendet, um die Realisierbarkeit und Prioritäten der identifizierten Bedürfnisse in Abstimmung mit dem Team praxisnah umzusetzen. Im Rahmen der Arbeit entsteht ein Minimal Viable Product (MVP) (vgl. Definition 6.1), welches die wichtigsten Bedarfe des Teams adressiert.

Um die Ausrichtung dieser Masterarbeit zu entwickeln, wird unter Verwendung der Zielvorlage von Basili et. al [57] das Forschungsziel konkretisiert:

---

<sup>2</sup><https://matomo.org/>

#### Forschungsziel

**Analysiere** die Bedarfe des Kuratierungs- und Gemeinschaftsbildungsteams **mit dem Zweck** der Unterstützung des Prozesses der Kuratierung und Gemeinschaftsbildung von Forschungswissensgraphen  
**im Hinblick auf** Umsetzung eines Dashboards in Form eines MVPs  
**aus der Perspektive des ORKG-Teams**, insbesondere des Teams „Kuratierung und Gemeinschaftsbildung“  
**im Kontext vom** Open Research Knowledge Graph und in Anlehnung an andere Forschungswissensgraphen.

Es ist wichtig zu betonen, dass das Forschungsziel während des Verlaufs der Arbeit regelmäßig überprüft wird, um sicherzustellen, dass es weiterhin die aktuellen Ziele widerspiegelt. Am Ende wird das entwickelte Dashboard in einem abschließenden Experiment evaluiert, um seine Wirksamkeit bei der Unterstützung des Kuratierungs- und Gemeinschaftsbildungsprozesses zu überprüfen.

## 1.3 Struktur

Die vorliegende Arbeit beginnt mit einem Überblick (vgl. Kapitel 2) über grundlegende Konzepte, um die Rahmenbedingungen für die Umsetzung des Dashboards zu schaffen. Dabei wird erklärt, was Forschungswissensgraphen sind und wie ihre Kuratierung aussieht, einschließlich der Herausforderungen, mit denen sie konfrontiert ist. Im Grundlagentext wird ebenfalls das Kuratierungs- und Gemeinschaftsbildungsteam des ORKG beleuchtet, um zu verstehen, mit welchem Partner im praktischen Kontext zusammengearbeitet wird. Anschließend werden dashboardspezifische Methoden erläutert und die Aktionsforschung und ihr zugrunde liegender Ansatz vorgestellt.

Im nächsten Abschnitt werden verwandte Arbeiten (vgl. Kapitel 3) referenziert und die vorliegende Arbeit von diesen abgegrenzt. Das Methodik-Kapitel (vgl. Kapitel 4) beschreibt die Umsetzung der Aktionsforschung, die im Design-Science-Ansatz eingebettet ist.

Die technische Basis (vgl. Kapitel 5) gibt einen Überblick über den Aufbau des ORKG und wie die Dienste davon genutzt werden können. Weiterhin wird die technische Umsetzung des Dashboards und die Ausführung einer lokalen Instanz erklärt.

Im Hauptteil der Arbeit wird die Durchführung der Aktionsforschung (vgl. Kapitel 6) behandelt. Hier werden die Zyklen instanziiert und im Detail erklärt, wann implementiert wurde und welche Probleme auftraten.

Dann widmet sich das folgende Kapitel dem Experiment (vgl. Kapitel 7) mit zukünftigen Nutzer:innen und der Analyse desselben, um die Bedienbarkeit und Umsetzung der Bedarfe des Dashboards zu untersuchen.

Das Diskussionskapitel (vgl. Kapitel 8) setzt die Datenanalyse und Ergebnisse des Experiments in den Kontext des Forschungsziels für das Experiment und beantwortet die Frage, inwiefern die Umsetzung erfolgreich war.

Abschließend wird die Arbeit zusammengefasst (vgl. Kapitel 9), ihre Grenzen aufgezeigt und auf zukünftige Forschungsbereiche hingewiesen.

## 1.4 Mehrwert

Das entwickelte Dashboard bietet einen Mehrwert, indem es das Kuratierungs- und Gemeinschaftsbildungsteam unterstützt. Durch den partizipativen Ansatz der Aktionsforschung wird Praxisnähe gewährleistet, indem das Team aktiv in den Entwicklungsprozess einbezogen wird. Diese Interaktion ermöglicht es, die spezifischen Bedürfnisse des Teams zu identifizieren und das Dashboard entsprechend anzupassen, um ihre Arbeit effektiver zu gestalten. Die Existenz des Dashboards allein bringt dem Team bereits einen deutlichen Mehrwert.

Die Arbeit ermöglicht es auch, notwendige Anpassungen im ORKG zu identifizieren, um nicht nur die Bedürfnisse der Nutzenden, sondern auch die der Kurator:innen zu berücksichtigen. Dies legt eine Grundlage für zukünftige Kommunikation innerhalb der ORKG-Teams und ermöglicht eine gezieltere Gestaltung für die Kurator:innen, da die Entwickler:innen nun besser informiert sind, was für sie relevant ist.

Durch die Entwicklung, insbesondere der Besucherpfade, sind neue Ideen aufgetaucht, die für die Kuratierung und Gemeinschaftsbildung interessant sein könnten. Zum Beispiel wäre es interessant, Profile von Nutzer:innen zu identifizieren, insbesondere von Neueinsteiger:innen oder langjährigen Nutzer:innen, um ihr Verhalten auf der ORKG-Seite besser zu verstehen.

Im größeren Kontext trägt die Arbeit zum Verständnis der Kuratierung von Wissensgraphen bei, insbesondere des ORKGs, und verdeutlicht die damit verbundenen Herausforderungen. Es wurde festgestellt, dass es keine universelle Anwendung für alle Wissensgraphen geben kann, da sie sehr anwendungsspezifisch sind.

Darüber hinaus hat der Aktionsforschungsansatz gezeigt, dass die Integration des praktischen Kontexts die Entwicklung eines maßgeschneiderten Produkts ermöglicht, wobei die Flexibilität durch Iterationen gewahrt bleibt. Diese Arbeit liefert einen Entwurf für ein solches Verfahren und könnte beispielhaft für andere sein.



# Kapitel 2

## Grundlagen

### 2.1 (Forschungs-)wissensgraphen

In „Towards a Definition of Knowledge Graphs“ von Ehrlinger und Wöß [13] wird die Notwendigkeit betont, klare Definitionen für Wissensgraphen zu entwickeln und sie von anderen Konzepten abzugrenzen. Dies ist entscheidend, um fundierte Aussagen über die Einführung und Verbreitung von Wissensgraphen treffen zu können.

Für eine Definition von Wissensgraphen wird die Arbeit von Hogan et. al [22] herangezogen:

#### Wissensgraph

*We define a knowledge graph as a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent potentially different relations between these entities. The graph of data (a.k.a. data graph) conforms to a graph-based data model, which may be a directed edge-labelled graph, a heterogeneous graph, a property graph [...]*

Weiterhin existieren formalere Definitionen, die eine explizite Verwendung von RDF (Resource Description Framework) Tripeln postulieren [15]. In Färber et. al wird der Begriff *Wissensgraph* synonym für jeden RDF-Graphen verwendet. Jedes RDF-Tripel besteht aus (Subjekt, Prädikat, Objekt), wobei diese geordnete Menge Knoten vom Typ IRI (Internationalised Resource Identifier), Literal und Blank Node enthält [15, 22].

Forschungswissensgraphen werden in dieser Arbeit zu allgemeinen Wissensgraphen abgegrenzt, da sie eine spezielle Form von Wissensgraphen darstellen. Sie fokus-

sieren sich auf den wissenschaftlichen Kontext und enthalten typischerweise Metadaten, wie Autor:innen, Tagungsorte, Organisationen, Forschungsthemen und Zitate um die Forschungspublikationen zu beschreiben [9].

Ausschließlich Metadaten sind für Wissenschaftler:innen bei der Beantwortung von Forschungsfragen von geringem direkten Wert, so argumentieren Stocker et. al [54]. Ihnen zufolge benötigt es Infrastrukturen, die den tatsächlichen Inhalt der Forschungsartikel in großem Umfang digital zur Verfügung stellen, so zum Beispiel der Open Research Knowledge Graph (ORKG)<sup>1</sup>.

Laut Karras et. al [30] repräsentiert ein Wissensgraph im Kontext wissenschaftlicher Kommunikation semantisch, das heißt explizit und formal, originale Forschungsergebnisse und verknüpft umfassend bestehende Daten, Metadaten, Wissen und Informationsressourcen .

Weiterhin können Forschungswissensgraphen in *generisch* und *spezifisch* unterteilt werden [29]. Demnach fokussieren sich generische Forschungswissensgraphen auf die Beschreibung von bibliografischen Metadaten (z.B. Titel, Autor:innen, Tagungsorte). Beispiele für generische Forschungswissensgraphen sind der OpenAIRE ResearchGraph [36], der Microsoft Academic Knowledge Graph [14] oder OpenAlex [41]. Wohingegen spezifische Forschungswissensgraphen auf wissenschaftliche Daten (z.B. Forschungsgestaltung, Methoden, Messungen und Ergebnisse) konzentriert sind. Beispiele für spezifische Forschungswissensgraphen sind der CovidGraph [11] oder SoftwareKG [46].

Im Allgemeinen, müssen Wissensgraphen gewartet und aktualisiert werden um sich den wandelnden Anforderungen anzupassen [56]. Dies ist nach Wang et. al [56] notwendig, da Wissensgraphen zwangsläufig Fehler enthalten und daher die Verbesserung ihrer Qualität von essenzieller Bedeutung ist.

Laut Yaser et. al [26] ist im Gegensatz zur typischen Verwendung von Wissensgraphen für enzyklopädisches und Allgemeinwissen die Verwaltung wissenschaftlichen Wissens erheblich anspruchsvoller aufgrund der Heterogenität, Inhomogenität und Evolution der wissenschaftlichen Kommunikation.

Manuell und automatisiert generierte Wissensgraphen werden über die Zeit sehr groß. Das führt unweigerlich zu Qualitätsproblemen. In „Temporal Dimensions of Quality in Knowledge Graph Evolution“ von Atig et. al [2] wird erläutert, wie durch die Integration von diversen Datenquellen die Qualität von Wissensgraphen über die Zeit kompromittiert wird:

Ein Problem in Bezug auf die Qualität besteht in der Präzision, die sowohl die grammatikalische als auch die semantische Richtigkeit betrifft und einen Vergleich mit der Realität ermöglicht. Zusätzlich spielt die Aktualität der Präzision eine Rolle,

---

<sup>1</sup><https://orkg.org/>

da ein Wissensgraph ungenau werden kann, wenn er nicht regelmäßig aktualisiert wird. Eine weitere Qualitätsmetrik betrifft die Kompaktheit, wobei darauf geachtet wird, dass kein Informationsüberfluss entsteht und keine redundanten Daten in den Wissensgraphen eingefügt werden. In ihrer Studie betonen Atig et al. [2], dass bei Frameworks, die Qualitätsmängel in Wissensgraphen aufdecken, sorgfältig zwischen unabhängigen und spezifischen Ansätzen abgewogen werden sollte. Hierbei findet ein Kompromiss zwischen hoher Genauigkeit an den jeweils angepassten Wissensgraphen und geringer Generalisierbarkeit sowie umgekehrt statt.

Des Weiteren adressieren Frameworks meist entweder die Detektion von fehlerhaften Inhalten oder die Korrektur dessen. Um eine kontinuierliche Qualitätssicherung, diese kann unter dem Begriff der Kuratierung zusammengefasst werden [25], zu gewährleisten, benötigt es einen angepassten Wartungsprozess, der auf den vorliegenden Wissensgraphen zugeschnitten ist.

## 2.2 Kuratierung von Wissensgraphen

Um einen Einblick in den Kuratierungsprozess von Wissensgraphen zu geben, wird die Arbeit von Fensel et. al herangezogen [16]. In ihrem Buch illustrieren Fensel et. al die Kuratierung von Wissensgraphen als einen repetitiven Prozess.

Hierbei gliedert sich die Wissensbewertung (knowledge assessment) in die Wissensbereinigung (knowledge cleaning) und Wissensbereicherung (knowledge enrichment) anhand zwei zentraler Qualitätsdimensionen: Korrektheit und Vollständigkeit. Anhand weiterer Eigenschaften wie Zugänglichkeit, Flexibilität oder Interoperabilität wird die Qualität eines Wissensgraphen beurteilt. Das Ziel besteht darin, einen Überblick über falsche oder fehlende Aussagen zu treffen.

Die *Wissensbereinigung* strebt danach, Richtigkeit zu verbessern indem Fehler erkannt und korrigiert werden. Hierbei ist es wichtig, zwischen Verifikation und Validierung zu unterscheiden. Die Verifikation erfolgt durch die Evaluierung anhand formaler Spezifikationen von Integritätsbedingungen, während die Validierung sicherstellt, dass die Aussagen das jeweilige Fachgebiet korrekt beschreiben.

Die *Wissensanreicherung* hat den Anspruch, die Vollständigkeit zu verbessern, indem neue Aussagen hinzugefügt werden. Dabei ist es eine Herausforderung, potenzielle Wissensquellen zu identifizieren, bedingt durch interne und externe Heterogenität, überlappende und fehlende Informationen. Die Automatisierung dieses Prozesses ist begrenzt möglich, da bestehende Tools vorherige Konfigurationen benötigen und auf bestimmte Anwendungsfälle zugeschnitten sind. Die Bewältigung verschiedener Probleme, darunter die entity resolution, die Erkennung von Duplikaten und

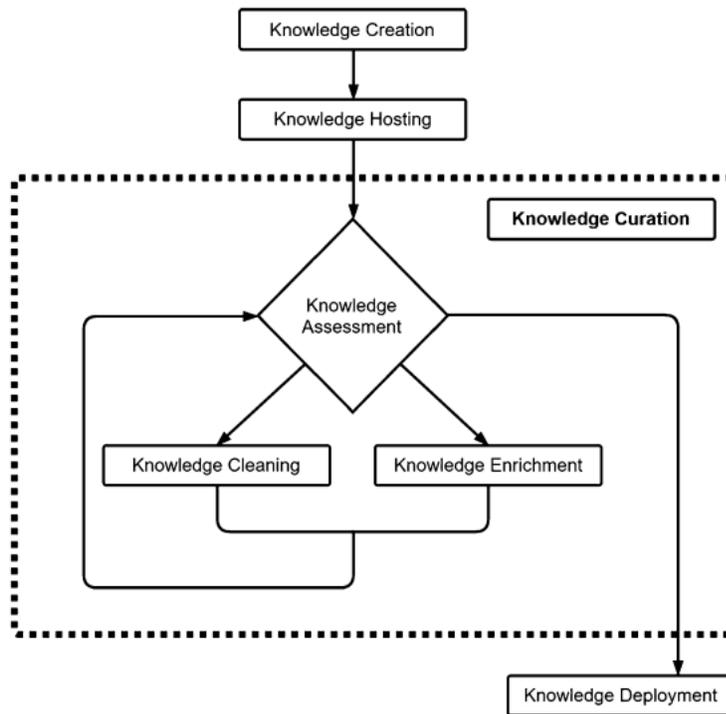


Abbildung 2.1: Prozessmodell für Wissensgraphgenerierung

author-name disambiguation, unter Beachtung von Skalierbarkeit sind zentrale Aufgaben in der Wissensanreicherung.

Das Errichten eines Wissensgraphen stellt eine gewisse Herausforderung dar, jedoch erfordert die Umwandlung desselben in eine wirkungsvolle Ressource zusätzlichen Aufwand. Huaman et al. [25] betonen, dass Wissensgraphen, die nicht aktiv gepflegt werden, im Laufe der Zeit nicht nur an Qualität einbüßen, sondern ebenso Anwendungen hervorbringen, die auf einer unzuverlässigen Wissensquelle basieren und somit ebenfalls von minderwertiger Qualität sind.

## 2.3 ORKG Kuratierung und Gemeinschaftsbildung

Nachdem wir die Herausforderungen der Pflege von Wissensgraphen allgemein betrachtet haben, werfen wir nun einen genaueren Blick auf die besondere Dynamik der Kuratierung im Kontext des Open Research Knowledge Graphs (ORKG).

Seit seiner Entstehung im Jahr 2018 [21] verfolgt der ORKG das Ziel, dokumen-

tenbasierte Informationen in eine wissensbasierte, semantische und kontextsensitive Form zu transformieren [3]. Das ORKG-Projekt revolutioniert die wissenschaftliche Zusammenarbeit, indem es eine Plattform bereitstellt, auf der Forschende ihre Erkenntnisse teilen, vergleichen und visuell darstellen können. Forschende können in einem zentralen System agieren, was die gemeinsame Bearbeitung und das Teilen von Daten erleichtert. Der Zugriff auf relevante Beiträge ermöglicht es, schnell auf aktuelle Erkenntnisse zuzugreifen und sie zu vergleichen. Zudem kann das Wissen aus verschiedenen Bereichen interdisziplinäre Forschung akzelerieren und innovative Ansätze fördern. Durch die transparente Darstellung von Forschungsergebnissen im ORKG wird Reproduzierbarkeit von Studien unterstützt sowie die Vertrauenswürdigkeit der Forschung gefördert.

Der Open Research Knowledge Graph als wissenschaftliche Ressource erfordert eine fortlaufende anwendungsspezifische Pflege, insbesondere vor dem Hintergrund einer stetig wachsenden Nutzergemeinschaft. Dabei wird deutlich, dass menschliche Kuratierung durch Domänenexpert:innen von entscheidender Bedeutung ist. Der ORKG handhabt Daten mit hoher semantischer Korrektheit, und die Herausforderung besteht darin, die Qualitätseigenschaften zu detektieren um sie genauer zu untersuchen. Dies wird umso anspruchsvoller, da der ORKG kontinuierlich an Größe zunimmt. Eine automatisierte Herangehensweise allein mag nicht ausreichen, da menschliche Expertise erforderlich ist, um die Vielschichtigkeit und Kontextualität der Daten angemessen zu bewerten und sicherzustellen. In diesem Zusammenhang spielt die Entwicklung und Implementierung eines auf den ORKG zugeschnittenen Wartungsprozesses eine zentrale Rolle, um eine nachhaltige Qualitätssicherung zu gewährleisten.

So werden im ORKG Daten aus Veröffentlichungen vorrangig durch Crowdsourcing von Paperautor:innen und Domänenexpert:innen extrahiert, strukturiert und manuell in den Wissensgraphen eingefügt.

Um (semantische) Korrektheit und Vollständigkeit sicher zu stellen, hat sich ein Team gebildet, was sich dieser Aufgabe annimmt. Im Kuratierungsteam des Open Research Knowledge Graph (ORKG) liegen vielfältige Aufgaben vor, die über die einfache Beurteilung von Qualität und Sinnhaftigkeit des erstellten Wissens hinausgehen. Um nicht nur inhaltliche Richtigkeit zu bewerten, sondern ebenso die korrekte Anwendung des ORKG als Tool durch Nutzende sicher zu stellen, erfordert es eine hohe Domänenexpertise.

In der fortlaufenden Evaluation der Effektivität des ORKG werden verschiedene Aspekte von besonderer Bedeutung betrachtet. Die Untersuchung der Wiederverwendbarkeit von Ressourcen ist dabei von zentraler Relevanz. Das übergeordnete Ziel besteht vielmehr darin, die bestehenden Ressourcen kontinuierlich zu vertiefen

und auszubauen, anstatt fortlaufend neue zu erstellen und dadurch den Wissensgraphen unnötig zu überladen. Es stellt sich die Frage, inwieweit der aufgebaute Graph in der Lage ist, kontextbezogene Informationen angemessen zu repräsentieren, was eine entscheidende Rolle für die Nutzung des ORKG spielt.

Das Kuratierungsteam legt zusätzlich einen Fokus darauf, Muster zu erkennen, um zu verstehen, warum bestimmte Elemente besser funktionieren als andere. Auf Basis dieser Erkenntnisse werden Empfehlungen für die Optimierung und Weiterentwicklung des ORKG abgeleitet. In diesem Zusammenhang besteht eine enge Verknüpfung mit der Gemeinschaftsbildung, deren Ziel es ist, Nutzerdaten zu analysieren. Hierbei steht insbesondere die Frage im Fokus, ob die Nutzer:innen den ORKG als intuitiv empfinden. Anhand vielversprechend modellierten Ressourcen wird eine Orientierung für andere Nutzende gegeben. Gleichzeitig strebt das Team danach, mehr Nutzer:innen zur aktiven Beteiligung zu motivieren. Dazu wird nach interessanten und vielfältigen Quellen gesucht, die den Open Research Knowledge Graph (ORKG) bereichern könnten.

Aufgrund der Diskrepanz von Daten, die für den Kuratierungsprozess notwendig sind und der Art, wie diese in der Benutzeroberfläche zur Verfügung gestellt sind, ist eine zugeschnittene Lösung für das Team nötig. Das entwickelte Dashboard ruft vorhandene Datensätze aus dem ORKG ab und bereitet diese visuell auf um weitere relevante Zusammenhänge und zusätzliche Möglichkeiten für die Kuratierung zu erschließen.

## 2.4 Dashboardspezifische Methoden

Um den Begriff Dashboard zu kontextualisieren, greift diese Arbeit auf die Erkenntnisse von Stephen Few zurück [17]:

### **Dashboard**

*A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance.*

Dieses Kapitel lehnt sich weiterhin an „Information Dashboard Design: The Effective Visual Communication Of Data“ von Stephen Few [17] an.

Das Design von Dashboards stellt eine erhebliche Herausforderung dar, da es darum geht, umfangreiche Daten so zu kondensieren, dass sie auf einen einzigen Bildschirm passen. Stephen Few kategorisiert Dashboards zunächst in drei Typen: strategisch, operativ und analytisch. Diese Kategorien dienen dazu, Dashboards nach ihrer Rolle und Funktion zu unterscheiden und beeinflussen somit maßgeblich die

Designausrichtung.

*Strategische* Dashboards legen den Fokus auf die Hervorhebung langfristiger Trends, um Entscheidungsprozesse von Führungskräften zu unterstützen. Auf diesen Dashboards werden zentrale Gesamtmetriken präsentiert, die einen Einblick in zukünftige Entwicklungen ermöglichen. Diese Dashboards zeichnen sich durch eine einfache, statische Darstellung von Momentaufnahmen aus, ohne interaktive Funktionen.

*Analytische* Dashboards setzen auf komplexe Visualisierungen, um eine tiefgehende Analyse der Daten zu ermöglichen. Die Diagramme bieten die Flexibilität zur Anpassung, um detaillierte Zusammenhänge eingehend zu erforschen. In diesem Kontext spielt die Bereitstellung von relevanten Kontextinformationen eine entscheidende Rolle, um eine präzise Einordnung der Daten zu gewährleisten.

*Operative* Dashboards setzen auf die Integration von Echtzeitdaten, um sicherzustellen, dass die präsentierten Informationen kontinuierlich auf dem neuesten Stand sind. Diese Dashboards zeichnen sich durch ihre dynamischen Funktionen aus, indem sie klare und einfach verständliche Informationen bereitstellen, die die Aufmerksamkeit unmittelbar auf sich ziehen. Sie ermöglichen es den Nutzern, direkte Operationen durchzuführen und auf diese Weise effizient mit den aktuellen operativen Prozessen zu interagieren.

Ein zentrales Prinzip für wirksame Dashboards ist die Anpassung an spezifische Anforderungen von Personen, Gruppen oder Funktionen. Die Anordnung der Informationen auf dem Bildschirm spielt eine entscheidende Rolle. Dabei sollten Elemente in Gruppen organisiert werden, die anhand von Geschäftsfunktionen (Aufgabenbereiche), Entitäten (Abteilungen, Projekte oder Systeme) und Verwendungszweck unterschieden werden. Dabei erfolgt die Fragmentierung der Daten in einer Weise, die sicherstellt, dass lediglich solche Daten zusammengeführt werden, die sinnvolle Vergleiche ermöglichen.

Das zu entwickelnde Dashboard orientiert sich an den zuvor erläuterten Designprinzipien. Es handelt sich um eine maßgeschneiderte Anwendung, die speziell für das ORKG Kuratierungs- und Gemeinschaftsbildungsteam konzipiert ist. Dabei werden Optionen zur Modularität und Erweiterbarkeit implementiert, um auch anderen Wissensgraphen gerecht zu werden. Im weiteren Verlauf werde ich gegebene Use Cases nach den drei präsentierten Typen kategorisieren um eine möglichst sinnvolle Strukturierung des Dashboards zu ermöglichen. Diese Zuordnung wird eine Überlappung der vorgestellten Kategorien offen legen (vgl. Tabelle 6.3), wobei hauptsächlich User Stories des operativen und strategischen Typs umgesetzt werden.

## 2.5 Aktionsforschung

Diese Arbeit orientiert sich an der Aktionsforschung von Miroslaw Staron [53]. Diese Forschungsmethode zeichnet sich durch einen Zyklus aus, der als Aktionsforschungszyklus bekannt ist (vgl. Abbildung 2.2). Jeder Zyklus beginnt mit der Diagnose eines Problems und schließt mit einer dokumentierten Reflexion über das Gelernte ab.

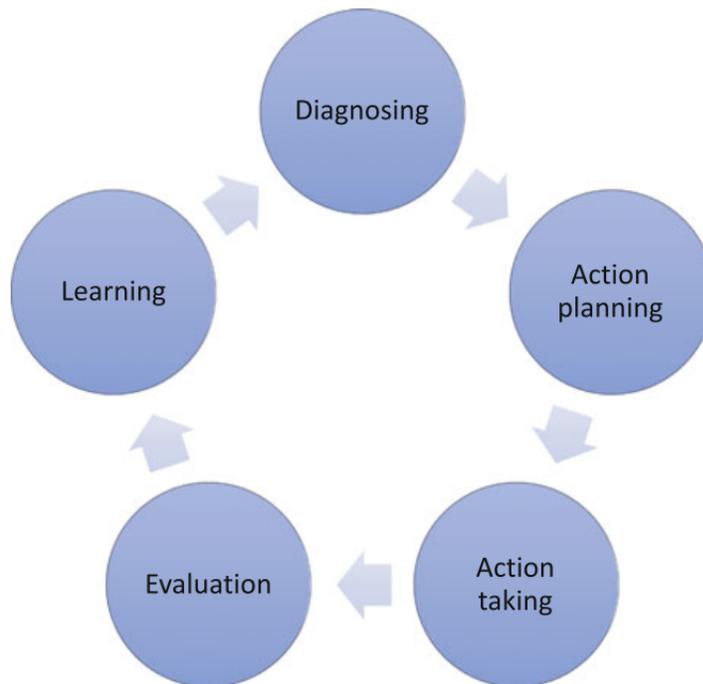


Abbildung 2.2: Aktionsforschungszyklus

Ein charakteristisches Merkmal der Aktionsforschung ist die Zusammenarbeit zwischen Industrie und Akademie. Die Praktiker:innen aus dem industriellen Umfeld haben den Bedarf, ihre Organisation, ihre Produkte oder ihre Abläufe zu verbessern. Die Forschenden hingegen möchten Methoden im industriellen Kontext auf wissenschaftlichem Niveau evaluieren oder validieren. Dadurch entsteht eine kollaborative Arbeit, von der beide Seiten mit Wissen, Theorien, Methoden und Werkzeugen profitieren. Durch diese Herangehensweise können akademische Forschungsergebnisse anwendungsorientierter und die Forschung von Praktiker:innen anspruchsvoller gestaltet werden.

Das Team, das für die Planung, Umsetzung und Evaluierung der Forschung zuständig ist, wird als Aktionsteam bezeichnet. Die Referenzgruppe steht bereit,

um das Aktionsteam zu beraten und Feedback zu geben. Auf diese Weise wird das Aktionsteam in den industriellen Kontext eingebunden, und der Forschungszyklus kann mit der einzigartigen *Diagnosephase* eingeleitet werden. Es ist üblich, mit einer formulierten Problemstellung in ein Projekt zu starten, jedoch steht hier die Erkundung des Kontextes im Vordergrund. Durch Diskussionen mit der Referenzgruppe wird herausgearbeitet, was im Aktionszyklus umgesetzt werden soll. Die Diagnose kann mit der Anforderungserhebung (requirements elicitation) im allgemeinen Software Engineering gleich gesetzt werden. Es gibt zwei Arten der Bestimmung von Problemstellungen: entweder erkennt ein:e Aktionsforscher:in die Probleme selbst (forschungsgetriebene Initiation) oder die Probleme werden dem:der Aktionsforscher:in präsentiert (problemgetriebene Initiation).

Nachdem für beide Parteien klar ist, welche Herausforderung im Zyklus angegangen werden soll, geht es in die nächste Phase: die *Planung*. Die kooperative Natur dieser Phase ermöglicht es Praktiker:innen und Forschenden, gemeinsam über Lösungswege für das diagnostizierte Problem zu diskutieren, Ressourcen, Produkte sowie Prozesse zu identifizieren und anzupassen. Während der Aktionsplanungsphase stimmt das Forschungsteam seine Pläne mit der Referenzgruppen ab und benötigt die Zustimmung der Managementebene für erforderliche Ressourcen. Dabei werden die Pläne an die theoretischen Grundlagen der Arbeit angepasst, und es wird festgelegt, welche Analysemethoden verwendet werden, um den Erfolg der Maßnahmen zu bewerten.

Die *Maßnahmengreifungsphase* ist charakteristisch für die Aktionsforschung, da sie Veränderungen direkt im laufenden Betrieb ermöglicht. In dieser flexiblen Methodik kann das Aktionsteam gemäß den zuvor erstellten Plänen sowohl Anpassungen vornehmen als auch die Auswirkungen beobachten. Die Interventionen werden aktiv vom Aktionsteam umgesetzt, wobei die Referenzgruppe regelmäßig eingebunden wird, um Feedback zu geben und bei der Bewältigung auftretender Herausforderungen zu unterstützen. Hierbei werden quantitative Nachweise geliefert, die in der nächsten Evaluierungsphase verwendet werden, um die Wirksamkeit der durchgeführten Maßnahmen zu beurteilen.

In der *Evaluierung* werden die zuvor gesammelten Daten statistisch analysiert. Die Ergebnisse werden der Referenzgruppe präsentiert. Falls die Daten keine klare Schlussfolgerung zulassen, muss das Aktionsteam entweder zusätzliche Analysen vornehmen oder die aktuelle Phase abschließen, das Gelernte spezifizieren und basierend auf den neuen Daten eine neue Problemstellung diagnostizieren. Zusätzlich ist es wichtig, dass die Datenanalyse mit den in diesem Zyklus verwendeten Theorien übereinstimmt, um einen Beitrag zur Theoriebildung in der nächsten Phase zu leisten.

Der abschließende Teil des Aktionsforschungszyklus ist die Spezifikation des *Lernens*. Dies geschieht sowohl als praktische Leitlinien für die beteiligten industriellen Kontexte als auch als Beitrag zur Theoriebildung für die Forschungsgemeinschaft. Die praktischen Leitlinien spezifizieren bewährte Praktiken und dokumentieren gute Beispiele. Der Beitrag zur Theoriebildung wird oft in wissenschaftlichen Arbeiten festgehalten.

# Kapitel 3

## Verwandte Arbeiten

Dieses Kapitel bietet einen umfassenden Überblick über die gegenwärtigen Methoden und Werkzeuge zur Bewertung von Wissensgraphen, wobei ein besonderer Fokus auf der Qualitätsbewertung und den visuellen Darstellungen dieser liegt.

Da die manuelle Erstellung von Wissensgraphen nicht skalierbar ist und die vollständig automatische Erstellung fehleranfällig ist, sind die meisten Ansätze zur Erstellung von Wissensgraphen halbautomatisch [45, 35]. Hierbei wird meist ein Kompromiss zwischen Genauigkeit und Effizienz angestrebt.

Im Rahmen der Erstellung von Wissensgraphen ist es wichtig, die Kuratierung von Wissensgraphen zu betrachten. Hierbei geht es um die Beurteilung der Qualität des Inhalts, die sich in zwei Hauptaspekte gliedert: Bereinigung und Bereicherung (vgl. Kapitel 2).

Schatz et. al [45] stellen einen Workflow (KG-CURATE) vor, um Herausforderungen im Zusammenhang mit der Benutzerfreundlichkeit von Wissensgraphen anzugehen. Dabei erfolgen die Schritte der Datenextraktion aus einer Textdatei und die Datenintegration in den Wissensgraphen automatisiert. Der abschließende Schritt des Workflows umfasst eine Phase der Datenbereinigung, in der eine repräsentative Stichprobe der Daten manuell überprüft wird, um präzise erneute Iterationen der Datenintegration zu ermöglichen.

Damit wird deutlich, dass trotz fortschrittlicher Technologie die menschliche Expertise weiterhin unverzichtbar ist, insbesondere in Bezug auf die Sicherstellung der Datenqualität.

Im Zuge dessen geben numerische Indikatoren einen Überblick über Größe und Struktur des Wissensgraphen, während Graphmetriken die Qualität hinsichtlich Organisation und Beziehungsvielfalt bewerten, jedoch sind diese nicht speziell auf die Qualitätsbewertung von Wissensgraphen ausgerichtet [48].

---

Huaman [24] stellt ein Framework vor, das Nutzer:innen ermöglicht, die Qualität eines Wissensgraphen anhand quantitativer und qualitativer Metriken zu bewerten. Dieses benutzerzentrierte Bewertungsframework gewichtet verschiedene Qualitätsdimensionen und -metriken, um einen Qualitätsstatus des Wissensgraphen für spezifische Anwendungsfälle zu ermitteln und Empfehlungen zu geben, wann solche Wissensgraphen eingesetzt werden sollten.

Im Kontext des Open Research Knowledge Graph (ORKG) ist das Crowdsourcing von entscheidender Bedeutung, da es maßgeblich dazu beiträgt, das Niveau der vorhandenen Datenqualität zu bestimmen. Durch die Beteiligung der Nutzer:innen erhalten diese direkten Einfluss darauf, welche Daten in den Wissensgraphen integriert werden.

Die Arbeit von Ahmeti et. al [1] stellt einen Crowdsourcing-Ansatz vor, der es Nutzer:innen ermöglicht, durch ein Human-in-the-Loop-Modell Wissen zu einem großen Wissensgraphen beizutragen. Automatische Schlussfolgerungsmechanismen prüfen dabei die Konsistenz der von Nutzer:innen eingebrachten Informationen, während verschiedene Abstimmungsverfahren es den Teilnehmenden ermöglichen, über die Korrektheit und Relevanz der neuen Beiträge abzustimmen. Nur wenn die Differenz zwischen positiven und negativen Stimmen einen bestimmten Schwellenwert erreicht, werden die vorgeschlagenen Änderungen in den Wissensgraphen integriert, wodurch die Qualität der eingetragenen Daten sichergestellt wird.

Für bestehende Ansätze zeigt sich, dass einfache Aufgaben durch Automatisierung (z.B. Extraktion von strukturierten Daten) erledigt werden können, während komplexe Probleme (z.B. Kuratierung) weiterhin menschliche Expertise erfordern.

Im Rahmen des ORKG werden verschiedene Anwendungen entwickelt, die die Daten im ORKG nutzen, um sie visuell darzustellen:

Das Dashboard von Mahsa Shamsabad<sup>1</sup> nutzt künstliche Intelligenz um virologische Kennzahlen zu vergleichen und in geeigneten visuellen Formaten darzustellen [49]. Diese Visualisierungen bieten einen übersichtlichen Einblick in die Daten und erleichtern die Interpretation.

Das Dashboard von Olga Lezhnina<sup>2</sup> konzentriert sich auf die Zusammenstellung relevanter Forschungsberichte im Zusammenhang mit dem PISA-Rahmenwerk [33]. Durch die Nutzung des ORKG können diese Berichte effizient gesammelt und präsentiert werden, um einen umfassenden Überblick über das Forschungsfeld zu ermöglichen.

Die R Shiny App<sup>3</sup> von Bernard-Verdier et. al [5] stellt eine interaktive Visua-

---

<sup>1</sup><https://orkg.org/usecases/r0-estimates>

<sup>2</sup><https://orkg.org/usecases/pisa-dashboard/>

<sup>3</sup><https://maudbernardverdier.shinyapps.io/Hypothesis-evidence-explorer/>

lisierung und Exploration von Daten zur Invasionsbiologie dar. Sie zeigt kuratierte Abbildungen und Zusammenfassungsverstatistiken zu zehn Hypothesen in diesem Bereich. Die App ermöglicht es, einen Überblick über den aktuellen Wissensstand zu gewinnen und die Daten nach relevanten Eigenschaften zu filtern. Dabei wurde der Open Research Knowledge Graph (ORKG) genutzt, indem die Daten mithilfe des ORKG-Pakets für Python exportiert und in die App integriert wurden. So konnte die Strukturierung und Visualisierung der Daten effizient durchgeführt werden, um einen umfassenden Einblick in die Literatur zur Invasionsbiologie zu bieten.

Kürzlich wurde ein neues Tool namens ORKG Ask<sup>4</sup> vom ORKG veröffentlicht. ORKG Ask ist ein wissenschaftliches Such- und Explorationssystem, das auf Vector Search, Large Language Models und Wissensgraphen basiert. Es ermöglicht es Nutzer:innen, gezielt nach wissenschaftlichen Inhalten zu suchen und diese zu erkunden. Mit ORKG Ask können Informationen effizient gefunden und Zusammenhänge zwischen verschiedenen wissenschaftlichen Konzepten aufgedeckt werden.

Ein genauerer Blick darauf, wie andere Wissensgraphen Dashboards einsetzen, bietet wertvolle Einsichten in die Vielfalt der Anwendungsmöglichkeiten von Visualisierungen.

DBpedia bietet ein interaktives Dashboard<sup>5</sup> an, mit dem Nutzer:innen Hierarchien und Instanzen erkunden können. Durch dieses Tool können Nutzende so ein besseres Verständnis für die Beziehungen zwischen den verschiedenen Begriffen und Entitäten in der DBpedia-Ontologie gewinnen.

Wikimedia stellt eine Übersicht in Form eines Dashboards zur Verfügung, welches vorrangig Nutzerstatistiken listet<sup>6</sup>.

Gemäß dem aktuellen Stand der Literatur wird in bisherigen Arbeiten nicht speziell auf Dashboards eingegangen, die auf einem Forschungswissensgraphen basieren [33]. Die vorliegende Arbeit präsentiert daher einen innovativen Ansatz zur Entwicklung von Dashboards. Das zu entwickelnde Dashboard nutzt Daten aus dem Forschungswissensgraphen ORKG, der auf Crowdsourcing basiert. Das Dashboard ist speziell für Kurator:innen entwickelt worden und legt den Schwerpunkt auf ihre Bedürfnisse. Im Gegensatz zu bisherigen Dashboards, die hauptsächlich darauf abzielten, den Nutzenden eine verbesserte Visualisierung der Daten zu bieten, bietet dieses Dashboard eine maßgeschneiderte Lösung für die Kuratierung und Gemeinschaftsbildung, die hier als eine spezielle Gruppe von Nutzenden betrachtet werden. Dabei sind modulare Erweiterungsmöglichkeiten vorgesehen, um auch andere Wissensgraphen integrieren zu können.

---

<sup>4</sup><https://ask.orkg.org/>

<sup>5</sup><https://www.dbpedia.org/resources/ontology/>

<sup>6</sup><https://stats.wikimedia.org/#/wikidata.org>



# Kapitel 4

## Methodik

Diese Arbeit nutzt das Design-Science-Paradigma [44] als Forschungsansatz, welches sich als wirksames Werkzeug zur Lösung praktischer Probleme im Bereich des Software Engineering erwiesen hat. Die übergeordnete Methodik umfasst die Entwicklung eines Minimal Viable Products (MVP) sowie dessen Evaluierung. Dabei werden zwei Zyklen des Aktionsforschungsansatzes integriert, um eine iterative und kooperative Verbesserung des MVP zu ermöglichen. Durch die Kombination von Design-Science und Aktionsforschung wird sichergestellt, dass das entwickelte MVP nicht nur theoretisch fundiert ist, sondern auch praktisch relevant und anwendbar in realen Szenarien. Das gewünschte Ergebnis besteht darin, das Dashboard gemäß den Anforderungen des Teams und im Rahmen der technischen Möglichkeiten zu optimieren. Dabei hat das Dashboard den Anspruch, so weit entwickelt zu sein, dass das Team die Anwendung effektiv für ihre zukünftige Arbeit nutzen kann und es Möglichkeiten zur Modularität und Erweiterung bietet. Das Dashboard fungiert zunächst als Werkzeug um bestehende Daten aufzubereiten und sie geeignet darzustellen. Durch die Anwendung werden keine Änderungen im zugrunde liegenden Wissensgraphen des ORKG vorgenommen.

Wie bereits in den Grundlagen beschrieben (vgl. Abschnitt 2.5), setzt die Aktionsforschung auf eine enge Zusammenarbeit von einer realen Organisation und Forschung. Im vorliegenden Kontext fungiert das Team des Open Research Knowledge Graph der TIB-Leibniz Informationszentrum Technik und Naturwissenschaften als Partner der realen Organisation. Die aktive Beteiligung des Teams an der Dashboard-Entwicklung ist entscheidend für den Erfolg der Arbeit. Die anwendungsorientierte Natur der Aktionsforschung ermöglicht es, konkrete Bedürfnisse des Teams zu identifizieren und sie in Form von User Stories umzusetzen um praxisnahe Lösungen zu entwickeln. Das Ziel der Anwendung ist es, Kuratierungsprozesse zu verbessern und

---

die Gemeinschaftsbildung zu unterstützen.

Weiterhin verfolgt die Arbeit den Zweck, das Dashboard in bestehende Kuratierungsprozesse zu integrieren und zukünftig neue Abläufe für die Gemeinschaftsbildung zu ermöglichen. Durch diesen Arbeitsfluss kann das Dashboard langfristig bei der Optimierung des ORKG für Nutzer:innen mit einbezogen werden. Es handelt sich hierbei um eine maßgeschneiderte Lösung für das Kuratierungs- und Gemeinschaftsbildungsteam des ORKG. Wie im Abschnitt 2.2 über die allgemeine Kuratierung von Wissensgraphen postuliert, erfordert dies oft einen anwendungsspezifischen Ansatz, da die Struktur von Wissensgraphen einzigartig ist. Ebenso ermöglichen die iterativen Zyklen der Aktionsforschung, kontinuierlich Feedback zu sammeln, darauf zu reagieren und somit eine optimierte Anwendung zu entwickeln. Die Dynamik von Wissensgraphen, insbesondere des ORKG, stellt eine komplexe Herausforderung dar, die eine adaptive Forschungsmethode wie die Aktionsforschung erfordert.

Die Abbildung 4.1 visualisiert die angewendete Methodologie bestehend aus dem übergeordneten Design Science-Ansatz und der eingebetteten Aktionsforschung. Das Design Science-Paradigma enthält hier eine abschließende Evaluation des erstellten MVP in Form eines Dashboards für das Kuratierungs- und Gemeinschaftsbildungsteams.

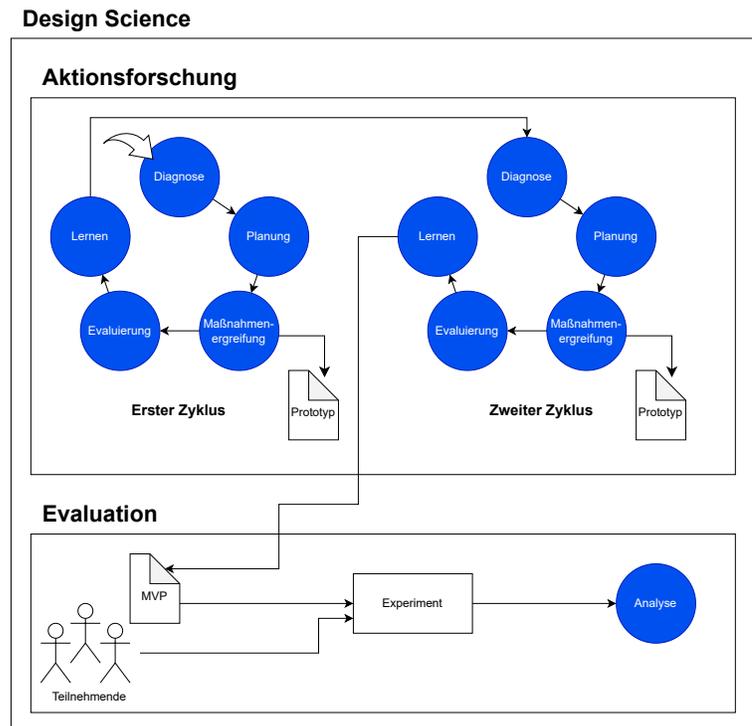


Abbildung 4.1: Angewendete Methodologie

Die Aktionsforschung wird mit zwei Zyklen umgesetzt. In der Diagnosephase werden zunächst die User Stories des Teams analysiert und priorisiert, später werden die Erkenntnisse des ersten Zyklus interpretiert und als Eingabe verwendet. Hierbei werden User Stories vorrangig in die Tiefe hingehend entwickelt, weniger als dass möglichst viele User Stories umgesetzt werden. Anschließend werden ausgewählte User Stories (Planung) implementiert und in einem Prototypen umgesetzt (Maßnahmenergreifung). Die Evaluierung umfasst einen *Pluralistic Usability Walkthrough* [23], der über Skype durchgeführt und aufgezeichnet wird. Im Gegensatz zum herkömmlichen Usability Walkthrough werden hier die Perspektiven von repräsentativen Nutzer:innen, Produktentwickler:innen und Usability-Expert:innen berücksichtigt. Diese pluralistischen Perspektiven ermöglichen eine Vielfalt an Blickwinkeln und einen partizipativen Ansatz. Insbesondere die Perspektive der zukünftigen Nutzenden ist in diesem Kontext von Interesse.

Diese Art der Evaluation kann wertvolle Einblicke für die Entwickler:innen liefern, da bereits in frühen Entwicklungsstadien Informationen über die Benutzererfahrung gesammelt und unmittelbare, flexible Designänderungen vorgenommen wer-

den können. Allerdings ist der Walkthrough bei der Bewertung eines Prototypen auf repräsentative Benutzerpfade beschränkt, was bedeutet, dass nur eine Auswahl typischer Benutzerinteraktionen durchgeführt wird, anstatt alle potenziellen Szenarien abzudecken. Somit wird in den Evaluierungen zunächst der aktuelle Stand des Prototypen präsentiert, woraufhin das Feedback des Teams dokumentiert wird und gegebenenfalls Diskussionen über Designentscheidungen und Realisierungsvorschläge entstehen. Bei der Evaluierung kristallisiert sich heraus, welche Komponenten Potential für eine zukünftige Benutzung und Erweiterung haben.

Durch die Aufzeichnung der Evaluierung ist es im Nachhinein möglich, Lehren und Erkenntnisse aus den Rückmeldungen zu gewinnen (Lernen), diese festzuhalten und damit eine Grundlage für den folgenden Zyklus zu schaffen. Basierend auf der Reflexion der Erkenntnisse werden iterative Verbesserungen am MVP vorgenommen, um dessen Leistungsfähigkeit und Nützlichkeit weiter zu optimieren. Während der Durchführung der Aktionsforschung hat das Team jederzeit die Möglichkeit Einfluss in die Entwicklung zu nehmen.

Welche **Akteur:innen** im Rahmen der Arbeit beteiligt sind, zeigt Tabelle 4.1. Als

Akteur:in	Team	Rolle
Lena John (ich)	Aktionsteam	Forscherin und Entwicklerin
Oliver Karras	Aktionsteam	Betreuer, Forscher und Moderator
Ildar Baimuratov	Aktionsteam	Betreuer, Forscher
Lars Vogt	Praxisteam	Leiter Kuratierungs- und Gemeinschaftsbildungsteam
Anna-Lena Lorenz	Praxisteam	Leiterin Gemeinschaftsbildung
Vinodh Ilangovan	Praxisteam	Koordinator Kuratierungsarbeit
Tim Wittenborg	Praxisteam	Doktorand
Marina Wurzbacher	Praxisteam	Forscherin
Markus Stocker	Praxisteam	Forscher, allgemeines ORKG-Team
Manuel Prinz	Praxisteam	Leiter Backendprogrammierung des ORKG
Yaser Jaradeh	Praxisteam	Programmierer des ORKG

Tabelle 4.1: Akteur:innen der Aktionsforschung

Entwicklerin werde ich das Dashboard implementieren und Evaluierungszeitpunkte festlegen. Oliver Karras übernimmt in seiner Rolle als Betreuer dieser Arbeit eine unterstützende Funktion, um mir genauere Einblicke in die Abläufe des Teams zu

geben und allgemeine Empfehlungen für die Prozesse zu bieten. Zudem fungiert er als Moderator in den Evaluierungsdiskussionen, um zwischen mir und dem Team zu vermitteln. Weiterhin ist Ildar Baimuratov im Aktionsteam tätig und unterstützt bei Fragen zur Implementierung. Wöchentliche Treffen innerhalb des Aktionsteams dienen dazu, akute Probleme zu lösen und den Fortschritt sicherzustellen.

Manuel Prinz und Yaser Jaradeh sind als Entwickler des ORKG die Ansprechpartner für alle Belange, die das Abfragen von Daten des ORKGs betreffen. Sie ermöglichen individuelle Anpassungen in der Struktur des ORKGs, nehmen jedoch nicht an den Evaluierungen über Skype teil. Somit gelten sie als Mitglieder des erweiterten Praxisteam.

Das übrige Praxisteam setzt sich hauptsächlich aus den Mitgliedern des Kuratierungs- und Gemeinschaftsbildungsteams zusammen, die an den Evaluierungen teilnehmen. Lars Vogt spielt dabei eine besondere Rolle, da er teilweise auch in Designfragen einbezogen wird und maßgeblichen Einfluss auf die Realisierung der Anwendung hat, was für das Kuratierungsteam von großer Bedeutung ist.

Die Aktionsforschung endet mit der Erstellung eines Minimal Viable Products (MVP), das die Erkenntnisse aus dem zweiten Zyklus umsetzt. Dieser Prototyp wird dann im abschließenden praktischen **Experiment** im Rahmen des Design Science evaluiert. Dabei nehmen zukünftige Nutzer:innen des Dashboards teil, um den tatsächlichen Nutzen bestmöglich zu bewerten. Die Analyse erfolgt mithilfe statistischer Auswertungsverfahren, und alle Daten werden anonymisiert verarbeitet.



# Kapitel 5

## Technische Basis

### 5.1 Technischer Aufbau des ORKG

Die Architektur des Open Research Knowledge Graph (vgl. Abbildung 5.2) basiert auf einer REST-API. Eine REST-API (Representational State Transfer - Application Programming Interface) ist eine Schnittstelle für die Kommunikation zwischen Anwendungen über das Internet (vgl. Abbildung 5.1). Sie ermöglicht den Austausch von Daten in einem standardisierten Format wie JSON oder XML. Über die API können Anwendungen auf Ressourcen zugreifen und Aktionen wie das Abrufen von Daten (GET), das Erstellen neuer Daten (POST), das Aktualisieren (PUT) oder Löschen (DELETE) durchführen. Die vorliegende Arbeit wird nur GET-Methoden verwenden, was im Falle des ORKG keine Authorisierung voraussetzt. Jede Ressource wird durch eine eindeutige URL identifiziert, und die API nutzt HTTP-Methoden, um mit ihnen zu interagieren. RESTful APIs sind zustandslos, was bedeutet, dass jede Anfrage unabhängig ist und keine vorherige Kontextinformationen speichert.

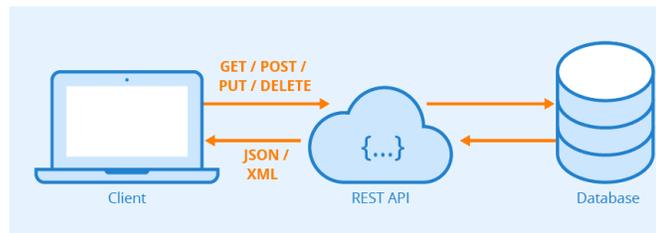


Abbildung 5.1: REST-API<sup>1</sup>

---

<sup>1</sup><https://www.astera.com/de/type/blog/rest-api-definition/>

## 5.1. Technischer Aufbau des ORKG

Die REST-API des Open Research Knowledge Graph (ORKG) bildet die Basis des Systems und ermöglicht die Interaktion mit den darin enthaltenen Ressourcen. Sie basiert auf den Datenbanken Neo4j und PostgreSQL. Während die Neo4j-Datenbank die Graphdaten enthält, in denen Entitäten miteinander verknüpft werden, bildet sie die Grundlage des ORKG als Wissensgraph. In diesem Graphen werden komplexe Beziehungen zwischen verschiedenen Wissensentitäten dargestellt.

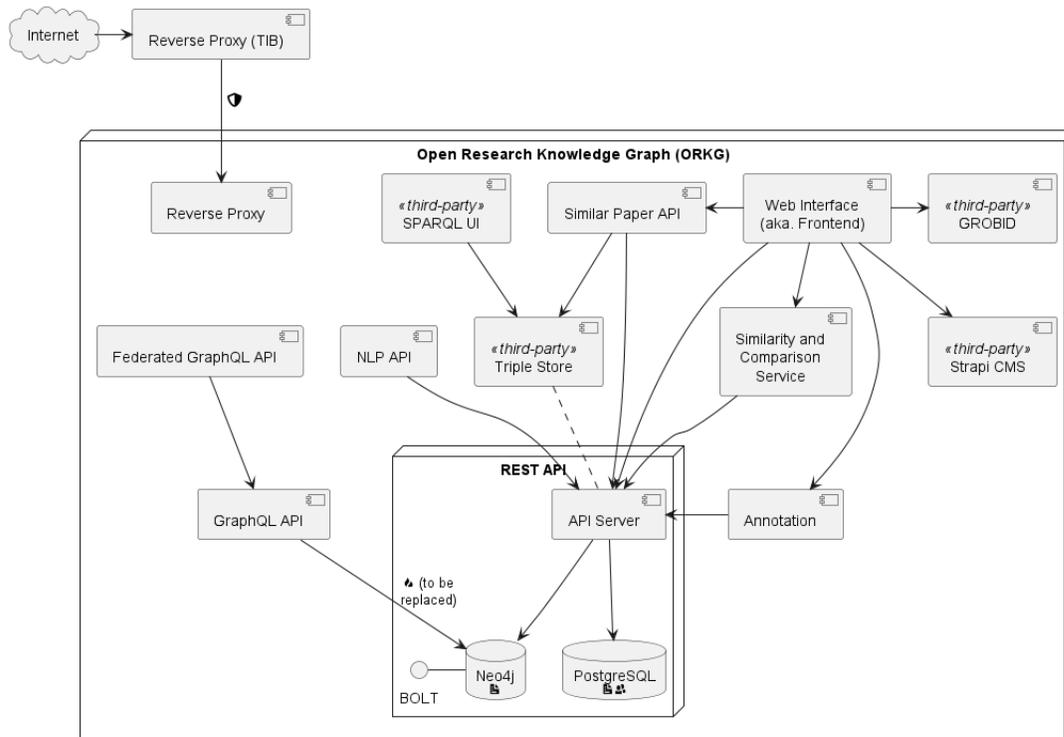


Abbildung 5.2: ORKG Komponenten

Die PostgreSQL-Datenbank hingegen speichert strukturierte Daten über die Nutzeraccounts. Diese Trennung besteht, da sensible Informationen aus Sicherheitsgründen separat abgespeichert werden und es sich um Daten handelt, die strukturierte Beziehungen haben, die sich nicht ändern.

Der API-Server ist die Komponente, die Anfragen von Benutzer:innen oder anderen Diensten entgegennimmt, verarbeitet und dann entsprechende Abfragen an die Datenbanken sendet, um die benötigten Informationen abzurufen oder zu aktualisieren. Er fungiert als Vermittler zwischen den Benutzer:innen und den Datenbanken und stellt sicher, dass die angeforderten Daten effizient und korrekt bereitgestellt

werden.

Darüber hinaus gibt es andere APIs, die auf die REST-API des ORKG zugreifen, um zusätzliche Funktionalitäten anzubieten. Zum Beispiel die Federated GraphQL-API, die es ermöglicht, Daten aus verschiedenen Quellen abzurufen und zu kombinieren, um einen umfassenderen Blick auf das Wissen im ORKG zu erhalten. Ebenso greift die NLP-API auf die REST-API zu, um Textanalysefunktionen anzubieten um die automatische Wissensextraktion aus wissenschaftlichen Berichten in den ORKG einzubetten.

In dieser Arbeit ist auch der Third-Party Triplestore von Bedeutung. Die Daten werden einmal täglich aus dem API-Server in RDF serialisiert und in den Triplestore importiert. RDF (Resource Description Framework) ist ein Standardformat zur Darstellung von Wissen im Web, das auf Tripeln basiert (Subjekt-Prädikat-Objekt). Die Serialisierung in RDF ermöglicht eine standardisierte Repräsentation der Daten, die sowohl maschinenlesbar als auch leicht verständlich ist. Durch den Triplestore wird ein SPARQL-Endpunkt bereitgestellt, über den individuelle gefilterte Anfragen an den Wissensgraphen gestellt werden können. Es ist jedoch wichtig anzumerken, dass bei diesem Prozess keine Provenance-Informationen übertragen werden. Das bedeutet, dass keine Daten über die Erstellungsdaten oder die Benutzer:innen verfügbar sind.

Des Weiteren werden in der Implementierung Ressourcen aus dem Similarity and Comparison Service (SimComp-API) abgefragt. Dieser Service wird zusammen mit dem zugrunde liegenden API-Server vom Frontend verwendet. Er ist unter anderem dafür verantwortlich, Daten für Comparisons bereitzustellen, die sich aus Papern und ihren Contributions zusammensetzen.

Die gesamte Architektur wird von außen durch einen Reverse Proxy geschützt. Dieser ermöglicht es, schädlichen Datenverkehr oder Angriffe vorab zu filtern. Der Proxy dient weiterhin dazu, alle Dienste unter einer Domain bereitzustellen, in dem er Pfade auf die einzelnen Dienste abbildet und die Anfragen an die richtigen Stellen leitet. Somit werden die einzelnen Dienste für Nutzer:innen transparent.

## 5.2 Technische Umsetzung des Dashboards

Das Dashboard wurde mit der JavaScript-Bibliothek React umgesetzt. Die Wahl von React wurde aufgrund seiner Fähigkeit, ein komponentenbasiertes Entwicklungsmodell umzusetzen, getroffen. Dies ermöglicht die Gliederung der Benutzeroberfläche in wiederverwendbare Komponenten, die unabhängig voneinander entwickelt und getestet werden können. Ein weiterer Vorteil von React ist die Verwendung eines Virtual DOMs, das nur die tatsächlich geänderten Teile der Benutzeroberfläche neu rendert. Dies trägt zu einer besseren Leistung der Anwendung bei.

Darüber hinaus wurde React bereits für die Benutzeroberfläche des ORKG verwendet, was bedeutet, dass das Dashboard leichter in die bestehende Struktur integriert werden kann. Dies ermöglicht es den Entwickler:innen, ihr Fachwissen einzubringen und die Implementierung des Dashboards zu unterstützen.

Die Entscheidung, direkte HTTP-Anfragen an das ORKG-Backend zu senden, wurde getroffen, obwohl der ORKG auch eine Python-Bibliothek<sup>2</sup> zur Verfügung stellt, die Daten vom ORKG abfragen kann. Diese Bibliothek erleichtert die Interaktion mit dem ORKG, da sie eine abstrahierte Schnittstelle bereitstellt und einige der Komplexitäten der direkten HTTP-Anfragen verbirgt. Insgesamt wurde die Entscheidung für direkte HTTP-Anfragen getroffen, um die Flexibilität, Kontrolle und Transparenz bei der Interaktion mit dem ORKG zu maximieren, während gleichzeitig die Abhängigkeit von zusätzlichen Bibliotheken minimiert wird.

Bestehende Plattformen zur Entwicklung von Dashboards, wie in Tabelle 5.1 gelistet, bieten zwar Optionen zur Anpassung, doch diese Möglichkeiten sind bei komplexeren Datenvisualisierungen begrenzt. Zudem sind die Architekturen teilweise auf bestimmte Anwendungsfälle zugeschnitten, beispielsweise die Integration von JupyterNotebooks. Daher kann ich durch die Verwendung von React und verschiedenen Bibliotheken das Dashboard nach den Wünschen und Bedürfnissen des Teams umfassend anpassen.

---

<sup>2</sup><https://pypi.org/project/orkg/>

Name	Beschreibung	Vorteile	Nachteile
Redash <sup>3</sup>	Plattform für Datenvisualisierung	benutzerfreundliche Oberfläche, Unterstützung für Vielzahl von Datenquellen	komplexere, individuellere Darstellungen nicht möglich
Panel <sup>4</sup>	Python-Framework	besonders geeignet für die Darstellung von mathematischen Daten	abhängig von externen Bibliotheken wie Plotly, Matplotlib
Dash <sup>5</sup>	Python-Framework für webbasierte Dashboards	einfache und schnelle Entwicklung	Einschränkungen bei Anpassungen im Vergleich zu JavaScript-Frameworks
Shiny <sup>6</sup>	R-basiertes Framework	nahtlose Integration mit R	begrenzte Anpassungsmöglichkeiten
Voila <sup>7</sup>	OpenSource-Bibliothek zur Umwandlung von JupyterNotebooks in Webanwendungen	schnelle Erstellung von interaktiven Dashboards aus bestehenden Notebooks	Einschränkungen bei Anpassung und Gestaltung von Benutzeroberflächen, begrenzte Unterstützung für nicht-Pythoncode
React <sup>8</sup>	JavaScript-Bibliothek zur Erstellung von Benutzeroberflächen	hohe Leistung und Reaktivität	erfordert zusätzliche Bibliotheken für die Entwicklung vollständiger Anwendungen

Tabelle 5.1: Auswahlkriterien Dashboardarchitektur

Der grundlegende programmatische Aufbau ist in Abbildung 5.3 ersichtlich. Hierbei stellen Rechtecke mit abgerundeten Ecken Komponenten innerhalb der Architektur dar und die Ellipsen stehen für die verwendeten Bibliotheken. Im Dashboard dient die Datei *App.js* als Einstiegspunkt der Anwendung. Hier wird die Navigationsleiste (*NavBar.jsx*) erstellt, die es Benutzer:innen ermöglicht, zwischen verschiedenen Wissensgraphen zu wählen. Je nach Auswahl wird eine entsprechende Benutzeroberfläche gerendert. Um einen globalen Zustand in React zu verwalten, der in allen Komponenten verfügbar ist, wird normalerweise die Weitergabe von Daten über Props verwendet. Die Verwendung der *state-pool*-Bibliothek<sup>9</sup> ermöglicht jedoch die Verwaltung eines globalen Zustands, der Informationen über den ausgewählten Wis-

---

<sup>3</sup><https://redash.io/>

<sup>4</sup><https://panel.holoviz.org/>

<sup>5</sup><https://dash.plotly.com/>

<sup>6</sup><https://shiny.posit.co/>

<sup>7</sup><https://voila.readthedocs.io/en/stable/using.html>

<sup>8</sup><https://react.dev/>

<sup>9</sup><https://github.com/yezyilomo/state-pool>

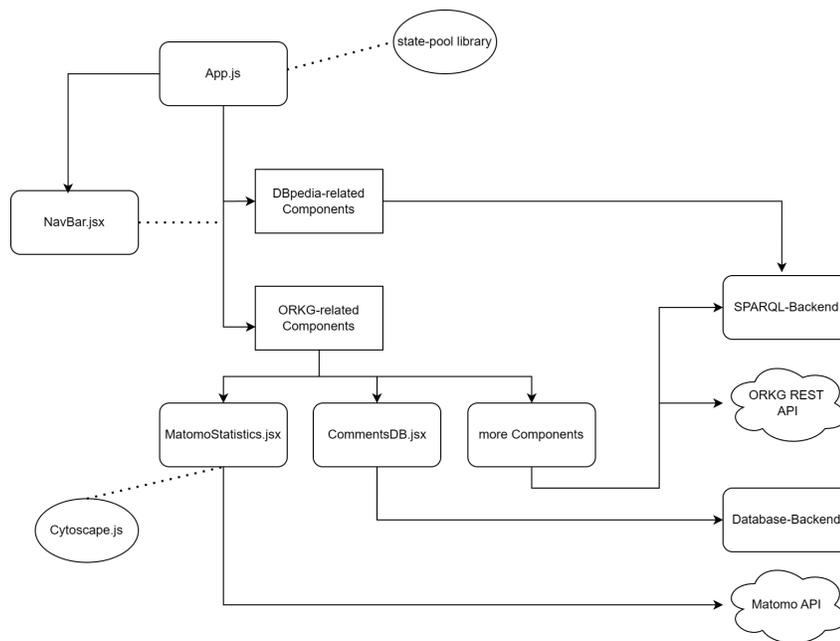


Abbildung 5.3: Dashboard Architektur

sensgraphen enthält. Dies legt die Grundlage dafür, die Komponenten modular zu gestalten, sodass nur der Endpunkt geändert werden muss.

Es hat sich jedoch herausgestellt, dass die Modularisierung der SPARQL-Abfragen eine Herausforderung darstellt, da die Wissensgraphen unterschiedliche Konzepte enthalten, die nicht einfach aufeinander abgebildet werden können. Dies bedeutet, dass die Abfragen für jeden Wissensgraphen individuell angepasst werden müssen, was die Modularisierung erschwert.

Die *state-pool*-Bibliothek bietet zwar eine Möglichkeit zur Verwaltung des globalen Zustands, aber die Komplexität der unterschiedlichen Wissensgraphen erfordert immer noch eine spezifische Anpassung der Abfragen. Dies kann die Modularität der Anwendung beeinträchtigen, da nicht alle Teile des Codes einfach austauschbar sind. Es ist daher notwendig, spezifische Lösungen für jeden Wissensgraphen zu implementieren.

Im Projekt gibt es zwei separate Backends: eines für die Verwaltung von SPARQL-Abfragen an die Wissensgraphen-Endpunkte und ein weiteres für eine interne Kommentarfunktionalität. Diese Aufteilung erleichtert die Fehlersuche und macht die Anwendung wiederverwendbarer. Für die Kommentarfunktion wurde ein Backend erstellt, das Kommentare über ORKG-Ressourcen in einer Datenbank verwaltet. Da-

durch gibt es eine gemeinsame Datenbank für alle Nutzer:innen, die nicht bei jedem Zugriff neu erstellt werden muss.

Eine weitere Komponente des Dashboards verwaltet die Besucherstatistiken und greift dazu auf die Matomo-API<sup>10</sup> zu, um Daten im JSON-Format zu erhalten. Zur Visualisierung wird die *Cytoscape*-Bibliothek<sup>11</sup> für JavaScript verwendet. Diese ermöglicht die komplexe Visualisierung eines Netzwerkgraphen, die stark angepasst werden kann.

Viele Komponenten, die mit dem ORKG in Verbindung stehen, greifen sowohl auf den SPARQL-Endpunkt als auch auf die REST-API zu. Die Dokumentation dazu findet sich auf der Website des ORKG<sup>12</sup>.

Die grafische Oberfläche wurde mithilfe von Bootstrap<sup>13</sup> für React umgesetzt. Dieses Frontend-Framework stellt vorgefertigte Komponenten wie Listen, Navigationsleisten, Buttons oder Overlays bereit, die nur noch mit CSS und den erforderlichen Daten angepasst werden müssen. Dies erleichtert die Entwicklung und sorgt für eine konsistente Benutzererfahrung.

Der implementierte Code ist auf Zenodo veröffentlicht [27].

## 5.3 Ausführung

Um das Dashboard lokal zu starten, ist es erforderlich, dass *Node.js*<sup>14</sup> und *Python*<sup>15</sup> auf dem System installiert sind. Diese Programme werden benötigt, um alle erforderlichen Module aus dem Projekt zu installieren. Durch Ausführen des Befehls `npm install` werden die erforderlichen Node-Module installiert. Zusätzlich wird der Befehl `pip install -r requirements.txt` innerhalb der beiden Backend-Ordner verwendet, um die erforderlichen Python-Module herunterzuladen.

Die Ausführung der beiden Backends wurde in einer Python-Datei zusammengefasst. Daher muss zunächst die Datei `executeBackends.py` gestartet werden, um die Backends zu initialisieren. Sobald die Backends gestartet sind, kann das Frontend mithilfe des Befehls `npm start` gestartet werden.

Zum Zeitpunkt der Abgabe wird aktiv daran gearbeitet, das Dashboard mithilfe von *Docker*<sup>16</sup> zu einer containerisierten Anwendung zu deployen. Die Verwendung

---

<sup>10</sup><https://developer.matomo.org/api-reference>

<sup>11</sup><https://js.cytoscape.org/>

<sup>12</sup><https://orkg.org/data>

<sup>13</sup><https://react-bootstrap.netlify.app/>

<sup>14</sup><https://nodejs.org/en>

<sup>15</sup><https://www.python.org/>

<sup>16</sup><https://www.docker.com/>

von *Docker* ermöglicht es, das Dashboard in einem isolierten Container auszuführen. Dies verbessert die Portabilität und Konsistenz der Anwendung erheblich, da alle erforderlichen Abhängigkeiten und Konfigurationen im Container gebündelt sind. Dadurch wird sichergestellt, dass das Dashboard unabhängig von der Umgebung und den Systemkonfigurationen konsistent funktioniert und kann damit gehostet werden, sodass es über eine URL verfügbar ist.

# Kapitel 6

## Durchführung der Aktionsforschung

Im Kontext des gesamten Forschungsvorgehens (vgl. Abbildung 4.1) bestehend aus Design Science und Aktionsforschung wird im nachstehenden Kapitel auf die Umsetzung der Aktionsforschung eingegangen.

Die entwickelte Anwendung im Rahmen dieser Masterarbeit hat den Anspruch, den Bedarfen des Kuratierungs- und Gemeinschaftsbildungsteams bestmöglich entgegen zu kommen und im Zusammenhang mit den technischen Möglichkeiten umzusetzen. Durch den intensiven Austausch mit den Partner:innen ermöglicht die Aktionsforschung eine gezielte Fokussierung auf die wesentlichen Merkmale. Dadurch wird ein Minimal Viable Product (MVP) entwickelt, deren Umsetzung bereits einen klaren Mehrwert für die zukünftigen Nutzer:innen bietet. Die Definition eines MVP kann gemäß Lanarduzzi und Taibi [32] auf verschiedene Arten erfolgen. Wir orientieren uns an folgender Definition:

### Minimal Viable Product

*MVP is a set of "minimal requirements", which meet the needs of the core group of early adopters or users.*

Die folgenden Unterkapitel erläutern explizit die zwei umgesetzten Zyklen (vgl. Abbildung 6.1) der Aktionsforschung und setzen damit die Grundlage für die abschließende Evaluation des entwickelten Dashboards. Screenshots des entstandenen Prototypen befinden sich im Anhang A. Die folgenden Abschnitte werden diese referenzieren.

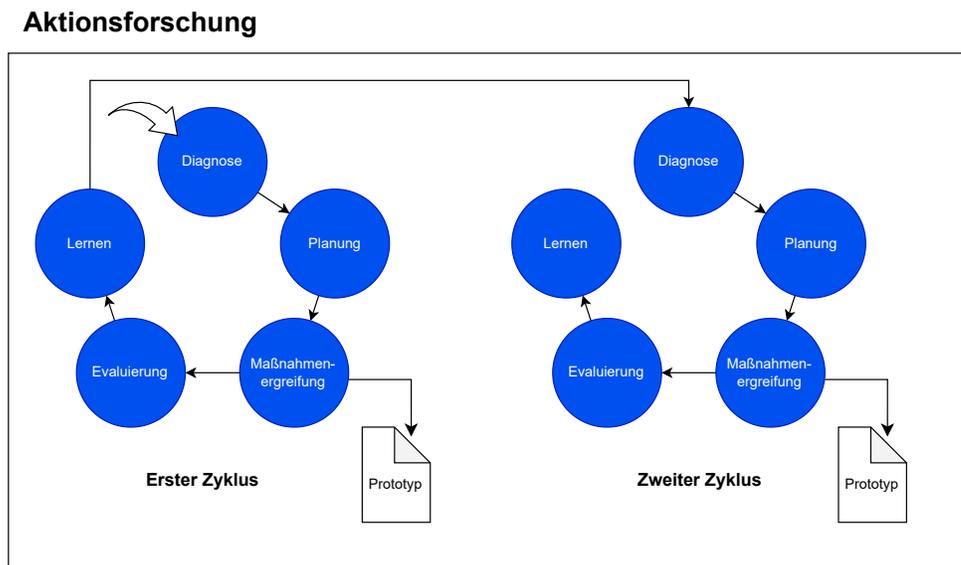


Abbildung 6.1: Umsetzung der Aktionsforschung

## 6.1 User Stories

Die vorliegende Arbeit basiert auf User Stories, die vorab von Mitgliedern des Kuratierungs- und Gemeinschaftsbildungsteam erstellt wurden. User Stories verorten sich im Requirements Engineering und repräsentieren kurz, einfach und prägnant die Anforderung der Kund:innen [6]. Dabei ist zu betonen, dass diese User Stories aus Sicht der zukünftigen Nutzenden formuliert werden und somit nicht nur ein zu entwickelndes Feature beschreiben, sondern den tatsächlichen Wert aus Sicht der Nutzer:innen repräsentieren [39]. Diese nutzerzentrierte Herangehensweise spiegelt sich in der üblichen Struktur wider: *As a <role>, I want <goal>, [so that <benefit>]* (Als <Rolle>, möchte ich <Ziel>, [sodass <Nutzen>]) [34]. Laut Mike Cohn [8] sind User Stories insofern von Use Cases abzugrenzen, da sie unvollständiger sind und geeigneter für eine iterative Softwareentwicklung. Weiterhin betont Cohn, dass User Stories durch ihr kurzlebiges Konzept selber iteriert werden und nicht als permanente Artefakte intendiert sind. Dieser Umstand akzeleriert nach Cohn die verbale Kommunikation zwischen Nutzer:innen und Entwickler:innen um zügige Feedbackschleifen zu bieten und führt zu einem besseres Verständnis der Anforderungen.

Die User Stories des Teams dienten als Ausgangspunkt für diese Arbeit und wurden im Verlauf iterativ angepasst. Die adressierten und iterierten User Stories im Rahmen der Dashboardentwicklung finden sich in Tabelle 6.1 und Tabelle 6.2 wie-

der. Die in den bereitgestellten User Stories identifizierten Themen spiegeln die Herausforderungen wider, die mit der Kuratierung von Wissensgraphen, insbesondere der des ORKG, einher gehen. Ein wesentlicher Fokus liegt auf der Entwicklung von Ansichten, die vorhandene Daten so aufbereiten, dass sie für das Kuratierungs- und Gemeinschaftsbildungsteam besser zugänglich sind. Diese Daten sind momentan zwar verfügbar, befinden sich jedoch nicht in einer geeigneten Form, um effektiv genutzt zu werden. Durch diese Ansichten können sowohl qualitativ hochwertige als auch qualitativ minderwertige Komponenten innerhalb des Wissensgraphen registriert werden. Des Weiteren konzentriert sich die Arbeit des Teams auf die Identifizierung von Mustern im Modellierverhalten der Nutzer:innen aus denen Schlussfolgerungen für die Verbesserung der Benutzeroberfläche abgeleitet werden können. Hierbei ist es wichtig, die gewonnenen Informationen zu nutzen, um Nutzende in bestimmten Aufgabenstellungen zu unterstützen und den ORKG damit bereinigt zu halten und eine Wissensbereicherung zu intensivieren (vgl. Kapitel Kuratierung 2.2).

Eine User Story besteht aus verschiedenen Elementen, die zuvor erwähnt wurden. Zunächst wird die Rolle beschrieben, aus deren Perspektive sie zu betrachten ist. Darüber hinaus wird der Artefakttyp des ORKG angegeben, auf den sich die Story bezieht. Weiterhin wird festgehalten, welche Beobachtung darüber gemacht werden kann und welches Ziel damit verfolgt wird um eine Änderung/Nutzen hervor zu rufen. Zudem hat das Team jeder Story eine Priorität zugewiesen. Die Rollen betreffen entweder Kurator:innen, allgemeine Benutzer:innen des ORKG oder beide. Vorrangig wird das Dashboard für Mitglieder des Kuratierungs- und Gemeinschaftsbildungsteams entwickelt. Die Perspektive, die Anwendung für alle Nutzer:innen des ORKG zur Verfügung zu stellen, beeinflusst in Teilen die Implementierung, insbesondere in Bezug auf die Verwendung personenbezogener Daten.

ID	Role	ORKG Artefact type	Observation	Goal	Benefit	Priority
1	all	Predicate	#Predicates without a description	identify missing descriptions	add missing descriptions	high
2	all	Class	#Classes without a description	identify missing descriptions	add missing descriptions	high
3	User	Predicate	duplicate Predicates	identify duplicates	merge the duplicates	medium
4	User	Resource	unused Resources	report the Resource	Resource gets deleted	high
5	User	Resource	unlabeled Resources	report the Resource	Resource gets deleted	high
6	Curator	Visitor	map of visitor paths	identify visit scenarios		medium
7*	all	Research field	#Papers in a research field	list research fields and the number of Papers in each of it	answer requests by user on how many Papers are assigned to a specific research field	low
8	all	Paper	#Statements per Paper	list the #Statements per Paper	identify patterns of modeling	medium

Tabelle 6.1: User Stories

\* Dieser User Story habe ich in Abstimmung mit dem Team eigenständig hinzugefügt.

ID	Role	ORKG Artefact type	Observation	Goal	Benefit	Priority
9	User	Comparison	#empty cells in a Comparison	get ratio of #empty cells over #all cells in a Comparison	model data for human and machine actionability	medium
10	all	Template	overview of Template creation	list creators of Templates and the number of Templates	identify users that frequently create Templates	low
11	User	Statement	I don't agree with a Statement and want to indicate that without changing it	make qualified comment of type "accuracy questioned"	original creator of Statement can identify all their tagged Statements	medium
12	User	Paper	I see a Paper representation that does not cover all information that I think is relevant	make qualified comment of type "lacking completeness"	information becomes available in the UI	low
13	User	Paper, Comparison	I see a Paper or Comparison that I don't understand due to bad modeling	make qualified comment of type "bad modeling"	information becomes available in the UI	low

Tabelle 6.2: User Stories

Die implementierten User Stories sind in der nachfolgenden Tabelle 6.3 nach Zyklen aufgliedert und den Dashboardkategorien zugeordnet, die zuvor in den Grundlagen (vgl. Kapitel 2.4) vorgestellt wurden. Wenn hier beide Zyklen angegeben sind, bezieht sich das auf eine zunächst grundlegende Implementierung von Komponenten im ersten Zyklus, die nach der ersten Evaluation und Feedback weiter in die Tiefe im zweiten Zyklus entwickelt wurden.

Die Einteilung der Stories in die einzelnen Dashboardkategorien legt eine Überlapung der Typen offen. Am meisten vertreten sind strategische und operative Komponenten. Das entwickelte Dashboard präsentiert demnach vorrangig Metriken als Momentaufnahme und integriert Echtzeitdaten, die mithilfe von Kontextinformationen eingeordnet werden können. Lediglich die Darstellung der Besucherpfade setzt auf eine komplexe Visualisierung um eine tiefgehende Analyse zu ermöglichen und somit in die Dashboardkategorie analytisch eingeordnet zu werden. Dieser Umstand spiegelt sich im Folgenden ebenso in der Implementierung wieder, da das analytische Fragment eine weitaus anspruchsvollere Entwicklung fordert.

ID	User Story	Zyklus	Dashboardtyp
1	#Predicates without a description	1	strategisch
2	#Classes without a description	1	strategisch
3	duplicate Predicates	1 und 2	strategisch/operativ
4	unused Resources	1	strategisch
5	unlabeled Resources	1	strategisch
6	map of visitor paths	2	analytisch
7	#Papers in a research field	2	operativ
8	#Statements per Paper	2	operativ
9	#empty cells in a Comparison	1 und 2	operativ
10	overview of Template creation	1 und 2	operativ
11, 12, 13	make comments about resources	2	operativ

Tabelle 6.3: Implementierte User Stories nach Zyklen und Dashboardkategorien

## 6.2 Erster Zyklus

Im ersten Zyklus habe ich mich sowohl mit der (technischen) Struktur des ORKG als auch mit den Grundlagen von React beschäftigt. Hierbei haben sich zwei zentrale Möglichkeiten ergeben, Daten aus dem ORKG abzufragen: über den SPARQL-Endpoint und als HTTP-Anfrage über die REST-API. Hierbei ist anzumerken, dass die technische Struktur des ORKG in Kapitel 5.1 detailliert erläutert wird und der programmatische Aufbau des Dashboards in Kapitel 5.2. Die Visualisierung in Abbildung 6.2 zeigt den ersten Zyklus, der mit den bereitgestellten User Stories des Teams beginnt und Teile davon in einem Prototypen umsetzt. Bei der Evaluierung wird das Team und der Moderator anwesend sein. Dabei entsteht eine Videoaufzeichnung um das Feedback detaillierter heraus zu arbeiten und daraus konkrete Erkenntnisse zu gewinnen. Diese Erkenntnisse dienen danach als Eingabe für den zweiten Zyklus.

Erster Zyklus

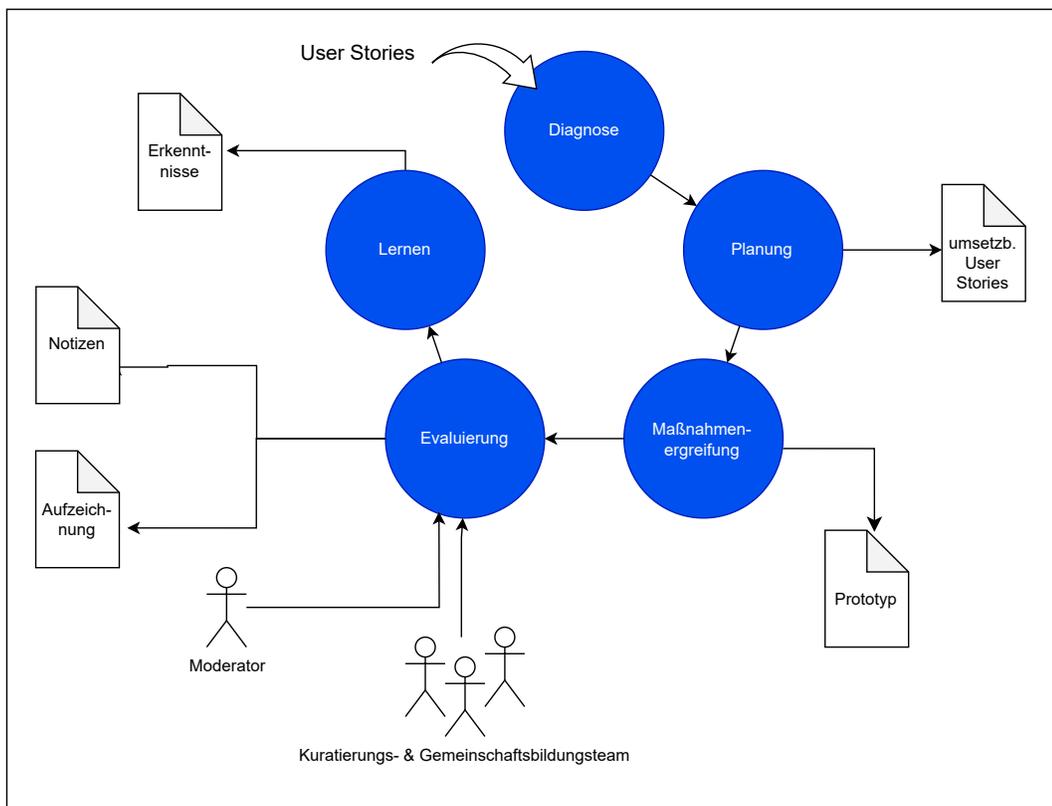


Abbildung 6.2: Erster Zyklus der Aktionsforschung

### 6.2.1 Diagnose

Während der Diagnosephase wurden die User Stories als Ausgangspunkt für die problemorientierte Identifizierung von Herausforderungen verwendet. Auf dieser Grundlage erfolgte zunächst eine Gruppierung der User Stories nach ihrer zugewiesenen Priorität und ihrer inhaltlichen Ähnlichkeit. Dadurch ergaben sich rasch übergeordnete Themen, anhand derer die Stories anschließend strukturiert werden konnten. Diese Anordnung ist zudem durch die Klassifizierung in die Dashboardkategorien geprägt. Somit ergeben sich fünf Reiter für die Benutzeroberfläche: Overview, Visitors, Insights, Content, Actions. Die folgende Tabelle 6.4 zeigt die Aufteilung, die Zahlen stellen die IDs der Stories dar. Der Overview-Reiter beinhaltet alle strategischen User Stories und dient somit als allgemeiner Überblick. Die Darstellung der Besucherpfade findet sich im Visitors-Reiter wieder. Der Insights-Reiter dient zur Übersicht aktueller Statistiken. Im Content-Reiter befinden sich die übrigen operativen User Stories bezüglich Comparisons und Templates. Abschließend bietet der Actions-Reiter die Option, Kommentare über ORKG Ressourcen zu erstellen und zu listen.

Overview	Visitors	Insights	Content	Actions
1, 2, 3, 4, 5	6	7, 8	9, 10	11, 12, 13

Tabelle 6.4: User Stories gruppiert in Reiter

Anhand dieser Gruppierung wurden die Stories zunächst mit den technischen Möglichkeiten abgestimmt um Entwicklungsschritte zu planen und zu priorisieren.

### 6.2.2 Planung

Für diesen Entwicklungszyklus werden mehrere User Stories adressiert, die sich auf die Funktionen des Overview- und Content-Reiters beziehen. Diese umfassen die Stories 1, 2, 3, 4, 5, 9 und 10. Insbesondere die User Stories des Overview-Reiters werden bevorzugt behandelt, da sie hauptsächlich mit SPARQL-Anfragen umgesetzt werden können und eine hohe Priorität haben. Die Stories 9 und 10 betreffen die Comparisons und die Templates. Die ursprüngliche User Story für die Templates steht in Tabelle 6.5.

Im ersten Zyklus wird zunächst von der Version der User Story ausgegangen, die einen Überblick darüber bietet, wie häufig Autor:innen ihre individuell erstellten Templates verwenden und wie oft diese von anderen Nutzer:innen des ORKG genutzt werden.

ID	Role	ORKG Artefact type	Observation	Action	Change	Priority
10	Curator	Template	distribution of Template usage (by author and others)	sort Templates based on usage	identify frequently used Templates	high

Tabelle 6.5: Initiale User Story für die Templates

Viele User Stories haben gemeinsam, dass sie in Listen dargestellt werden. Bei einem Klick auf ein Element, eine Ressource, wird eine Weiterleitung zu ihrem entsprechenden ORKG-Eintrag ermöglicht. Die Daten innerhalb der Listen, sowie alle Daten im Dashboard insgesamt, werden beim Starten und Neuladen der Anwendung vom ORKG-Backend in Echtzeit angefragt und aktualisiert. Um die Listen übersichtlich zu halten, sind Sortier- und Filterfunktionen vorgesehen, um die Ergebnisse zu strukturieren.

Ein weiterer wichtiger Aspekt, der im Hinblick auf die langfristige Entwicklung von Bedeutung ist, besteht darin, die Implementierung von Anfang an möglichst modular zu gestalten. Dadurch soll es beispielsweise möglich sein, Daten von anderen Wissensgraphen wie DBpedia abzufragen, um deren Metriken vergleichen zu können.

### 6.2.3 Maßnahmenereifung

Die Implementierung des Dashboards habe ich in Frontend und Backend aufgeteilt. Das Frontend wird in React programmiert und ermöglicht den komponentenbasierten Ansatz, bei dem einzelne Komponenten unterschiedlich voneinander entwickelt und getestet werden können. Im Frontend werden alle relevanten Teile für die Benutzeroberfläche programmiert. Das Backend besteht aus einem Python-Flask-Server, der dazu dient, Anfragen aus dem Frontend entgegen zu nehmen und an den SPARQL-Endpunkt des entsprechenden Wissensgraphen weiterzuleiten.

Zunächst wurden die Stories 1 und 3, die sich auf Predicates beziehen, zu einer Komponente zusammengeführt. Die Daten wurden vom SPARQL-Endpunkt abgefragt. Dadurch entstand eine Liste von Predicates ohne Beschreibungen sowie eine Liste von Duplikaten, die nach der Anzahl der Duplikate pro Predicate sortiert werden kann (vgl. Screenshot A.2). Diese Duplikate basieren auf einem String-Matching des Labels. Ähnlich wurden die Stories 1 und 5, die ORKG Classes und Resources betreffen, ebenfalls in einer Liste umgesetzt. Für Story 4, die ungenutzten ORKG-

Ressourcen, wurden lediglich das Verhältnis und die absolute Anzahl festgehalten, um einen Überblick zu gewinnen, da es sich um eine große Menge von Ressourcen handelt. Der Overview-Reiter ist in Screenshot A.1 dargestellt. Schnell wurde ersichtlich, dass die Daten aus dem SPARQL-Endpunkt nicht alle Details enthalten. Insbesondere betrifft das die Provenance-Informationen, welche nicht in das SPARQL-Image übernommen werden. So ist es beispielsweise momentan nicht möglich, die Ressourcen nach dem Datum ihrer Erstellung zu sortieren.

Im Content-Reiter wurden die Stories 9 und 10 durch HTTP-Anfragen an das ORKG-Backend umgesetzt (vgl. Screenshot A.9). Für die Comparisons wurden zunächst alle IDs über den SPARQL-Endpunkt erfragt, und anschließend wurde über die Simcomp-API jede Comparison über ihre ID abgerufen. Das erhaltene Objekt wurde dann programmatisch nach der Anzahl der Gesamtzellen und leeren Zellen in einer Comparison gezählt. Beide Komponenten wurden als Tabelle umgesetzt, die Sortierfunktionen für jede Spalte bietet.

Bei der Komponente für die Templates, stellte sich heraus, dass es nicht trivial war, herauszufinden, wie oft ein Template verwendet wird. Dafür mussten zunächst alle Artefakte des ORKG durchlaufen werden, um zu zählen, wie oft jedes Template verwendet wird. Zusätzlich gab es in dieser Iteration keinen speziellen Template-Endpunkt, wodurch zunächst über alle Statements nach Ressourcen gefiltert werden musste, die eine Klasse eines speziellen Typs enthalten. Anschließend konnte durch eine erneute Abfrage ermittelt werden, wie oft dieses Template in der Objektposition eines Statements auftritt, folglich, wie oft es verwendet wird. Erst im Verlauf der Arbeit wurde vom Entwicklungsteam der spezielle Endpunkt bereitgestellt, sodass nur noch eine Anfrage dafür nötig war.

### 6.2.4 Evaluierung

Bei der Evaluation des Dashboards mit dem Team wurde zunächst eine Bildschirmfreigabe genutzt, um das Dashboard in Form eines Pluralistic Walkthroughs vorzustellen. Im Anschluss daran gaben die Anwesenden Rückmeldungen zur Verbesserung. Eine der Rückmeldungen betraf die Duplikate der Predicates, die nach dem Vorhandensein einer Beschreibung filterbar sein sollten. Es wurde angemerkt, dass Predicates möglicherweise dasselbe Label haben können, aber durch eine Beschreibung voneinander unterschieden werden, da sie verschiedene Konzepte beschreiben könnten. Deswegen würden nur Duplikate zusammengeführt werden, die zusätzlich keine Beschreibung enthalten.

Des Weiteren wünschte sich das Team zusätzliche Spalten für die Comparison-Tabelle, um zu sehen, wie viele Predicates und wie viele Contributions eine Compa-

rierson enthält. Dies würde es erleichtern, Auffälligkeiten in den Daten auf einen Blick zu erkennen.

Es entstand eine Diskussion darüber, in welche Richtung sich das Dashboard in Zukunft entwickeln könnte. Es wurde beschlossen, dass das Dashboard vorerst keine Daten im ORKG selbst verändern soll und sich auf die Visualisierung von Daten und die Offenlegung problematischer Ressourcen sowie die Qualität des Wissensgraphen des ORKG konzentrieren soll. Die Möglichkeit einer automatischen Zusammenführung (merging) von Duplikaten könnte in Betracht gezogen werden, nachdem das Dashboard etabliert ist und sichergestellt werden kann, was genau durch den Eingriff der Maschine geschieht.

Dennoch betonte das Team den Wunsch nach mehr Interaktion mit dem Dashboard und unterstrich, dass die Möglichkeit, Kommentare zu Ressourcen abzugeben, ein nächster Schritt in der Implementierung sein sollte.

### **6.2.5 Lernen**

Die Evaluation verdeutlichte, dass die Kuratierung ein komplexer Prozess ist, der nicht allein aufgrund einer simplen statistischen Zahl erfolgen kann, sondern verschiedene Aspekte berücksichtigen muss. Gegenwärtig ist die Automatisierung dieser Aufgabe zu anspruchsvoll. Das Dashboard soll die Kuratierung optimal unterstützen und die Arbeit erleichtern, um damit die Nutzer:innen des ORKG effektiver zu unterstützen.

Ein grundlegender Überblick über die Qualität des ORKG wurde implementiert und soll nun durch die Möglichkeit, auffällige Ressourcen durch Kommentare festzuhalten, erweitert werden. Die Dokumentation von Artefakten, die Verbesserungen erfordern oder positiv modelliert wurden, ist entscheidend, um die Qualität zu gewährleisten.

Aufgrund des stetig wachsenden Wissens im ORKG stellt die menschliche Kuratierung eine zeit- und ressourcenaufwendige Aufgabe dar, die nun die Möglichkeiten der Unterstützung durch das Dashboard erkunden will. Damit werden die Grenzen der maschinellen Unterstützung erweitert, ohne dass dabei direkt in den Wissensgraphen eingegriffen wird. Insbesondere die gewünschten Möglichkeiten zum Handeln im Dashboard sollen dies ermöglichen.

## 6.3 Zweiter Zyklus

Im zweiten Zyklus der Aktionsforschung liegt der Fokus darauf, die Erkenntnisse aus der ersten Iteration zu nutzen und die Implementierung des Dashboards weiter voranzutreiben. Während zuvor grundlegende Funktionen umgesetzt wurden, werden jetzt bestehende Funktionen verbessert und tiefere Konzepte wie Visualisierungsmöglichkeiten der Besucherstatistiken und maßgeschneiderte Komponenten entwickelt. Ziel ist es, das Dashboard zu einem echten Mehrwert für das Kuratierungs- und Gemeinschaftsbildungsteam zu machen, indem individuell auf die Bedarfe eingegangen wird. Die Abbildung 6.3 veranschaulicht den Ablauf des zweiten Zyklus. Dabei werden die Erkenntnisse aus dem vorherigen Zyklus als Eingabe genutzt, um einen verbesserten Prototypen zu entwickeln. Am Ende dieses Zyklus werden abschließende Erkenntnisse gewonnen, die zur Entwicklung eines finalen MVPs dienen.

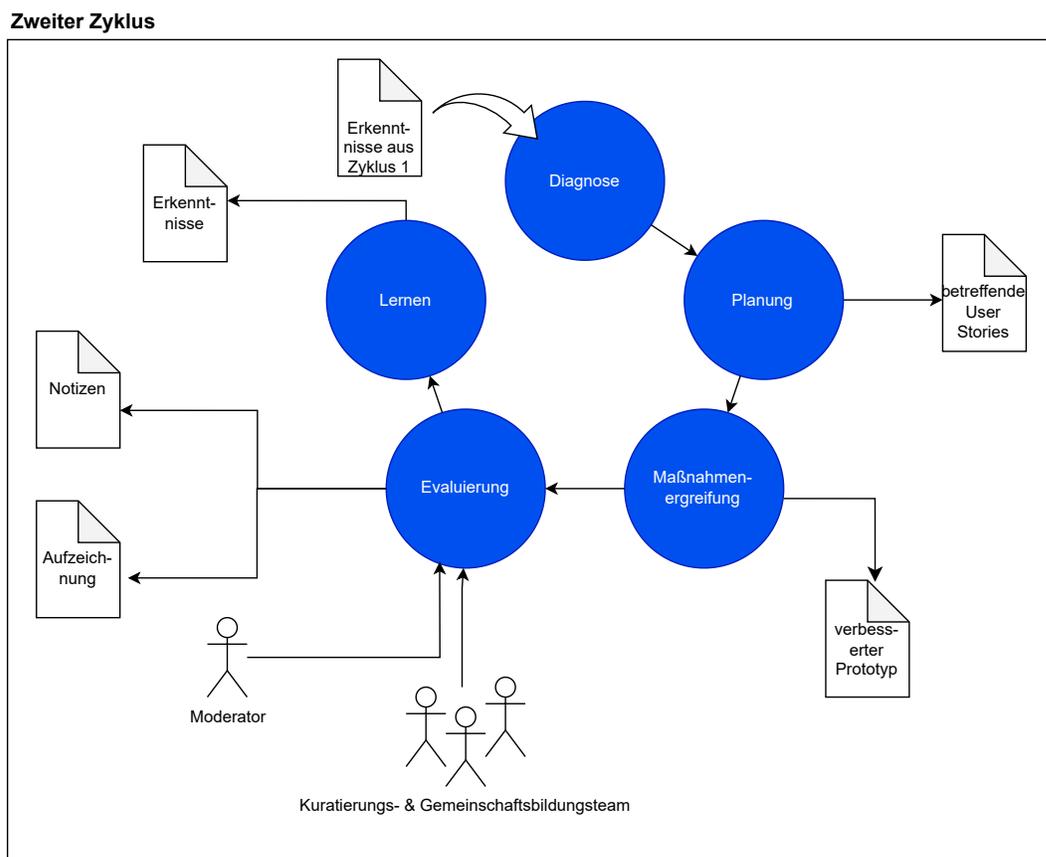


Abbildung 6.3: Zweiter Zyklus der Aktionsforschung

### 6.3.1 Diagnose

In der Evaluation wurde deutlich, dass das Dashboard mehr Interaktionsmöglichkeiten bieten sollte, die über die einfache Verknüpfung und Öffnung von Ressourcen hinausgehen. Sowohl die Kuratierung als auch die Gemeinschaftsbildung sind daran interessiert, die Herausforderungen der Nutzenden des ORKG zu identifizieren, die Benutzeroberfläche zu verbessern und die Nutzenden damit besser bei bestimmten Aufgaben zu unterstützen. Um diese Szenarien zu bestimmen, liegt der Fokus in dieser Iteration auf der Analyse der Besucherstatistiken.

Aufgrund der Anfrage, zusätzliche Statistiken für die Comparisons bereitzustellen - die Anzahl der Predicates und der Contributions - hat sich eine weitere wichtige Metrik herauskristallisiert. Diese Metrik ist für die Bewertung der Qualität des ORKG von zentraler Bedeutung, da sie Informationen über die Anzahl der Statements pro Paper liefert. Comparisons werden durch Contributions erstellt, diese Contributions basieren auf Papern. Demnach können die Statements pro Paper als fundamentale Einheit des ORKG gesehen werden. Bisher ist diese Metrik nicht direkt über die Benutzeroberfläche des ORKG ersichtlich. Daher ist die Story 8 in dieser Iteration besonders relevant.

### 6.3.2 Planung

In dieser Iteration werden die Stories 6, 7, 8, 11, 12 und 13 umgesetzt. Die Stories 3, 9 und 10 werden ebenfalls in dieser Phase detailliert implementiert. Besonderes Augenmerk liegt auf der Umsetzung von Story 6, der Visualisierung der Besucherpfade, die mithilfe eines Netzwerkgraphen realisiert werden soll.

Des Weiteren wird die Funktion zum Abgeben von Kommentaren zu Ressourcen integriert. Diese Kommentare sind ausschließlich im Dashboard sichtbar und haben keine Auswirkungen auf die Inhalte des ORKG.

Die Umsetzung der Story 8, die die Anzahl der Statements pro Paper betrifft, erfordert eine eigenständige Programmierung eines Endpunkts, da diese Daten derzeit nicht direkt vom Backend abrufbar sind.

### 6.3.3 Maßnahmenereifung

**Predicates und Comparison.** Die grundlegenden Implementierungen umfassen die Bereitstellung eines Toggle für die Duplikate der Predicates (vgl. Screenshot A.2), um festzulegen, ob sie eine Beschreibung enthalten oder nicht, sowie das Hinzufügen zusätzlicher Spalten zur der Comparison-Tabelle (vgl. Screenshot A.9).

**Research fields.** Die Erfassung der Anzahl der Paper pro Forschungsbereich (Story 7) erfolgt durch eine Anfrage an den SPARQL-Endpunkt. Die erhaltenen Daten werden anschließend in einem Balkendiagramm visualisiert, das es ermöglicht, sowohl nach der Anzahl der Paper als auch nach dem Label des Forschungsbereichs zu filtern (vgl. Screenshot A.6). Die Ansicht passt sich dabei dynamisch an, um eine benutzerfreundliche Interaktion zu gewährleisten.

**Statements per Paper.** Um die Anzahl der Statements pro Paper zu ermitteln, habe ich im ORKG-Backend einen spezifischen Endpunkt entwickelt. Dieser Endpunkt ermöglicht die Generierung eines Pageable-Objekts, das sämtliche Paper mit ihrer jeweiligen Anzahl an Statements (inklusive der Metdaten) zurückgibt. Dabei werden für jede Contribution die Statements gezählt und summiert. Diese Information lässt sich in der ORKG Benutzeroberfläche durch die Graph-Ansicht eines Papers nachvollziehen. Die so gewonnenen Daten dienen der Erstellung einer Komponente, die verschiedene Funktionen bereitstellt: sie erstellt einen Boxplot zur Visualisierung der Verteilung (vgl. Screenshot A.9), präsentiert die Top 10 Paper mit den meisten Statements, die Bottom 10 mit den wenigsten Statements (vgl. Screenshot A.7) und bietet zudem eine filterbare Liste zur weiteren Analyse (vgl. Screenshots A.8). Zum Zeitpunkt der Evaluation war meine merge request noch nicht bearbeitet, somit enthielt diese Komponente zunächst Dummy-Daten. Beim Experiment wurden die korrekten Paper aus dem ORKG gelistet.

**Templates.** Durch die Einführung des neuen Template-Endpunkts ist es nun weiterhin nicht möglich festzustellen, welches Template von wem und wie oft verwendet wird. Daher wurde die Komponente für die Template-Statistiken modifiziert, um trotz dieser Einschränkung relevante Daten zu erfassen. Im Rahmen der verfügbaren technischen Mittel wurde diese zu einer Tabelle umgestaltet, die aufführt, welche Nutzer:innen wie viele Templates erstellt haben. Durch Auswahl einer bestimmten Zeile ist es möglich, die genaueren Details der erstellten Templates einzusehen und diese nach ihrem Erstellungsdatum zu sortieren (vgl. Screenshot A.10). Dies ermöglicht beispielsweise die Identifizierung von Nutzer:innen, die kontinuierlich zum Inhalt des ORKG beitragen. Somit ist diese Komponente die einzige, die auf ORKG Nutzer:innen zugreift.

**Matomo Besucherstatistiken.** Eine der aufwändigeren Implementierungen in diesem Zyklus betraf die Umsetzung von Story 6, die Visualisierung von Besucherpfeilen. Zunächst wurden die Besucherstatistiken über einen Matomo API-Key abgefragt. Diese Statistiken können für einen bestimmten Zeitraum abgerufen werden und

liefern für jede:n Besucher:in ein JSON-Objekt mit den besuchten URLs sowie der Verweildauer auf diesen Seiten. Im Rahmen dieser Daten besteht keine Möglichkeit, zwischen einfachen Besucher:innen und eingeloggten Nutzer:innen zu unterscheiden.

Ursprünglich war geplant, die Besucherpfade mittels eines Sankey-Diagramms (vgl. Abbildung 6.4) darzustellen. Jedoch stellte sich schnell heraus, dass dieses Diagramm nicht in der Lage ist, zirkuläre Pfade anzuzeigen und eher für lineare Abläufe geeignet ist. Es kommt häufig vor, dass Besucher:innen während ihres Aufenthalts auf dem ORKG Seiten mehrfach besuchen, was die Darstellung komplexer gestaltet. Aufgrund dessen habe ich mich für einen Netzwerkgraphen (vgl. Abbildung 6.5) entschieden, da dieser flexibler ist und den Vorteil bietet, dass er komplexe Beziehungen zwischen den Elementen visualisieren kann, indem er Knoten (ORKG Seiten) und (gerichtete) Kanten (Besucherpfade) darstellt (vgl. Screenshot A.3). Für die Darstellung des Diagramms habe ich die Cytoscape-Bibliothek für JavaScript verwendet<sup>1</sup>.

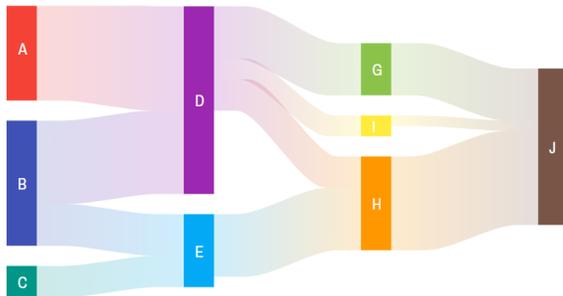


Abbildung 6.4: Sankey-Diagramm<sup>2</sup>

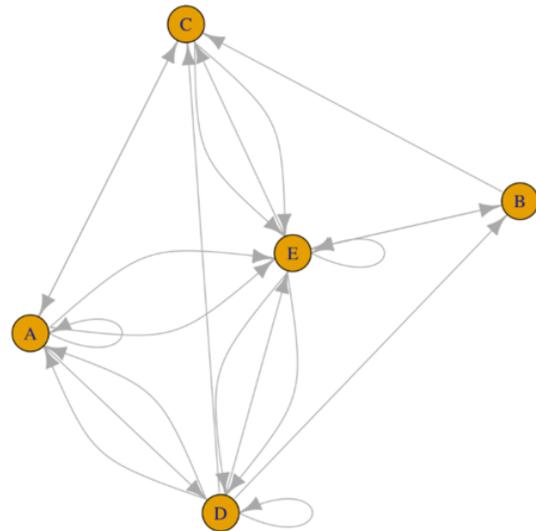


Abbildung 6.5: Netzwerkgraph<sup>3</sup>

Zusätzlich ergab sich das Problem der hohen Einzigartigkeit der URLs, was zu einer starken Heterogenität der Daten führte und die übersichtliche Einbettung in den Netzwerkgraphen erschwerte. Um dieses Problem anzugehen, wurden die besuchten

<sup>1</sup><https://js.cytoscape.org/>

<sup>3</sup><https://www.amcharts.com/demos-v4/sankey-diagram-v4/>

<sup>3</sup><https://r-graph-gallery.com/network.html>

Seiten nach ihren Sub-URLs aufgeteilt. Dies bedeutet, dass die URLs auf ein gemeinsames Muster reduziert wurden, zum Beispiel wird aus `https://orkg.org/paper/R690304` der Knoten *paper*. Dieser Ansatz ermöglichte eine deutlich bessere Strukturierung der Daten und eine übersichtlichere Darstellung im Netzwerkgraphen.

Diese Modifikation der URLs trägt auch zum Datenschutz bei, da konkrete Besucherpfade nicht mehr rückverfolgt werden können. Da noch nicht feststeht, ob das Dashboard auch für allgemeine Nutzer:innen des ORKG zur Verfügung gestellt werden soll, wurde zunächst diese Implementierung gewählt. Zusätzlich hat diese Modifikation den Effekt, dass viele Besucher:innen ähnliche Pfade nehmen werden, zum Beispiel den Pfad *paper->comparison->paper*. Infolgedessen habe ich beschlossen, die Kanten zwischen den Knoten zusammenzufassen und sie farblich nach ihrer Häufigkeit zu markieren. Auf diese Weise wird schnell deutlich, welche Übergänge von den Besucher:innen am häufigsten gewählt werden. Der Netzwerkgraph liefert somit einen allgemeinen Überblick über das Besucherverhalten auf den Seiten des ORKG, wobei einzelne Besucherpfade nicht individuell verfolgt werden können, da die Daten aggregiert werden.

Darüber hinaus bietet der Graph die Möglichkeit, die dargestellte Zeitspanne anzupassen, aus der die Statistiken abgeleitet werden sollen. Außerdem kann zwischen verschiedenen Darstellungen für das Graphlayout gewählt werden, beispielsweise *breadthfirst*, *circle* oder *grid*. Um den Graphen weiter zu analysieren, steht eine Filteroption für die Kanten zur Verfügung. Diese ermöglicht es, Kanten auszublenden, die nicht mindestens einen bestimmten Schwellenwert erreichen. Durch Hovern über die Legende (Farben der Kanten) können zeitweise nur Kanten einer bestimmten Häufigkeit angezeigt werden.

Einzelne Knoten können durch Klick inspiziert werden, wobei zwei Tabellen angezeigt werden: eine für die ausgehenden und eine für die eingehenden Kanten sowie Informationen darüber, zu welchem Knoten sie führen. Darüber hinaus gibt es die Möglichkeit, externe Knoten auszublenden, wie zum Beispiel DOI-Seiten oder ein *git-Repository*, über die Besucher den ORKG verlassen.

**Kommentarfunktion.** Um die Interaktivität im Dashboard weiter zu verbessern und die Stories 11, 12 und 13 zu bearbeiten, habe ich ein zweites Backend hinzugefügt, das eine Datenbank verwaltet. Diese Datenbank enthält Kommentare zu Ressourcen. Dadurch ist es möglich, in einer bestimmten Komponente ein Paper oder eine Comparison auszuwählen, indem die entsprechende ID übergeben wird. Momentan ist es im ORKG nicht möglich, ein spezielles Statement über eine ID zu identifizieren, diese ist nur im Backend ersichtlich. Anschließend kann der:die Nutzer:in aus sechs vorgegebenen Kommentartypen wählen, wie beispielsweise *accuracy*

*questioned*, und eine individuelle Beschreibung hinzufügen. Hier kann beispielsweise das betreffende Statement spezifiziert werden (vgl. Screenshot A.12). Wenn die ID korrekt ist, wird der Titel der Ressource vom ORKG abgefragt und die eingetragenen Kommentare werden in einer Tabelle dargestellt (vgl. Screenshot A.11), so dass Kurator:innen diese Angelegenheiten bearbeiten und lösen können. Über einen Link kann die Ressource im ORKG geöffnet werden. Gleichzeitig können diese Ressourcen auch als positive Beispiele durch einen Kommentartypen wie *reproducible results* aufgelistet werden. Sobald ein Kommentar über eine Ressource hinzugefügt wurde, ist es möglich die Beschreibung im Nachhinein zu aktualisieren. Zudem werden Kommentare zur selben Ressource in derselben Tabellenzeile gelistet. Durch die Verwendung einer lokalen Datenbank bleiben die Daten beim Neuladen der Anwendung erhalten und werden zwischen allen Nutzer:innen des Dashboards geteilt.

**DBpedia.** In dieser Iteration des Dashboards habe ich die Modularität erweitert, indem ich zusätzliche Daten aus DBpedias SPARQL-Endpunkt abgefragt habe (vgl. Screenshot A.13). Dabei ging es speziell um die Anzahl der Prädikate und Klassen ohne Beschreibung. Diese Implementierung stellt eine Grundlage dar und bietet einen ersten Einblick, wie das entstehende Dashboard auf andere Wissensgraphen angewendet werden kann und Metriken zum Vergleich anbieten soll.

### 6.3.4 Evaluierung

In der Evaluation mit dem Team, einleitend mit dem Pluralistic Walkthrough, wurde besonders viel Zeit für das Feedback zum Netzwerkgraphen der Besucherstatistik aufgewendet. Die Leiterin der Gemeinschaftsbildung betonte, dass diese Übersicht eine Bereicherung für ihre Arbeit darstellt. Dadurch kann sie häufig besuchte Übergänge identifizieren und Schlussfolgerungen über das Besucherverhalten ziehen. Allerdings bemängelten die Kurator:innen, dass ihnen Informationen darüber fehlen, was die Besucher:innen auf der Seite tatsächlich tun, ob und welche Inhalte sie hinzufügen. Da diese Daten jedoch nicht von Matomo abrufbar sind, besteht hier noch Potenzial, den ORKG selbst zu erweitern, um diese Informationen bereitzustellen.

Des Weiteren wurde angemerkt, dass im Netzwerkgraphen aufgrund der aggregierten Darstellung nicht ersichtlich ist, welchen Pfad ein:e spezifische:r Besucher:in nimmt. Es wurde festgehalten, dass die tatsächlichen Pfade interessant wären, um später eine Art Muster für Benutzerprofile zu entwickeln, beispielsweise wie sich neue User im Vergleich zu langfristigen User verhalten. Allerdings würde dies zusätzlich die Einschränkung mit sich bringen, dass identifiziert werden kann, welche Besucher:innen als Benutzer:innen eingeloggt sind.

Für das Team war es weiterhin wichtig, dass die Kommentarfunktion nicht nur für Paper und Comparisons verfügbar ist, sondern für jegliche Inhalte im ORKG. Damit wird sichergestellt, dass die Funktionalität flexibel und umfassend genutzt werden kann, unabhängig von der Art des Inhalts, auf den sich die Kommentare beziehen.

### 6.3.5 Lernen

Bei der finalen Implementierung sollte besonders die Skalierbarkeit im Blick behalten werden. Es ist wichtig zu berücksichtigen, wie die bereitgestellten Daten tatsächlich in die bestehende Arbeit des Teams integriert werden können. Zudem sollte eine zukünftige Vision des Dashboards die Einbindung von Workflows beinhalten, die zumindest Teile der Kuratierung automatisieren können.

Eine genauere Analyse des Besucherverhaltens erfordert eine Abwägung zwischen Datenschutz und den Anforderungen an die Datenvisualisierung. Es gilt zu entscheiden, wie viele Informationen über Besucher:innen (öffentlich) zugänglich sein dürfen. Eine detailliertere Darstellung kann jedoch auch den Trade-off mit sich bringen, dass die Daten unübersichtlicher werden. Hier ist Fingerspitzengefühl gefragt, um die Benutzeroberfläche nicht zu überladen, während gleichzeitig sinnvolle Informationen für die Nutzenden bereitgestellt werden. Eine gewisse Eingewöhnungszeit in eine neue Oberfläche ist normal, sollte aber nicht zu Frustration oder einer schlechten Nutzererfahrung führen. Die Wirksamkeit des Dashboards bei der Bewältigung von Kuratierungsaufgaben und der Unterstützung der Gemeinschaftsbildung wird im Rahmen des abschließenden Experiments mit zukünftigen Endnutzenden untersucht.

Darüber hinaus wurde deutlich, dass die technische Struktur des ORKG möglicherweise angepasst werden muss, um den Anforderungen des Kuratierungs- und Gemeinschaftsbildungsteams gerecht zu werden und die gewünschten Funktionen in Zukunft umsetzen zu können.

## 6.4 MVP

Mit dem Abschluss der Lernphase endet der zweite Zyklus der Aktionsforschung. Um der Definition des Minimal Viable Products (MVP) gerecht zu werden, wurden die wichtigsten und einfachsten Änderungen im Dashboard umgesetzt. Der resultierende Prototyp dient als Basis für das folgende Experiment zur Bewertung der Benutzerfreundlichkeit des Dashboards und zur Überprüfung, inwieweit die Bedürfnisse des Teams darin umgesetzt wurden.

Die bedeutendsten Änderungen betreffen die Komponenten der Besucherstatistiken und der Kommentare. Für die Besucherstatistiken wurde ein separater Reiter

hinzugefügt, der konkrete Besucherpfade auflistet (vgl. Screenshot A.4). Dadurch können wir sicher sein, dass Besucher tatsächlich diese Pfade genommen haben. Im Gegensatz zum aggregierten Verhalten aller Besucher im Graphen liefert diese Liste spezifische Pfade, wie beispielsweise *search- >comparison- >resource- >paper*, und zeigt an, wie oft dieser Pfad genutzt wurde. Diese Informationen beziehen sich auf die angegebene Zeitspanne und können nach der Anzahl der Vorkommen und der Länge des Pfades gefiltert werden. Durch Klick auf einen Pfad wird dieser im Netzwerkgraphen hervorgehoben (vgl. Screenshot A.5), was es ermöglicht, den Pfad im Kontext der gesamten Besucherstatistik zu betrachten.

Darüber hinaus wurde die Kommentarfunktion so angepasst, dass nun Kommentare über jegliche Inhalte des ORKG möglich sind. Es ist nun nicht mehr erforderlich, nur die ID anzugeben, sondern einfach eine ORKG-URL einzufügen. Falls ein Endpunkt für diese Ressource existiert, wird der Titel und Typ automatisch gesetzt, ansonsten wird einfach die URL angezeigt (vgl. Screenshot A.12).

Im Anhang A finden sich alle Screenshots für den entstandenen MVP.



# Kapitel 7

## Experiment

Der Kern dieser Arbeit liegt in der Entwicklung einer maßgeschneiderten Anwendung für das Kuratierungs- und Gemeinschaftsbildungsteams des ORKG. Diese Anwendung wird in Form eines Dashboards als MVP (Minimal Viable Product) (vgl. Definition 6.1) umgesetzt.

Ziel ist es, vorhandene Daten des ORKG so aufzubereiten, dass sie in Bezug auf Visualisierung und Funktionalität den wichtigsten Bedürfnissen des Teams gerecht werden. Basierend auf diesem Kontext wird für die abschließende Evaluierung des Dashboards unter Verwendung der Zielvorlage von Wohlin et. al [57] folgendes Forschungsziel gesetzt:

### Forschungsziel Experiment

***Analysiere** das entwickelte Dashboard  
**mit dem Zweck** der Unterstützung des Prozesses der Kuratierung und Gemeinschaftsbildung von Forschungswissensgraphen  
**im Hinblick auf** Erfüllung von spezifizierten Bedürfnissen bezüglich Bedienbarkeit  
**aus der Perspektive des** ORKG-Teams, insbesondere des Teams „Kuratierung und Gemeinschaftsbildung“  
**im Kontext vom** Open Research Knowledge Graph.*

In diesem Zusammenhang wird Bedienbarkeit nach Dutsinma et. al anhand des ISO 9421-Standards definiert [12]:

### Usability

*Usability is defined as the degree to which a program may be utilized to achieve measurable objectives with effectiveness, efficiency and satisfaction in a specific context of usage.*

Bedienbarkeit wird demnach nicht als eindimensionale Metrik aufgefasst, sondern versteht sich als eine Kombination aus den Faktoren Effektivität, Effizienz und Zufriedenheit.

Um den Umfang des Experiments weiter zu definieren, wird die GQM (Goal Question Metric)-Methode von Wohlin et. al [57] angewandt. Im Kontext der Aktionsforschung, deren Ziel es ist, Forschungsergebnisse anwendungsorientiert zu gestalten, stellt sich die Frage, inwiefern das Dashboard den Bedarfen des Teams gerecht wird. Folglich ergibt sich aus dem Forschungsziel die folgende Forschungsfrage:

### Forschungsfrage

*Wie gut unterstützt das entwickelte Dashboard das Team des ORKG, insbesondere das Team „Kuratierung und Gemeinschaftsbildung“, hinsichtlich der spezifizierten Bedürfnisse bezüglich Bedienbarkeit im Prozess der Kuratierung und Gemeinschaftsbildung des ORKG?*

Um diese Forschungsfrage beantworten zu können, wird im Folgenden auf die Hypothesen, Variablen und den detaillierte Aufbau des Experiments eingegangen.

## 7.1 Hypothesen

In Bezug auf das vorangehende Forschungsziel und Forschungsfrage, wird die Bedienbarkeit des Dashboards anhand von Effizienz, Effektivität und Zufriedenheit evaluiert. Für diese drei Aspekte werden jeweils Null- und Alternativhypothesen aufgestellt. Die Effizienz differenziert hier zwischen den Aufgaben und die Effektivität betrachtet die Ergebnisse in aggregierter Form über alle Aufgaben.

**Effizienz:** Die erste Hypothese untersucht die Effizienz. Die Effizienz wird durch die objektive und subjektive Bewertung mittels zweier Metriken erfasst. Die objektive Effizienz wird anhand von benötigter Zeit der Teilnehmenden für die Vollendung einer Aufgabe gemessen. Für die objektive Effizienz wird ein erwarteter Mittelwert

von fünf Minuten pro Aufgabe gesetzt. Dieser setzt sich aus der Vertrautmachung mit der Benutzeroberfläche als auch der tatsächlichen Durchführung der Aufgabe zusammen. Das Dashboard soll leichtgewichtig sein und die Nutzer:innen nicht übermäßig lange beanspruchen. Angesichts der Komplexität der Aufgaben und der vermuteten Fähigkeiten der Teilnehmenden erscheint eine Zeitspanne von fünf Minuten als angemessen. Diese Zielsetzung soll sowohl erreichbar sein als auch die Effizienz der Teilnehmenden erfassbar machen.

- ( $H_{0,Effizienz}$ ): Die gemessene Zeit, die ein:e Teilnehmende:r für den Abschluss der Aufgabe [1, 2.1, 2.2, 3, 4] benötigt, übersteigt oder beträgt exakt 5 Minuten.
- ( $H_{1,Effizienz}$ ): Die gemessene Zeit, die ein:e Teilnehmende:r für den Abschluss der Aufgabe [1, 2.1, 2.2, 3, 4] benötigt, beträgt weniger als 5 Minuten.

Die subjektive Effizienz wird von den Teilnehmenden nach Beendigung der Aufgabe abgefragt und umfasst eine fünfstufige Likert-Skala, die die Zustimmung mit der Aussage „I finished the task fast“ auf Werte von 1 (Strongly disagree) bis 5 (Strongly agree) abbildet. Diese Abfrage wird herangezogen, um zu sehen, ob die gemessene Zeit mit der Wahrnehmung der Teilnehmenden übereinstimmt.

**Effektivität:** Die zweite Hypothese untersucht die Effektivität, mit der die Aufgaben mithilfe des Dashboards abgeschlossen werden. Nach Frøkjær et. al [19] kann Effektivität in Hinblick auf Benutzbarkeit als die Genauigkeit und Vollständigkeit interpretiert werden, mit der Nutzer:innen ein Ziel erreichen. Indikatoren der Effektivität beinhalten demnach die Qualität der Lösung und Fehlerraten. Die Effektivität wird ebenfalls anhand von objektiver und subjektiver Metriken untersucht. Die objektive Effektivität beschreibt, ob die Teilnehmenden die Aufgaben erfolgreich abgeschlossen haben. Hierbei liegt der Fokus auf dem tatsächlich erreichten Ziel und nicht, wie erfolgreich Teilnehmende dieses erreicht haben, da es keinen universellen Lösungsweg für die Aufgaben gibt. Wenn die Mehrheit der Teilnehmenden nicht in der Lage ist, mindestens drei Aufgaben erfolgreich zu bewältigen, könnte das darauf hinweisen, dass das Dashboard nicht in der Lage ist, die benötigte Unterstützung für eine effektive Nutzung zu bieten.

- ( $H_{0,Effektivität}$ ): Die Teilnehmenden schließen maximal 3 der 5 Aufgaben erfolgreich ab.
- ( $H_{1,Effektivität}$ ): Die Teilnehmenden schließen mehr als 3 der 5 Aufgaben erfolgreich ab.

Die subjektive Effektivität wird ebenfalls von den Teilnehmenden nach Abschluss einer Aufgabe abgefragt und umfasst eine fünfstufige Likert-Skala, die die Zustimmung mit der Aussage „I finished the task successfully“ auf Werte von 1 (Strongly disagree) bis 5 (Strongly agree) abbildet. Auch hier wird die Wahrnehmung der Teilnehmenden mit den tatsächlich erfolgreich abgeschlossenen Aufgaben abgeglichen.

**Zufriedenheit:** Die dritte Hypothese untersucht die Zufriedenheit der Teilnehmenden mit dem Dashboard. Um Aussagen über die Zufriedenheit zu treffen, wird der verkürzte User Experience Questionnaire (UEQ) einmal nach Abschluss aller Aufgaben verwendet. Ein Durchschnittswert von  $> 0.8$  kann laut Auswertungsverfahren des UEQ als positive Nutzererfahrung interpretiert werden.

- $(H_{0,Zufriedenheit})$ : Die Teilnehmenden haben keine positive Nutzererfahrung mit dem Dashboard (UEQ  $\leq 0,8$ ).
- $(H_{1,Zufriedenheit})$ : Die Teilnehmenden haben eine positive Nutzererfahrung mit dem Dashboard (UEQ  $> 0,8$ ).

**Bedarfe des Teams:** Die Teilnehmenden geben nach Abschluss der Aufgaben an, inwieweit sie den spezifizierten Bedarfen des Kuratierungs- und Gemeinschaftsbildungsteams zustimmen, die im Dashboard umgesetzt wurden. Hierbei werden fünf Bedarfe aufgelistet, zu denen die Teilnehmenden eine Aussage auf einer fünfstufigen Likert-Skala treffen können.

## 7.2 Abhängige und unabhängige Variablen

Basierend auf den vorgestellten Metriken und Hypothesen, ergeben sich folgende unabhängige und abhängige Variablen, dargestellt in Tabelle 7.1. Die unabhängige Variable ist das bereitgestellte Dashboard. Zu messen ist, wie gut das Dashboard die spezifizierten Bedürfnisse bezüglich Bedienbarkeit unterstützt. Zuvor haben wir Benutzerfreundlichkeit definiert, die aus den Elementen Effektivität, Effizienz und Zufriedenheit besteht. Darüber hinaus ist die Erfüllung der spezifizierten Bedarfe innerhalb des Dashboards eine subjektive Beurteilung und wird im Folgenden ebenfalls berücksichtigt. Mithilfe der GQM-Methode [57] werden sowohl objektive als auch subjektive Metriken aus der Forschungsfrage abgeleitet, um diese quantitativ zu beantworten.

Im Experiment werden verschiedene abhängige Variablen gemessen werden, um den potenziellen Einfluss des Dashboards zu untersuchen. Durch die Kombination

Variable		Erklärung	
unabhängig	Dashboard	Das entwickelte Dashboard wird bereitgestellt.	
abhängig	Effizienz	objektiv	Die gemessene Zeit, die Teilnehmende für die Durchführung einer Aufgabe benötigen.
		subjektiv	Die Zustimmung der Teilnehmenden nach der Bearbeitung jeder Aufgabe mit der Aussage „I finished the task fast“.
abhängig	Effektivität	objektiv	Ein boolescher Wert (1 oder 0), der angibt, ob die Teilnehmenden das Ziel der Aufgabe erreicht haben.
		subjektiv	Die Zustimmung der Teilnehmenden nach der Bearbeitung jeder Aufgabe mit der Aussage „I finished the task successfully“.
abhängig	Zufriedenheit	subjektiv	Der Wert des UEQ nach Bearbeitung aller Aufgaben.
abhängig	erfüllte Bedarfe	subjektiv	Die Zustimmung der Teilnehmenden nach Bearbeitung aller Aufgaben bezüglich der Umsetzung von fünf ausgewählten Bedarfen des Teams.

Tabelle 7.1: Abhängige und unabhängige Variablen mit ihrer Erklärung

von objektiven und subjektiven Messungen wird ein umfassendes Bild davon erzeugt, wie effektiv das Dashboard in Bezug auf die Erfüllung der Bedarfe ist und welche Bereiche verbessert werden müssen.

### 7.3 Material

Für die Durchführung des Experiments wird den Teilnehmenden ein Laptop mit dem entwickelten Dashboard und eine Maus bereitgestellt. Zudem wird das Audio mit einem separaten Mikrophon aufgezeichnet, die eingebaute Webcam ist abgeklebt. Zusätzlich erhalten die Teilnehmenden ein Dokument, welches einen Fragebogen zu demografischen Daten, eine Einverständniserklärung, die Aufgabenstellungen und einen abschließenden Bewertungsbogen enthält.

Das verbleibende Kapitel behandelt das begleitende Dokument des Experiments (vgl. Anhang B). Das Experiment ist in drei Abschnitte unterteilt: die Erfassung demografischer Daten jedes/jeder Teilnehmenden, die Bearbeitung von vier Aufgaben

unter Verwendung des Dashboards und eine abschließende subjektive Bewertung der Benutzererfahrung sowie der Zustimmung zur Erfüllung der Bedarfe des Teams. Das Dokument ist in englischer Sprache verfasst, da Englisch die vorherrschende Sprache im gesamten ORKG-Team ist. Dies gewährleistet, dass alle Teammitglieder die Möglichkeit haben, an dem Experiment teilzunehmen.

Die demografischen Angaben umfassen den Namen, das Alter, die berufliche Tätigkeit und die Dauer der Tätigkeit beim ORKG. Zudem wird erfragt, welchem Team der oder die Teilnehmende angehört (Kuratierungs- & Gemeinschaftsbildungsteam, Generelles ORKG Team oder Extern). In diesem Zusammenhang steht ebenfalls die Angabe, an wievielen der zwei vorangegangenen Evaluationen des Dashboards der oder die Teilnehmende anwesend war um später eventuelle Rückschlüsse über eine mögliche Vertrautheit der Anwendung ziehen zu können. Bei der Erhebung der persönlichen Daten wird der Datenschutz der Teilnehmenden gewährleistet, sie haben jederzeit die Möglichkeit, ihre Teilnahme abzubrechen oder zurück zu ziehen und können wählen, Angaben unbeantwortet zu lassen. Alle Daten werden anonymisiert verarbeitet. Zudem ist die Zustimmung für die Aufnahme von Audio und Video erfragt um eine detailliertere Analyse zu ermöglichen.

Der Durchführungsteil besteht aus fünf Aufgaben, welche die Teilnehmenden mit dem Dashboard ausführen. Die Aufgaben beziehen sich auf unterschiedliche Reiter im Dashboard. Jede Aufgabe wird eingeleitet, aus welcher Perspektive sie zu betrachten ist (Kurator:in oder Gemeinschaftsbildungsexpert:in) und welchen Zweck sie verfolgt. Genauer, kann eine Aufgabe als eine Handlung betrachtet werden, die erforderlich ist, um ein bestimmtes Ziel zu erreichen, wie von Mori et al. [37] beschrieben. Darüber hinaus wird betont, dass ein Modell für die Aufgabenentwicklung erstellt werden muss, damit das Benutzer-Aufgabenmodell (wie die Benutzer:innen glauben, dass die Handlungen durchgeführt werden sollen) mit dem System-Aufgabenmodell (wie die Anwendung annimmt, dass die Handlungen durchgeführt werden sollen) übereinstimmt. Anhand möglicher Diskrepanzen in diesen Modellen können Verbesserungen für die Benutzeroberfläche abgeleitet werden. Aufgrund dessen werden die Teilnehmenden animiert ihre Entscheidungen und Gedanken laut zu äußern (Think-Aloud-Methode).

Nach der Durchführung jeder Aufgabe geben die Teilnehmenden ihre Einschätzung zu ihrer Effizienz und Effektivität ab. Hierbei reicht die ordinale Skala von *Strongly disagree* über *Disagree*, *Neutral* und *Agree* bis hin zu *Strongly agree*. Die Entscheidung, eine ungerade Anzahl von Antwortmöglichkeiten zu wählen, beruht darauf, die Option einer neutralen Antwort anzubieten, selbst wenn diese fälschlicherweise als Nichtbeantwortung interpretiert werden könnte [40]. Um dieses potenzielle Szenario zu vermeiden, wird auf die Think-Aloud-Methode verwiesen, auf die im Ablauf des

Experiments genauer eingegangen wird. Bei der Reihenfolge der Aufgaben wird bewusst auf Randomisierung verzichtet, da die Durchführung der Aufgaben auf einem Lerneffekt basiert und teilweise Ergebnisse aus vorangegangenen Aufgaben wiederverwendet werden.

Zum Abschluss erfolgt eine Bewertung der Erfahrung der Teilnehmenden mit dem Dashboard. Zur Beurteilung der Zufriedenheit wird der gekürzte User Experience Questionnaire (UEQ) verwendet, der eine effiziente Einschätzung der Nutzererfahrung durch Endnutzer:innen ermöglicht [47]. Laut Schrepp et. al [47] wird die verkürzte Version des UEQ in einem experimentellen Kontext angewendet, wobei der Fragebogen auf acht Paare gegensätzlicher Begriffe (z. B. conventional und inventive) beschränkt ist, deren Zustimmung mittels einer siebenstufigen Skala ausgedrückt werden kann. Der verkürzte UEQ konzentriert sich auf vier Aspekte der pragmatischen Qualität (Interaktionsqualität bezogen auf das Erreichen von Zielen) sowie auf vier Aspekte der hedonischen Qualität (Aspekte in Verbindung mit dem Vergnügen bei der Nutzung):

obstructive	<input type="radio"/>	supportive						
complicated	<input type="radio"/>	easy						
inefficient	<input type="radio"/>	efficient						
confusing	<input type="radio"/>	clear						
boring	<input type="radio"/>	exciting						
not interesting	<input type="radio"/>	interesting						
conventional	<input type="radio"/>	inventive						
usual	<input type="radio"/>	leading edge						

Für die Auswertung des UEQ werden die angegebenen Daten auf der siebenstufigen Skala auf Werte von -3 bis 3 abgebildet. Wie bereits erwähnt, kann ein Durchschnittswert von  $> 0.8$  als positive Evaluation interpretiert werden.

Abschließend wird von den Teilnehmenden die Zustimmung mit der Erfüllung von fünf Bedarfen des Kuratierungs- & Gemeinschaftsbildungsteams innerhalb des Dashboards abgefragt (vgl. Tabelle 7.2). Die Zustimmung wird hier anhand einer fünfstufigen Likert-Skala festgehalten.

Die Bedarfe B1 bis B4 wurden im Dashboard umgesetzt. B5 wurde ursprünglich vom Team mit hoher Priorität gefordert, konnte aufgrund fehlender Möglichkeit in der technischen Struktur des ORKG für die Differenzierung von internen und externen Nutzer:innen nicht umgesetzt werden. Hier dient B5 als Kontrollfrage, um

B1: The dashboard allows the tracking of visitor paths.

B2: The dashboard makes it possible to indicate if a resource is inaccurate or incomplete.

B3: The dashboard provides tools for monitoring the quality of the ORKG.

B4: The dashboard provides an overview of areas for improvement.

B5: The dashboard lists how active the internal (ORKG members) and external users are.

Tabelle 7.2: Ausgewählte Bedarfe des Teams

sicherzustellen, dass die Teilnehmenden aufmerksam waren und kein Muster beim Ankreuzen der Antworten gewählt haben. Kung et al. [31] betonen, dass die Einbeziehung einer solchen Kontrollfrage die Validität beeinflussen könnte, indem sie das Verhalten der Teilnehmenden bei den nachfolgenden Antworten beeinflusst. Aufgrunddessen wird die vorliegende Kontrollfrage zum Abschluss des Fragebogens verwendet.

## 7.4 Stichprobe

Ziel der Entwicklung des Dashboards ist es, ein MVP zu konzipieren, das im praktischen Rahmen des ORKG Kuratierungs- und Gemeinschaftsbildungsteams eingesetzt wird. Damit bestmöglich beurteilt werden kann, in welchem Ausmaß die zugeschnittene Anwendung verwendbar ist, werden Teilnehmende aus dem Umfeld des ORKG eingeladen teilzunehmen um eine gewisse Generalisierbarkeit zu gewährleisten.

Jedes Mitglied des ORKG-Teams hat die Rechte eines: einer Kurators:Kuratorin um den Wissensgraphen effizient zu pflegen und Nutzer:innen schnell zu unterstützen. Daher kann davon ausgegangen werden, dass alle Teilnehmende mit den Kuratierungs- und Gemeinschaftsbildungsprozessen vertraut sind, was die Validität des Experiments unterstützt. Es ist zu erwähnen, dass alle Mitglieder aus dem Kuratierungs- und Gemeinschaftsbildungsteams bei mindestens einer der zwei vorangegangenen Evaluationen anwesend waren. Dieser Umstand könnte auf eine gewisse Vertrautheit mit dem Dashboard dieser Mitglieder schließen und eine bessere Performance bei dem Experiment hervor rufen.

Die Stichprobenauswahl erfolgt durch *Convenience Sampling* im gesamten ORKG-Team, das aus 53 Personen besteht. Im gesamten Team des ORKG wurde für das Experiment geworben und um Teilnahme gebeten. Über ein Termintool wurden Interessierte gebeten, sich für das Experiment einzutragen. Dabei wurden 11 Teilnehmer:innen individuell anhand ihrer Verfügbarkeit ausgewählt, während weitere 4 Personen spontan an einem bestimmten Tag zur Teilnahme eingeladen wurden. Die Teilnahme am Experiment ist freiwillig.

Somit ergibt sich eine Stichprobe aus Teilnehmenden, die sich im wissenschaftlichen Umfeld befinden und vertraut mit dem ORKG sind. Insgesamt nahmen 15 Personen am Experiment teil, wobei sowohl Präsenz- als auch Fernzugriffsmöglichkeiten angeboten wurden, da das Team des ORKG teilweise verteilt über Deutschland arbeitet und flexible Home Office-Möglichkeiten hat. Von den Teilnehmenden nahmen 10 Personen persönlich und 5 Personen digital teil, um die Teilnahme zu erleichtern und sicherzustellen, dass eine vielfältige Gruppe von Personen vertreten ist.

Die Darstellung 7.1 veranschaulicht die Verteilung der Teilnahmen an den beiden vorangegangenen Evaluationen der Aktionsforschungszyklen, differenziert nach der Zugehörigkeit zu den Teams: dem Kuratierungs- und Gemeinschaftsteam (KuG-Team) und dem allgemeinen ORKG-Team. Die x-Achse repräsentiert die Anzahl der Teilnahmen, während die y-Achse die Anzahl der Personen darstellt, die jeweils an einer bestimmten Anzahl von vorangegangenen Evaluationen teilgenommen haben. Die Daten zeigen, dass die Mehrheit der Teilnehmenden an keiner Evaluation teilgenommen hat, wobei jedoch alle Mitglieder des KuG-Teams an mindestens einer Evaluation teilgenommen haben.

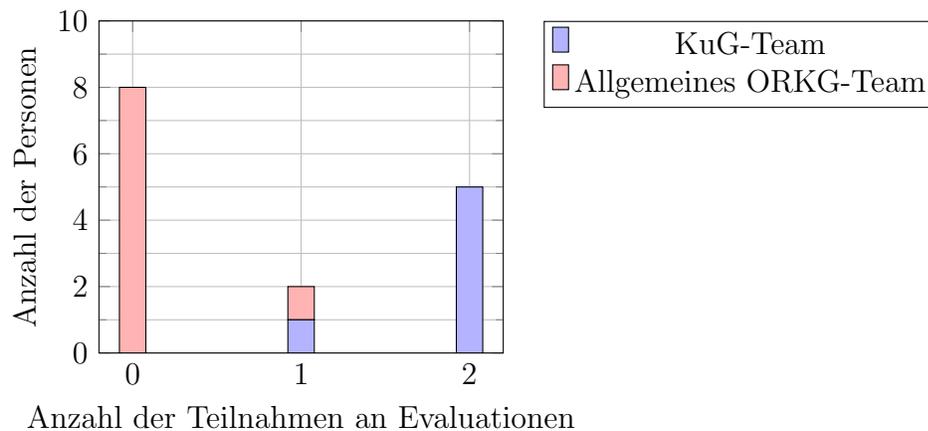


Abbildung 7.1: Verteilung der Anzahl der Teilnahmen an vorangegangenen Evaluationen

## 7.5 Ablauf des Experiments

Das Experiment wird sowohl in Präsenz als auch online durchgeführt, um allen Teilnehmenden gerecht zu werden, auch denen, die nicht persönlich anwesend sein können. Für die Online-Teilnahme wurde auf die Software TeamViewer und Skype zurückgegriffen. Mittels TeamViewer konnten die Teilnehmenden sich auf meinem Laptop einloggen und dort navigieren. Über Skype führten wir eine Videokonferenz durch, um miteinander zu kommunizieren.

Es startet mit einer Begrüßung und Einleitung, in der der Ablauf und Zweck des Experiments erläutert werden. Besonders betont wird dabei, dass es darum geht, zu untersuchen, wie gut das Dashboard dem Team dient und wie benutzerfreundlich es ist, insbesondere für die Durchführung verschiedener Aufgaben.

Anschließend erhalten die Teilnehmenden eine Einverständniserklärung aus dem beigelegten Dokument (vgl. Anhang B). Die Ausfüllung erfolgt bei den Online-Experimenten auf einem geteilten Bildschirm, auf dem die Teilnehmenden ihre Daten selbstständig eingeben können. Die Einverständniserklärung erfasst demografische Daten und bittet um Zustimmung zur anonymisierten Verwendung ihrer Daten sowie zur Aufzeichnung von Audio und Video während des Experiments. Die Aufzeichnung ermöglicht eine genauere Analyse im Anschluss und eine bessere Messung objektiver Metriken für Effizienz und Effektivität, die exakte Bearbeitungszeit und den erfolgreichen Abschluss einer Aufgabe.

Nach erfolgter Zustimmung beginnt der offizielle Teil, und die Aufzeichnung wird gestartet. Die Teilnehmenden werden ermutigt, ihre Gedanken während der Durchführung nach der Think-Aloud-Methode laut auszusprechen und jederzeit Fragen zu stellen, wenn etwas unklar ist. In *The think aloud method: a practical approach to modelling cognitive processes* von Someren et al. [51] wird deutlich, dass die Think-Aloud-Methode geeignet ist, um Einblicke in das menschliche Problemlösungsverhalten und kognitive Prozesse zu gewinnen. Die Autor:innen beschreiben ebenso, dass die Protokolle aus dieser Methode offenlegen, welche Schwierigkeiten verschiedene Personen mit Aufgaben haben und welche anderen Faktoren das Erreichen eines Ziels beeinflussen können. Darüber hinaus wird betont, dass die Think-Aloud-Methode den Denkprozess nicht stört, da das Sprechen automatisch nebenbei erfolgt.

Als nächstes erhalten die Teilnehmenden die Aufgabenblätter aus dem Begleitdokument. Sie können jederzeit Notizen auf dem Blatt machen. In Präsenz haben sie zusätzlich die Möglichkeit, gleichzeitig das Aufgabenblatt und das Dashboard zu betrachten. Online müssen sie jedoch zwischen den Anwendungen wechseln.

Das Experiment verwendet ein *within-subject*-Design, bei dem jede:r Teilnehmenden

de die definierten Aufgaben nacheinander mit dem Dashboard ausführt. Es gibt keine Kontrollgruppe oder eine andere unabhängige Variable zum Vergleich. Für die Dauer des Experiments sind 30 Minuten angesetzt.

Die Aufgaben werden schriftlich eingeführt. Zusätzlich habe ich für Aufgabe 2.1 eine mündliche Einführung in die grafische Oberfläche gegeben, um den Teilnehmenden zu helfen, den Netzwerkgraphen aus den Besucher:innenstatistiken zu verstehen. Wenn die Teilnehmenden der Ansicht sind, die Aufgabe abgeschlossen zu haben, bewerten sie ihre persönliche Einschätzung zur Geschwindigkeit und zum Erfolg des Aufgabenabschlusses auf zwei Likert-Skalen. In Online-Umgebungen wechseln sie erneut zur primären Anwendung, um die Skala auf dem Aufgabenblatt anzukreuzen. Der Ablauf, bei dem die Teilnehmenden die Aufgabenstellung lesen, die Aufgabe bearbeiten und anschließend ihre Einschätzung abgeben, wird für alle fünf Aufgaben durchgeführt.

Danach füllen die Teilnehmenden den UEQ in Bezug auf alle erledigten Aufgaben aus. Durch mein gezieltes Nachfragen beim Ausfüllen können Aspekte identifiziert werden, auf die sich die Teilnehmenden bei der Bewertung beziehen. Dadurch können auch extreme Antworten besser verstanden werden. Damit schließt der offizielle Teil des Experiments. Die Messungen werden hier beendet.

Abschließend haben die Teilnehmenden die Möglichkeit, sich das gesamte Dashboard anzusehen und Anmerkungen sowie Verbesserungsvorschläge zu machen. Diese sind ebenfalls relevant, um das Dashboard zukünftig weiterzuentwickeln und den Bedürfnissen der Zielnutzer:innen gerecht zu werden.

Das Experiment wurde Anfang April 2024 mit 15 Teilnehmenden durchgeführt.

## 7.6 Datenanalyse

Das Kapitel der Datenanalyse orientiert sich zunächst an den Faktoren aus der sich die Bedienbarkeit ableitet: Effizienz, Effektivität, Zufriedenheit. Zusätzlich wurde die Zustimmung von umgesetzten Bedarfen des Teams anhand einer ordinalen Likert-Skala gemessen. Bezüglich Effizienz und Effektivität liegen quantifizierbare Daten vor, die sowohl deskriptive als auch Inferenzstatistik ermöglichen. Hinsichtlich der Zufriedenheit wurde von allen Teilnehmenden der UEQ-Score erfasst. Die Bedarfe des Teams wurden ebenfalls anhand einer Likert-Skala gemessen. Die gesammelten Daten lassen sich somit in objektive und subjektive Kategorien unterteilen (vgl. Tabelle 7.1). Die Auswahl der statistischen Tests für die Daten hängt von der Normalverteilung der Daten ab, das kann mithilfe des *Shapiro-Wilk-Tests* [50] überprüft werden. Falls die Daten normalverteilt sind, wird der *Einstichproben-t-Test* [43] verwendet, andernfalls der *Einstichproben-Wilcoxon Signed Rank-Test* [42]. Die Hypothesentests

werden verwendet, um festzustellen, ob die gemessenen Indikatoren signifikant sind und ob das Dashboard die definierten Ziele erreicht hat.

Alle Rohdaten sind auf Zenodo [28] veröffentlicht.

### 7.6.1 Korrelation

Zunächst wird der Rangkorrelationskoeffizient nach Spearman [10] angewendet, um den Zusammenhang zwischen zwei Variablen zu bestimmen. Dabei wird untersucht, ob die Anzahl der teilgenommenen Evaluationen mit den zu untersuchenden Variablen der objektiven Metriken korreliert. Der Rangkoeffizient  $\rho$  liefert dabei eine Aussage darüber, wie stark diese Beziehung ist und in welche Richtung die Korrelation besteht. Der Rangkorrelationskoeffizient nimmt Werte zwischen -1 und 1 an. Ein positives Vorzeichen deutet darauf hin, dass eine Zunahme einer Variable mit einer Zunahme der anderen einhergeht, während ein negatives Vorzeichen auf eine Abnahme der einen Variable mit einer Zunahme der anderen hinweist. Mit anderen Worten, je höher der Koeffizient ist (positiv oder negativ), desto stärker ist die Beziehung zwischen den Variablen. Nach Cohen [7] stellen die Grenzen von 0,1 eine schwache Korrelation dar, 0,3 eine mittlere und 0,7 eine starke. Die Ergebnisse der untersuchten Zusammenhänge, die die Anzahl der teilgenommenen Evaluationen mit den Effizienzen der Aufgaben, der Effektivität und dem UEQ in Beziehung setzen, sind in Tabelle 7.3 dargestellt.

Korrelation	Spearman's $\rho$		
	Anzahl teilgenommener Evaluationen		
	$\rho$	$p$	signifikant?
Task 1 Effizienz	-0,349	0,202	×
Task 2.1 Effizienz	-0,089	0,752	×
Task 2.2 Effizienz	0,225	0,420	×
Task 3 Effizienz	-0,024	0,933	×
Task 4 Effizienz	-0,355	0,195	×
Effektivität	-0,0253	0,364	×
UEQ	0,427	0,112	×

Tabelle 7.3: Korrelation nach Spearman

Die Ergebnisse aller Untersuchungen zeigen keine Signifikanz, da für alle Resultate  $p > 0,05$  gilt. Dies legt nahe, dass kein ausreichend starker Zusammenhang zwischen einer erhöhten Anzahl an teilgenommenen Evaluationen (Mitglieder des

Kuratierungs- und Gemeinschaftsbildungsteam) und signifikant unterschiedlichen Ergebnissen im Experiment besteht. Die Werte in Bezug auf den UEQ stechen jedoch etwas heraus. Hier zeigt sich eine positive mittlere Korrelation, die darauf hinweist, dass eine Zunahme an teilgenommenen Evaluationen mit einer verbesserten Bewertung der Teilnehmenden im UEQ einhergeht. Jedoch ist zu beachten, dass diese Korrelation nicht als signifikant betrachtet werden kann.

Insgesamt legen die Befunde nahe, dass die beobachteten Unterschiede zwischen den Gruppen nicht deutlich genug sind, um eine klare Unterscheidung zwischen den Teams in der fortlaufenden Datenanalyse vorzunehmen.

### 7.6.2 Effizienz

Anhand der Video- und Audioaufzeichnungen der Experimente können die präzisen Zeiten für den Abschluss einer Aufgabe gemessen werden. Diese objektiven Daten basieren auf einer metrischen Verhältnisskala. Nach Beendigung jeder Aufgabe haben die Teilnehmenden ihre persönliche Einschätzung bezüglich der Geschwindigkeit der Aufgabenerledigung abgegeben. Diese Daten werden auf einer Ordinalskala festgehalten. Die vorliegende Grafik 7.2 veranschaulicht die Antworten in Bezug auf die Aussage „I finished the task fast“. Neben der Grafik stehen die jeweiligen Werte für den Median pro Aufgabe, diese liegen bei 4 und 5. Das weist darauf hin, dass die Mehrheit der Teilnehmenden der Aussage zustimmen, dass sie die Aufgaben schnell erledigt haben.

Für die erste Aufgabe bekundete die Mehrheit (8 Personen) ein neutrales bis leicht zustimmendes Empfinden bezüglich ihrer Schnelligkeit. Die restlichen Teilnehmenden gaben an, stark zuzustimmen in Bezug auf ihre Effizienz. Die zweite und dritte Aufgabe (2.1 und 2.2) zeigt eine breitere Streuung der Antworten. Die Mehrheit gibt an, die Aufgaben überwiegend schnell oder schnell abgeschlossen zu haben. Für die Aufgaben 3 und 4 äußerte die überwiegende Mehrheit, bestehend aus jeweils 14 Personen, eine klare Zustimmung zur schnellen Bewältigung der Aufgaben. Die Mehrzahl der Teilnehmenden signalisierte ihre Zustimmung durch die Auswahl von *Agree* und *Strongly agree*.

Hinsichtlich der objektiven Effizienz zeigt das folgende Diagramm 7.3 die Aufgabenabschlusszeiten für die fünf verschiedene Aufgaben. Die Daten wurden gesammelt, indem die Zeit gemessen wurde, die einzelne Teilnehmer:innen benötigten, um jede Aufgabe abzuschließen. Ein Boxplot wird verwendet, um die Verteilung der Abschlusszeiten für jede Aufgabe zu visualisieren.

Die horizontale Achse repräsentiert die verschiedenen Aufgaben, während die vertikale Achse die Zeit in Sekunden darstellt. In den Plots ist erkennbar, dass Teil-

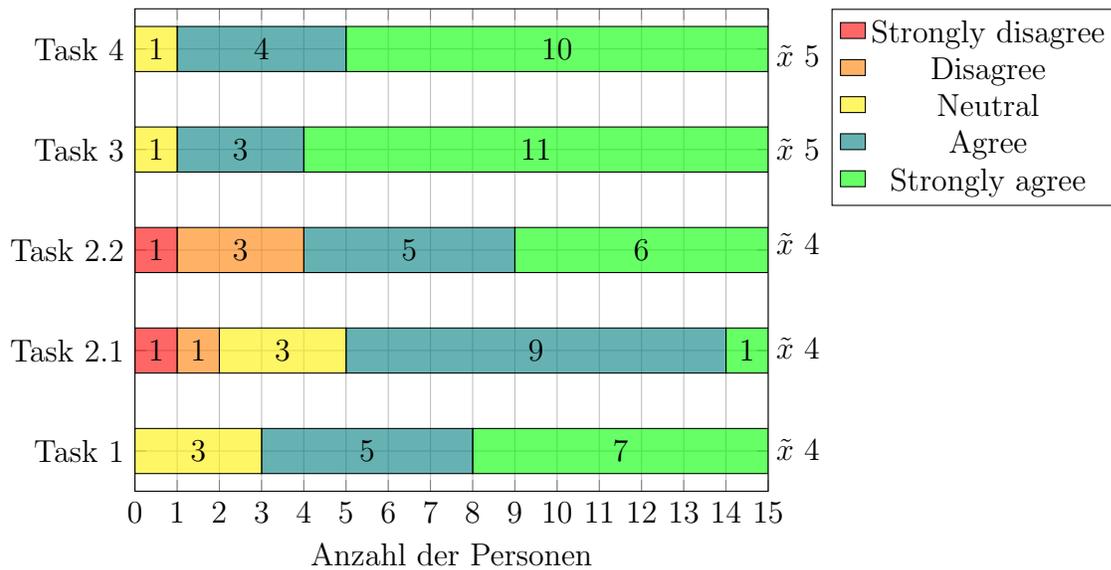


Abbildung 7.2: Likert-Skala „I finished the task fast.“ (Effizienz)

nehmende im Durchschnitt die geringste Zeit für die Bewältigung von Aufgabe 1 und Aufgabe 3 benötigten. Dies legt nahe, dass diese Aufgaben entweder weniger komplex waren oder dass die Teilnehmenden über mehr Erfahrung oder bereits vorhandenes Wissen verfügten, was zu einer schnelleren Bewältigung führte. Auffällig ist jedoch, dass Aufgabe 2.1 den Teilnehmenden im Durchschnitt am meisten Zeit abverlangte. Des Weiteren zeigen die Plots Ausreißer für lange Aufgabenabschlusszeiten bei Aufgabe 2.1, 2.2 und 3. Dies deutet darauf hin, dass es individuelle Fälle gab, in denen Teilnehmende signifikant länger für den Abschluss dieser Aufgaben benötigten als der Durchschnitt.

Die nachstehende Abbildung 7.4 zeigt die Aufgabenabschlusszeiten in aggregierter Form. Hier wurden die Zeiten zu Intervallen aus zweieinhalb und fünf Minuten zusammengefasst. Jeder farbige Balken repräsentiert ein Zeitintervall, und die Breite jedes Balkens zeigt die Anzahl der Personen an, die ihre Aufgabe innerhalb dieses Intervalls abgeschlossen haben. Der Mittelwert in [Minuten: Sekunden] steht jeweils neben dem entsprechenden Graph für die Aufgaben.

Die aggregierten Balkendiagramme bieten einen alternativen Blickwinkel auf die Verteilung der Aufgabenabschlusszeiten im Vergleich zu den Boxplots. Während die Boxplots detaillierte Einblicke in die gesamte Verteilung der Abschlusszeiten geben und Ausreißer deutlich hervorheben, vermitteln die Balkendiagramme eine klare Vorstellung darüber, wie viele Personen ihre Aufgaben innerhalb bestimmter Zeitinter-

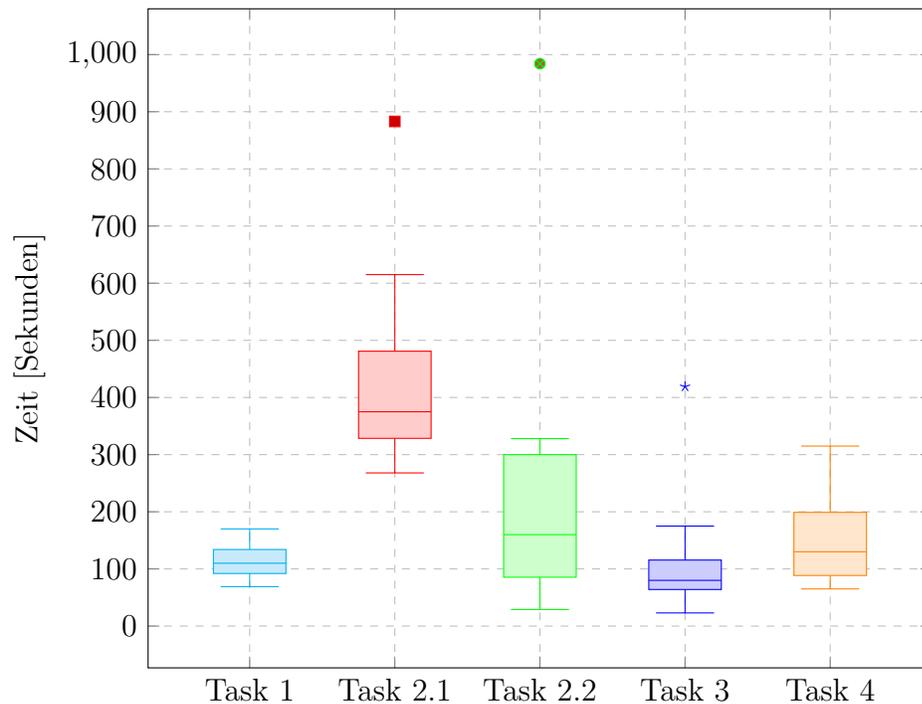


Abbildung 7.3: Boxplot für die Aufgabenabschlusszeiten (Effizienz)

alle abgeschlossen haben. Betrachten wir zunächst Aufgabe 1: Hier wurde von allen Teilnehmenden ein schneller Abschluss innerhalb von maximal fünf Minuten verzeichnet, was auf eine hohe Effizienz hinweist.

Für Aufgabe 2.1 ist eine interessante Variation zu erkennen: Während einige Teilnehmende die Aufgabe in weniger als fünf Minuten bewältigen konnten, benötigte etwa die Hälfte der Gruppe fünf bis sieben Minuten. Zwei Personen benötigten sogar mehr als zehn Minuten, was möglicherweise auf spezifische Schwierigkeiten oder die Komplexität der Aufgabenstellung hinweist.

Die Ergebnisse für Aufgabe 2.2 zeigen die breiteste Streuung der Zeiten. Die Mehrheit der Teilnehmenden schloss die Aufgabe innerhalb von fünf Minuten ab, jedoch gab es auch Personen, die zwischen sieben und fünfzehn Minuten benötigten. Dies deutet darauf hin, dass Aufgabe 2.2 möglicherweise weniger einheitlich war oder dass die Teilnehmenden unterschiedliche Herangehensweisen hatten.

Für Aufgabe 3 sehen wir, dass die meisten Teilnehmenden ihre Aufgabe innerhalb von zweieinhalb Minuten abschließen konnten, aber es gibt auch eine kleine Anzahl von Personen, die mehr Zeit benötigten. Dies könnte auf bestimmte Herausforderungen bei der Bewältigung der Aufgabe hinweisen.

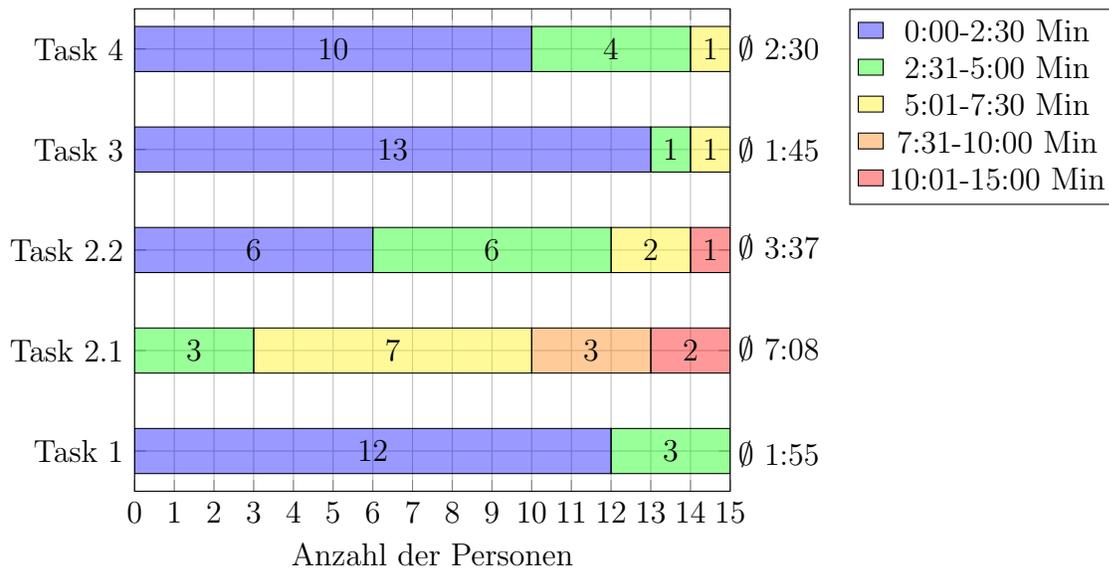


Abbildung 7.4: Zeiten für den Abschluss der Aufgaben in aggregierter Form (Effizienz)

Schließlich für Aufgabe 4: Die meisten Teilnehmenden schlossen diese innerhalb von null bis zweieinhalb Minuten ab, aber es gibt auch einige, die bis zu fünf Minuten benötigten. Hier könnten individuelle Unterschiede oder unterschiedliche Herangehensweisen zur Erklärung dieser Variation beitragen.

Die Metrik der Effizienz wird anhand mehrerer statistischer Tests getestet. Diese fünf Tests beinhalten das Problem, dass durch die wiederholte Testung die Wahrscheinlichkeit steigt, fälschlicherweise statistisch signifikante Resultate (Typ 1-Fehler) zu erhalten [20]. Deswegen wird auf die Ergebnisse abschließend die Bonferroni-Holm-Korrektur angewandt. Tabelle 7.4 zeigt die Ergebnisse der statistischen Tests auf die Effizienz.

Zunächst wurden alle Datensätze für die einzelnen Aufgaben auf eine Normalverteilung mithilfe des Shapiro-Wilk-Tests überprüft. Daraus lässt sich schließen, dass die Daten von Aufgabe 1 und 4 normalverteilt sind, woraufhin der t-Test für die statistische Analyse verwendet werden kann. Die Daten für die Aufgaben 2.1, 2.2 und 3 sind nicht normalverteilt, demnach wird hierfür der Wilcoxon Signed-Rank-Test verwendet. In der letzten Spalte der Tabelle befindet sich der angepasste  $p$ -Wert durch die Bonferroni-Holm-Korrektur.

Die Alternativhypothese bezüglich der Effizienz besagt, dass die Teilnehmenden weniger als fünf Minuten (300 Sekunden) für eine Aufgabe benötigen. Im Einzelnen

Effizienz	Shapiro-Wilk			t-Test		Wilcoxon Rank	Signed-	$p_a$	$H_0$ abl.?
	$W$	$p$	NV?	$t$	$p$	$Z$	$p$		
Task 1	0,956	0,617	✓	-22,186	< 0,001	—	—	< 0,001	✓
Task 2.1	0,83	0,01	×	—	—	3,249	0,999	0,999	×
Task 2.2	0,67	< 0,001	×	—	—	-2,027	0,021	0,043	✓
Task 3	0,067	< 0,001	×	—	—	-3,842	< 0,001	< 0,001	✓
Task 4	0,089	0,052	✓	-7,256	< 0,001	—	—	< 0,001	✓

Tabelle 7.4: Ergebnisse statistischer Tests in Bezug auf Effizienz

sehen die Ergebnisse der statistischen Tests in Bezug auf die Effizienz wie folgt aus:

**Task 1:** Die Ergebnisse des Einstichproben-t-Tests deuten darauf hin, dass es einen signifikant großen Unterschied zwischen Gruppe 1 ( $M = 115,8$ ,  $SD = 32,2$ ) und dem Populationsmittelwert ( $M = 300$ ) gibt,  $t(14) = 22,2$ ,  $p < 0,001$ , Cohen's  $d = 5,7$ . Die Nullhypothese  $H_{0,\text{Effizienz}}$  für Aufgabe 1 wird damit abgelehnt. Die Teilnehmenden haben weniger als fünf Minuten für Aufgabe 1 benötigt.

**Task 2.1:** Die Ergebnisse des Wilcoxon Signed-Rank-Tests zeigen, dass es einen nicht signifikanten großen Unterschied zwischen Vorher ( $Mdn = 300$ ,  $n = 15$ ) und Nachher ( $Mdn = 375$ ,  $n = 15$ ) gibt,  $W+ = 112$ ,  $p = .999$ ,  $r = 0,8$ . Die Nullhypothese  $H_{0,\text{Effizienz}}$  für Aufgabe 2.1 wird damit angenommen. Die Teilnehmenden haben genau oder mehr als fünf Minuten für Aufgabe 2.1 benötigt.

**Task 2.2:** Die Ergebnisse des Wilcoxon Signed-Rank-Tests zeigen, dass es einen signifikanten großen Unterschied zwischen Vorher ( $Mdn = 300$ ,  $n = 15$ ) und Nachher ( $Mdn = 160$ ,  $n = 15$ ) gibt,  $Z = -2$ ,  $p = .021$ ,  $r = -0,6$ . Die Nullhypothese  $H_{0,\text{Effizienz}}$  für Aufgabe 2.2 wird damit abgelehnt. Die Teilnehmenden haben weniger als fünf Minuten für Aufgabe 2.2 benötigt.

**Task 3:** Die Ergebnisse des Wilcoxon Signed-Rank-Tests deuten darauf hin, dass es einen signifikant großen Unterschied zwischen Vorher ( $Mdn = 300$ ,  $n = 15$ ) und Nachher ( $Mdn = 80$ ,  $n = 15$ ) gibt,  $W+ = 1$ ,  $p < .001$ ,  $r = -1$ . Die Nullhypothese  $H_{0,\text{Effizienz}}$  für Aufgabe 3 wird damit abgelehnt. Die Teilnehmenden haben weniger als fünf Minuten für Aufgabe 3 benötigt.

**Task 4:** Die Ergebnisse des Einstichproben-t-Tests zeigen, dass es einen signifikant großen Unterschied zwischen Gruppe 1 ( $M = 150,4$ ,  $SD = 79,9$ ) und dem Populationsschnitt ( $M = 300$ ) gibt,  $t(14) = 7,3$ ,  $p < .001$ , Cohen's  $d = 1,9$ . Die Nullhypothese  $H_{0, \text{Effizienz}}$  für Aufgabe 4 wird damit abgelehnt. Die Teilnehmenden haben weniger als fünf Minuten für Aufgabe 4 benötigt.

### 1. Erkenntnis: Effizienz

Die Teilnehmenden benötigten unter 5 Minuten für die Aufgaben 1, 2.2, 3 und 4. Für Aufgabe 2.1 haben die Teilnehmenden mehr als 5 Minuten benötigt.

## 7.6.3 Effektivität

Ebenso kann anhand der Video- und Audioaufzeichnung für die Effektivität gemessen werden, ob die Teilnehmenden die Aufgaben erfolgreich abgeschlossen haben. Hierbei gibt ein boolescher Wert (1 oder 0) an, ob die Aufgabe erfolgreich abgeschlossen wurde oder nicht. Nach Beendigung der Aufgabe haben die Teilnehmenden ihre persönliche Einschätzung bezüglich des erfolgreichen Abschlusses abgegeben. Diese Daten werden auf einer Ordinalskala festgehalten. Die vorliegende Grafik 7.5 veranschaulicht die Antworten in Bezug auf die Aussage „I finished the task successfully“. Neben der Grafik stehen die jeweiligen Werte für den Median pro Aufgabe, dieser liegt für alle Aufgaben bei 5. Dies weist darauf hin, dass die Mehrheit der Teilnehmenden der Aussage zustimmen, die Aufgabe erfolgreich abgeschlossen zu haben.

Bei jeder Aufgabe hat die überwiegende Mehrheit (jeweils mindestens 10 Personen) mit *Strongly agree* angegeben, dass sie die Aufgabe erfolgreich abgeschlossen haben. Für Aufgabe 1 haben lediglich zwei Personen mit *Agree* angegeben, dass sie die Aufgabe erfolgreich erledigt haben. Der Rest der Teilnehmenden haben die Antwort *Strongly agree* gegeben. Aufgabe 2.1 wurde in Bezug auf ihre erfolgreiche Bearbeitung von einer Person mit *Disagree* beantwortet, jeweils zwei haben *Neutral* und *Agree* angegeben. Der Rest der Teilnehmenden (10 Personen) waren der Meinung, dass sie die Aufgabe erfolgreich abgeschlossen (*Strongly agree*) haben. Demnach zeigt sich für Aufgabe 2.1 die größte Streuung in Hinblick auf die Antworten, was auf eine höhere Komplexität der Aufgabe oder auf eine unklare Aufgabenstellung hinweisen könnte. Aufgabe 2.2 wurde nach Beurteilung von 12 Personen erfolgreich abgeschlossen, die restlichen drei Personen haben entweder *Neutral* oder *Agree* geantwortet. Aufgabe 3 und 4 unterscheiden sich nicht stark von Aufgabe 1 in Hinblick

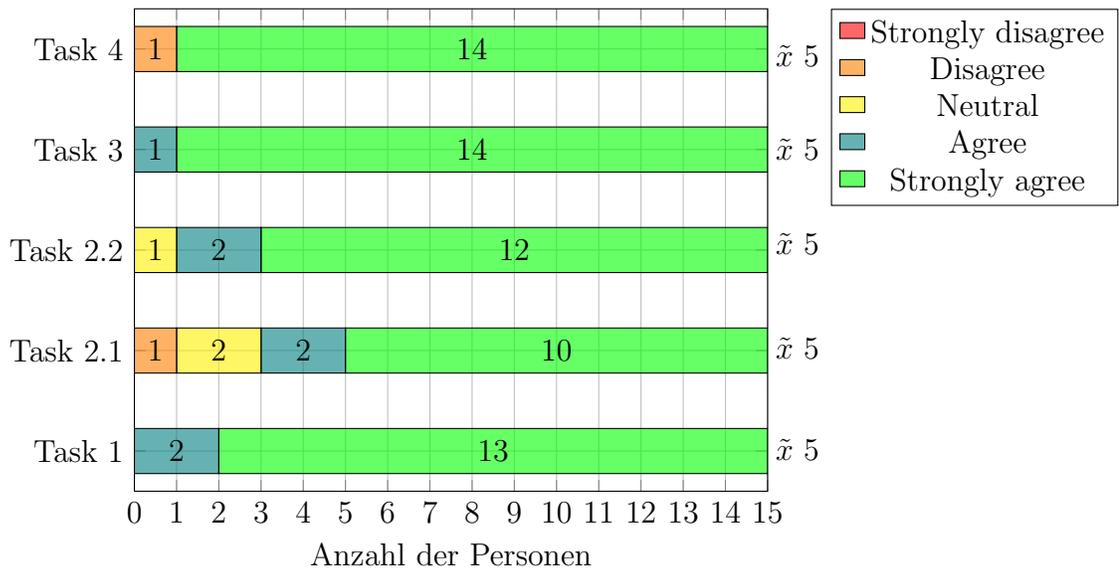


Abbildung 7.5: Likert-Skala „I finished the task successfully.” (Effektivität)

auf die Einschätzungen der Effektivität. Für beide Aufgaben haben 14 Personen mit *Strongly agree* geantwortet. Die eine Person hat für Aufgabe 3 ein *Agree* angekreuzt. Für Aufgabe 4 wurde hier ein *Disagree* angegeben, was hier als Ausreißer betrachtet werden kann und in der Diskussion weiter analysiert wird.

Hinsichtlich der objektiven Effektivität zeigt das folgende Diagramm 7.6 wieviele Personen die jeweiligen Aufgaben erfolgreich abgeschlossen haben. Jede Aufgabe wurde von der Mehrheit erfolgreich abgeschlossen, lediglich Aufgabe 2.2 und 3 stehen mit jeweils vier nicht erfolgreichen Abschlüssen heraus. Für die Aufgaben 1, 2.1 und 4 haben lediglich eine oder zwei Personen die Aufgabe nicht erfolgreich abgeschlossen.

Die Ergebnisse der objektiven Effektivität werden in aggregierter Form analysiert. Das bedeutet, dass wir überprüfen, ob die Teilnehmenden mindestens drei der fünf Aufgaben erfolgreich abschließen. Die nachfolgende Grafik 7.7 zeigt, wieviele Aufgaben die jeweiligen Teilnehmenden erfolgreich abgeschlossen haben. Hierbei spiegelt die Teilnehmenden-ID nicht die Reihenfolge der Teilnahme am Experiment wieder um die Anonymisierung zu gewährleisten.

Von allen Teilnehmenden haben 11 Personen mindestens vier der insgesamt fünf Aufgaben erfolgreich abgeschlossen. Drei Personen haben drei Aufgaben erfolgreich abgeschlossen und lediglich eine Person hat als Ausreißer zwei Aufgabe erfolgreich abgeschlossen. Dieses Ergebnis könnte auf die Intuitivität der Benutzeroberfläche

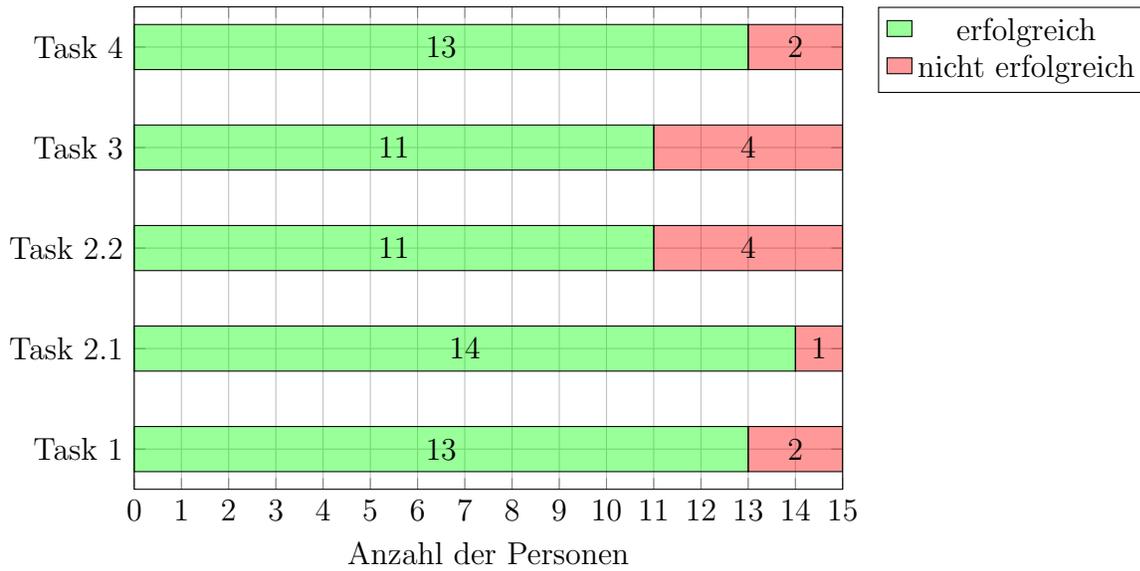


Abbildung 7.6: Anzahl erfolgreicher Aufgabenabschlüsse pro Aufgabe (Effektivität)

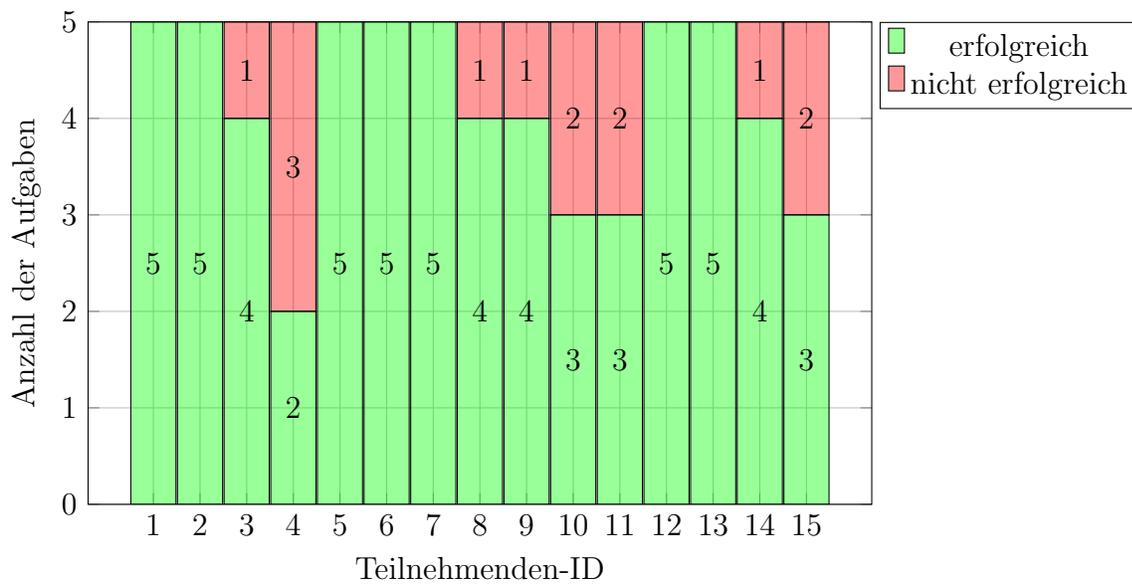


Abbildung 7.7: Anzahl erfolgreicher und nicht erfolgreicher Aufgabenabschlüsse pro Person (Effektivität)

hinweisen, sodass es den meisten Teilnehmenden gelang, die Aufgaben erfolgreich mit dem Dashboard abzuschließen. Dieser Datensatz dient als Eingabe für den folgenden statistischen Test, die Ergebnisse sind in Tabelle 7.5 dargestellt.

Effizienz	Shapiro-Wilk			Wilcoxon Signed-Rank		$H_0$ ablehnen?
	$W$	$p$	NV?	$Z$	$p$	
aggregierte Aufgabenabschlüsse (Task 1 bis 4)	0,816	0,00594	×	2,87	0,004	✓

Tabelle 7.5: Ergebnisse statistischer Tests in Bezug auf Effektivität

Zunächst wurden die Daten mithilfe des Shapiro-Wilk-Tests auf eine Normalverteilung geprüft. Die vorliegenden Daten sind demnach nicht normalverteilt, woraufhin der Wilcoxon Signed-Rank-Test verwendet werden kann. Die Alternativhypothese bezüglich der Effektivität besagt, dass die Teilnehmenden mehr als drei von fünf Aufgaben erfolgreich abschließen. Die Ergebnisse des Wilcoxon Signed-Rank-Tests zeigen, dass es einen signifikanten großen Unterschied zwischen Vorher (Mdn=3, n=15) und Nachher (Mdn=4, n=15) gibt,  $W+=75$ ,  $p = 0,004$ ,  $r=0,8$ . Die Nullhypothese  $H_{0,\text{Effektivität}}$  wird damit abgelehnt. Die Teilnehmenden haben mehr als drei der fünf Aufgaben erfolgreich abgeschlossen.

## 2. Erkenntnis: Effektivität

*Die Teilnehmenden haben mehr als die Hälfte der Aufgaben erfolgreich abgeschlossen.*

### 7.6.4 Zufriedenheit

Die Zufriedenheit wurde anhand des gekürzten User Experience Score einmalig nach Abschluss aller Aufgaben gemessen. Zur Auswertung wird hier das bereitgestellte Analyseverfahren <sup>1</sup> genutzt. In der folgenden Tabelle 7.6 befindet sich ein Überblick über die abgegebenen Antworten des UEQ mit den Angaben von Durchschnitt, Varianz, Standardabweichung und Anzahl der Angaben für jede Skala. Der gekürzte UEQ teilt sich in vier Qualitätseigenschaften pragmatischen und vier des hedonischen Typs. Schrepp et al. [47] charakterisieren die Aspekte des pragmatischen Typs als Merkmale, die die Interaktionsqualität beschreiben und in Bezug auf die Aufgaben

<sup>1</sup><https://www.ueq-online.org/>

oder Ziele stehen, die Nutzende beim Einsatz des Produkts erreichen möchten. Die hedonischen Qualitätseigenschaften hingegen stehen in Verbindung mit dem Vergnügen oder Spaß, den Nutzende bei der Nutzung des Produkts empfinden.

Item	Skala	Mittelwert	Varianz	Standardabweichung	Anzahl Angaben	Negativ	Positiv
1	Pragmatic Quality	2,5	0,6	0,7	15	obstructive	supportive
2	Pragmatic Quality	1,6	1,1	1,1	15	complicated	easy
3	Pragmatic Quality	1,9	1,4	1,2	15	inefficient	efficient
4	Pragmatic Quality	1,9	1,6	1,3	15	confusing	clear
5	Hedonic Quality	2,1	1,4	1,2	15	boring	exciting
6	Hedonic Quality	2,5	0,7	0,8	15	not interesting	interesting
7	Hedonic Quality	1,5	1,0	1,0	15	conventional	inventive
8	Hedonic Quality	1,4	1,8	1,3	14	usual	leading edge

Tabelle 7.6: Beschreibung der UEQ-Skalen und statistische Kennzahlen

Für fast alle Skalen wurden 15 Antworten gegeben, für die Skala *usual/leading edge* hat eine Person keine Antwort gegeben, da sie der Meinung war diesen Aspekt nicht beurteilen zu können. Die Antworten auf den siebenstufigen Skalen werden auf Werte von -3 bis 3 abgebildet. Anschließend ergeben sich daraus Mittelwert, Varianz und Standardabweichung.

In Bezug auf die Mittelwerte bedeutet ein Wert von  $< 0,8$  eine negative Evaluation, ein Wert von  $> 0,8$  eine positive, Werte dazwischen weisen auf ein neutrales Ergebnis hin. Laut Auswertungsverfahren werden Mittelwerte über 2 selten beobachtet, da Teilnehmende vermeiden, Extremwerte anzukreuzen. Somit kann ein Mittelwert von etwa 1,5 als äußerst positiv angesehen werden. Alle Skalen weisen hier eine positive Evaluation von  $> 0,8$  mit Mittelwerten von mindestens 1,4 (*usual/leading edge*) auf. Der höchste Mittelwert findet sich auf den Skalen *obstructive/supportive*

und *not interesting/interesting* mit 2,5. Die Teilnehmenden bewerten demnach, dass das Dashboard diese Kriterien besonders positiv umsetzt.

Die geringste Varianz mit 0,6 und der Standardabweichung von 0,7 lassen sich für die Skala *obstructive/supportive* beobachten. Daraus lässt sich schließen, dass die Teilnehmenden hier sehr ähnlich angekreuzt haben. Die Skala *usual/leading edge* verzeichnet die höchste Varianz mit 1,8, zugleich ebenso die höchste Standardabweichung mit 1,3 gleichauf mit der Skala *confusing/clear*. Bei diesen Skalen waren sich die Teilnehmenden am uneinigsten mit ihren Antworten. Das könnte auf die unterschiedliche Interpretation der Skalen durch die Teilnehmenden zurückzuführen sein. So haben Teilnehmende ihre Antwort teilweise auf die Dashboardtechnologie im Allgemeinen oder auf Dashboards im speziellen Wissensgraphenbereich bezogen.

Die durchschnittlichen Werte pro Teilnehmer:in für ihre Bewertungen im UEQ sind in Tabelle 7.7 dargestellt. Diese Werte umfassen sowohl die pragmatische als auch die hedonische Skala sowie den Gesamtdurchschnitt. Auch hier spiegelt die Teilnehmenden-ID nicht die Reihenfolge der Teilnahme im Experiment wieder um Anonymisierung zu gewährleisten.

Teilnehmenden-ID	$\emptyset$ Pragmatic Quality	$\emptyset$ Hedonic Quality	$\emptyset$ Gesamtskala
1	2,25	2,75	2,50
2	1,50	2,67	2,00
3	2,25	1,75	2,00
4	2,25	2,50	2,38
5	1,00	1,50	1,25
6	3,00	1,00	2,00
7	2,00	1,50	1,75
8	1,25	1,50	1,38
9	1,75	2,00	1,88
10	1,75	2,00	1,88
11	0,50	0,25	0,38
12	2,50	3,00	2,75
13	2,00	2,25	2,13
14	2,50	2,25	2,38
15	3,00	1,50	2,25
Gesamt	1,96	1,894	1,93

Tabelle 7.7: Durchschnittswerte pro Person für Bewertungen im UEQ

Die Gesamtmittelwerte werden nun für die statistische Analyse verwendet. Die

Alternativhypothese bezüglich der Zufriedenheit besagt, dass die Teilnehmenden eine positive Nutzererfahrung ( $UEQ > 0,8$ ) mit dem Dashboard haben. Die Daten werden zunächst auf Normalverteilung geprüft.

Zufriedenheit	Shapiro-Wilk			t-Test		$H_0$ ablehnen?
	$W$	$p$	NV?	$Z$	$p$	
UEQ-Bewertungen	0,90256	0,10415	✓	7,4847	< 0,001	✓

Tabelle 7.8: Ergebnisse statistischer Tests in Bezug auf Zufriedenheit

Dadurch, dass die Daten normalverteilt sind, kann der t-Test verwendet werden. Die Ergebnisse des Einstichproben-t-Tests deuten darauf hin, dass es einen signifikant großen Unterschied zwischen Gruppe 1 ( $M = 1,9$ ,  $SD = 0,6$ ) und dem Populationsmittelwert ( $M = 0,8$ ) gibt,  $t(14) = 7,5$ ,  $p < 0,001$ , Cohen's  $d = 1,9$ . Die Nullhypothese  $H_{0,Zufriedenheit}$  wird damit abgelehnt. Die Teilnehmenden haben eine positive Nutzererfahrung mit dem Dashboard ( $UEQ > 0,8$ ).

Das Analyseverfahren des UEQ beinhaltet einen Benchmark aus einem Vergleichsdatensatz mit 21175 Bewertungen. Dieser Datensatz basiert auf Angaben aus dem vollständigen UEQ, da nicht genügend Daten für die gekürzte Version vorhanden sind.

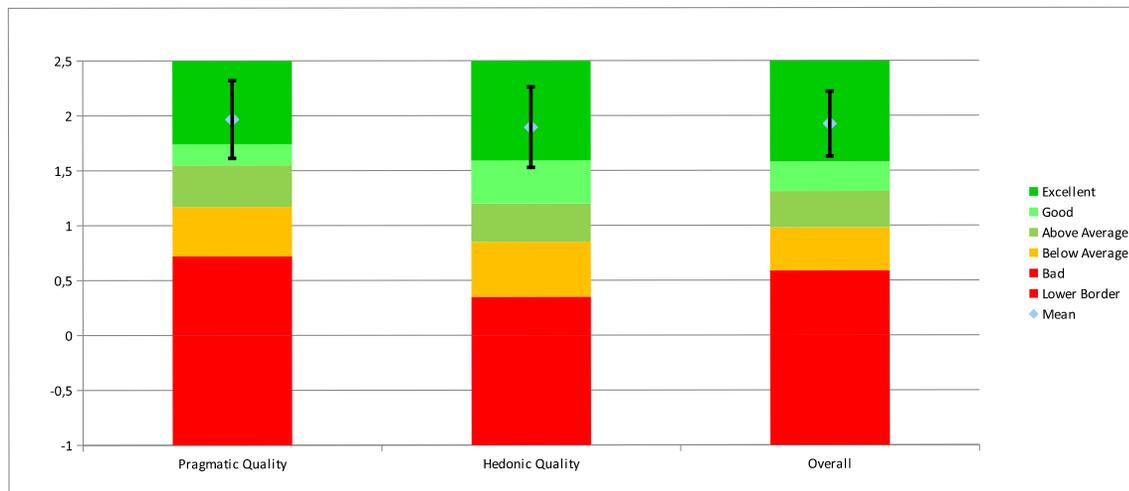


Abbildung 7.8: Benchmark des UEQ

Die Grafik 7.8 zeigt Balkendiagramme für die hedonische und pragmatische Skala sowie die Gesamtbewertung in Hinblick auf die Mittelwerte des UEQ. Die Streuung

des Datensatzes aus dem durchgeführten Experiment ist anhand der Fehlerbalken zu erkennen, die alle in einem positiven Bewertungsbereich liegen. Somit liegen die verzeichneten Ergebnisse in den oberen 10% im Benchmark.

### 3. Erkenntnis: Zufriedenheit

Die Teilnehmenden haben laut UEQ eine positive Nutzererfahrung mit dem Dashboard ( $UEQ > 0,8$ ).

## 7.6.5 Bedarfe des Kuratierungs- und Gemeinschaftsbildungsteams

Die Bedarfe des Kuratierungs- und Gemeinschaftsbildungsteams finden sich in Tabelle 7.2 wieder. Die Bedarfe B1 bis B4 wurden im Dashboard umgesetzt, der Bedarf B5 diente als Kontrollfrage und sollte ablehnend von den Teilnehmenden beantwortet werden. Das Diagramm 7.9 veranschaulicht die gegebenen Antworten der Teilnehmenden.

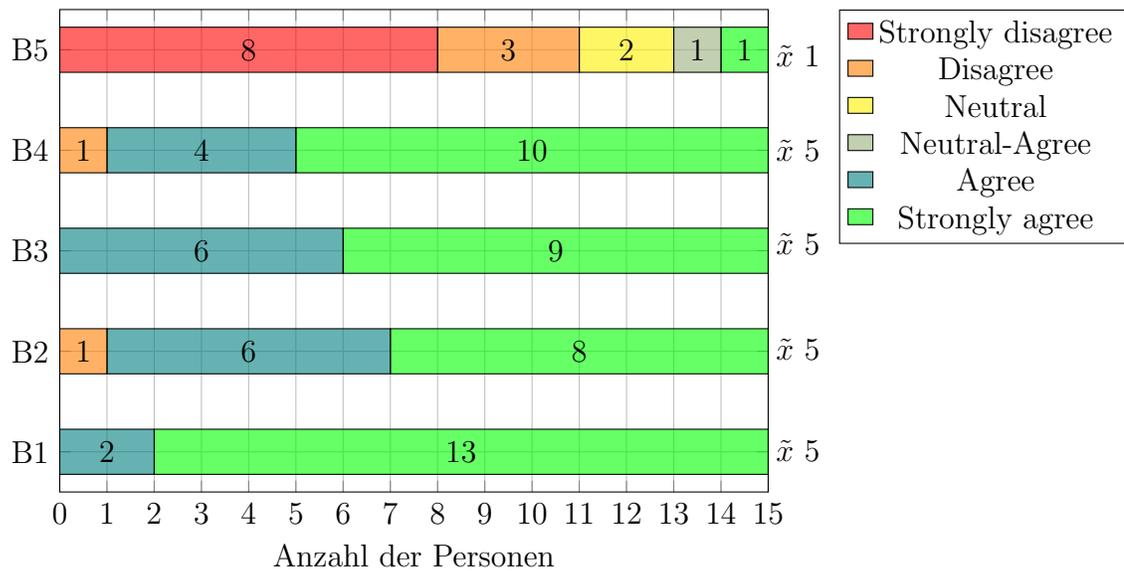


Abbildung 7.9: Likert-Skala (Bedarfe des Teams)

Die Medianwerte für die Bedarfe B1 bis B4 liegen bei 5, was auf eine allgemeine Zustimmung zur Erfüllung dieser Bedarfe hinweist. Hingegen liegt der Medianwert für Bedarf B5 bei 1, was auf eine Ablehnung dieser Umsetzung deutet. Der Bedarf

B1, das Tracking von Besucherpfeaden, erhält am meisten Zustimmung in der Umsetzung mit 13 Angaben von *Strongly agree* und 2 Angaben für *Agree*. Der Bedarf B3, Werkzeuge zur Überwachung der Qualität im ORKG bereitzustellen, erhält ähnlich viel Zustimmung mit 6 Angaben von *Agree* und 9 Angaben von *Strongly agree*. Die Bedarfe B2, Möglichkeiten Ressourcen als inakkurat oder unvollständig zu markieren, und B4, einen Überblick für Bereiche für Verbesserungen zu stellen, wurden jeweils ein Mal mit *Neutral* angekreuzt. Die Umsetzung des Bedarfs B5, Tracking von Aktivitäten der internen und externen Nutzenden, wurde vorrangig abgelehnt. Hier hat die Mehrheit aus 8 Personen *Strongly disagree* angegeben, eine Person hat ihr Kreuz zwischen *Neutral* und *Agree* gesetzt. Die Ergebnisse weisen darauf hin, dass die Bedarfe B1 bis B4 für Teilnehmenden des Experiments weitestgehend im Dashboard umgesetzt wurden, Bedarf B5 wurde als Kontrollbedarf von der Mehrheit erwartungsgemäß abgelehnt.

## 7.7 Einschränkungen der Validität

Um die Gültigkeit des Experiments genauer zu untersuchen, werden in den folgenden Unterkapiteln Bedrohungen der Validität beleuchtet. Diese Abschnitte lehnen an die vorgestellte Struktur von Wohlin et. al in *Experimentation in Software Engineering* [57] an. Die Validität, und damit die Vertrauenswürdigkeit, muss zu jedem Zeitpunkt des Experiments berücksichtigt werden.

### 7.7.1 Konstruktvalidität

Die Konstruktvalidität beschreibt, in wie weit die operativen Maßnahmen, die untersucht werden, das repräsentieren, was mit der Forschungsfrage untersucht wird. Sie verbindet somit die theoretischen Konzepte mit den beobachteten Daten und bestimmt, inwieweit die Ergebnisse eines Experiments auf die zugrunde liegende Theorie verallgemeinert werden können. Die Bedrohungen für die Konstruktvalidität können in zwei Kategorien unterteilt werden: *Design-Bedrohungen*, die Aspekte des Experimentdesigns betreffen, und *soziale Bedrohungen*, die das soziale Verhalten der Teilnehmenden unter den experimentellen Bedingungen betreffen.

**Design-Bedrohungen.** Ein Beispiel ist die Bedrohung durch eine *unzureichende voroperative Erläuterung der Konstrukte*. Dies bezieht sich darauf, dass die Konzepte und Variablen, die in einer Studie untersucht werden, nicht klar definiert sind, bevor das Experiment durchgeführt wird. Dies kann zu Missverständnissen und inkonsistenten Messungen führen und die Konstruktvalidität beeinträchtigen. Im Rahmen des durchgeführten Experiments betrifft dies insbesondere die objektive Messung

der Effektivität. Die Herausforderung besteht darin, zu bestimmen, wie die erfolgreiche Beendigung einer Aufgabe mithilfe des Dashboards gemessen werden kann. Dies ist komplex, da es oft verschiedene Wege gibt, um ein bestimmtes Ziel zu erreichen. Darüber hinaus kann es mehrere korrekte Antworten geben, und auch das erfolgreiche Navigieren oder das Verständnis eines Konzepts können als erfolgreiche Aufgabenabschlüsse betrachtet werden. In diesem Zusammenhang wurde ausschließlich auf das angestrebte Ergebnis geachtet, das die Teilnehmenden mit dem Abschluss der Aufgabe erreichen wollten. Es kam vor, dass sie gezielt eine andere Ressource wählten, weil dies für sie persönlich im Kontext der Kuratierungsarbeit sinnvoller war. Wenn die Teilnehmenden das beabsichtigte Ergebnis erreicht haben, wurde dies als erfolgreicher Abschluss gewertet.

Die *Verzerrung durch einen Einzelfall* tritt auf, wenn ein Experiment nur eine unabhängige Variable untersucht. Dadurch kann die Vielfalt der betrachteten Variablen eingeschränkt sein, was die Generalisierbarkeit der Ergebnisse beeinträchtigen kann. In dieser Studie war es nicht möglich, eine alternative Methode zu untersuchen, da es derzeit kein vergleichbares Dashboard gibt und die durch das Dashboard ermöglichte Arbeit ohne dieses nicht durchführbar ist. Daher werden die Ergebnisse des Experiments sorgfältig interpretiert, um die Genauigkeit der Schlussfolgerungen nicht zu gefährden.

Weiterhin gibt es die Bedrohung durch *Wechselwirkungen durch Test und Behandlung*. Diese bezieht sich auf Faktoren, die die Wirkung der unabhängigen Variablen auf die abhängigen Variablen beeinflussen können. Dazu gehört ebenfalls die Testumgebung. Im Experiment nahmen 10 Personen in Präsenz teil, während 5 Personen online teilnahmen. Die Interaktion in Präsenz und online ist naturgemäß unterschiedlich, da bei letzterer die Teilnehmenden ihr eigenes Setup haben und wir nicht nebeneinander vor dem gleichen Bildschirm sitzen. Um möglichst vergleichbare Bedingungen zu schaffen, wurden die Online-Experimente auf demselben Laptop durchgeführt wie die Präsenzexperimente. Dabei wurde die Möglichkeit geboten, sich per TeamViewer auf meinem Laptop einzuloggen um eigenständig zu interagieren. Damit konnte das Verhalten bestmöglich im Vergleich zur Präsenzteilnahme untersucht werden. In den Online-Experimenten besteht weiterhin die Einschränkung, dass die Teilnehmenden nur eine primäre Ansicht haben können, entweder die Aufgabenstellung oder das Dashboard. In Präsenzveranstaltungen konnten die Teilnehmenden beides gleichzeitig betrachten.

**Soziale Bedrohungen.** Die Bedrohung durch die *Erwartungen der Experimentator:innen* kann die Ergebnisse eines Experiments sowohl bewusst als auch unbewusst beeinflussen, basierend auf den Erwartungen, die sie an das Experiment haben. In diesem Experiment könnte zum Beispiel die Tatsache, dass es im Rahmen

einer Masterarbeit durchgeführt wurde, dazu führen, dass die Teilnehmerinnen und Teilnehmer dazu neigen, eher positive Rückmeldungen zu geben, damit die Arbeit ebenfalls positive Ergebnisse verzeichnen kann. Um dieser Bedrohung entgegenzuwirken, habe ich den Fokus der abgegebenen Bewertungen und der Gespräche darauf gelegt, dass das Ziel darin besteht, das Dashboard zu verbessern und es bestmöglich für das Team nutzbar zu machen. Daher wurden die Teilnehmenden ermutigt, ihre Antworten möglichst ehrlich zu geben. Da es weder Belohnungen noch externe Anreize für die Teilnahme gab, unterstützt es die Annahme, dass die Antworten ehrlich und unvoreingenommen waren.

### 7.7.2 Externe Validität

Die externe Validität bezieht sich darauf, inwieweit die Ergebnisse eines Experiments auf eine breitere Population oder Situation verallgemeinert werden können. In diesem Kontext lässt sich annehmen, dass die Stichprobe zunächst repräsentativ für potenzielle Nutzerinnen und Nutzer des Dashboards ist, insbesondere für Personen im Umfeld des ORKG, die mit Kuratierungsarbeit und Gemeinschaftsbildungsprozessen vertraut sind. Darüber hinaus haben nicht nur alle Mitglieder des Kuratierungs- und Gemeinschaftsbildungsteams, die die Kerngruppe darstellen, am Experiment teilgenommen, sondern auch Personen aus dem weiteren Umfeld, was die Stichprobe zusätzlich repräsentativ macht. Zukünftig könnten Teile des Dashboard für die allgemeinen Nutzerinnen und Nutzer des ORKG zur Verfügung gestellt werden, was die externe Validität erhöhen würde. Allerdings ist zu beachten, dass die Aufgaben im Experiment nur teilweise den tatsächlichen Arbeitsabläufen entsprechen, was bedeutet, dass die Ergebnisse lediglich einen ersten Einblick in die Verwendbarkeit des Dashboards bieten. Um die externe Validität zu verbessern, müssten zukünftige langfristige Studien in einer realen Umgebung durchgeführt werden, in der das Dashboard über eine URL zugänglich ist und sich die Teilnehmenden regelmäßig mit dem Dashboard beschäftigen um komplexere Aufgaben zu bewältigen.

### 7.7.3 Interne Validität

Interne Validität bezieht sich auf die Gültigkeit der Schlussfolgerungen innerhalb eines Experiments. Sie beantwortet die Frage, ob die beobachteten Effekte tatsächlich auf die Manipulation der unabhängigen Variable zurückzuführen sind und nicht auf andere Faktoren oder Störeinflüsse. In anderen Worten, interne Validität betrachtet, ob die durchgeführten Messungen und Beobachtungen tatsächlich das messen, was sie messen sollen, und ob die beobachteten Veränderungen oder Unterschiede auf die

experimentelle Intervention zurückzuführen sind. Die interne Validität lässt sich in *Bedrohungen durch eine einzelne Gruppe*, *Bedrohungen durch mehrere Gruppen* und *soziale Bedrohungen* aufteilen. Da das Experiment mit einer Gruppe, ohne Kontrollgruppe, durchgeführt wurde, schauen wir zunächst auf die Bedrohungen durch eine einzelne Gruppe.

Es gibt den Aspekt der *Reifung*, bei der sich die Performance der Teilnehmenden im Verlauf des Experiments positiv oder negativ entwickeln kann. In diesem Fall könnte ein Lerneffekt aufgetreten sein, da die Aufgaben nicht zufällig zugewiesen wurden und die Teilnehmenden im Laufe des Experiments möglicherweise effizienter oder effektiver wurden.

Der Aspekt der *Instrumentierung* bezeichnet die Bedrohung durch mangelhafte Artefakte oder Dokumente. Während des Experiments wurden einige unvorhergesehene Fehler im Prototypen des Dashboards festgestellt, die sich negativ auf die Leistung der Teilnehmenden auswirken konnten. Zum Beispiel traten verlängerte Ladezeiten beim Sortieren von Listen im einstelligen Sekundenbereich auf, ohne dass Ladeindikatoren angezeigt wurden. Dies führte dazu, dass die Teilnehmenden vermuteten, sie hätten einen Fehler gemacht. Des Weiteren war eine Sortierfunktionalität fehlerhaft implementiert, was dazu führte, dass die Teilnehmenden nicht sofort das gewünschte Ergebnis erzielen konnten. Darüber hinaus kam es einmal zu einer Ausnahme beim Laden des Netzwerkgraphen, was jedoch durch das Neuladen des Dashboards behoben werden konnte. Obwohl diese Probleme als Bugs betrachtet werden können, die bei der Entwicklung eines Prototypen auftreten können, sollten sie nicht als kritisch angesehen werden. Vielmehr bieten sie wertvolle Einblicke und Input, um das Dashboard benutzerfreundlicher und fehlerfreier zu gestalten.

Die *Ambiguität über die Richtung des kausalen Einflusses* beschreibt die Unklarheit darüber, ob beobachtete Effekte auf die Leistung der Teilnehmenden oder auf die Komplexität des Programms zurückzuführen sind. Dies kann die Interpretation der Ergebnisse erschweren. Die gemessenen Aufgabenabschlusszeiten für die Effizienz können durch die Performance der Teilnehmenden bestehen, der Komplexität des Programms oder der Aufgabenstellungen. Ebenso für die erfolgreich oder erfolglos abgeschlossenen Aufgaben.

Unter den sozialen Bedrohungen könnte die Bedrohung durch die *Diffusion oder Imitation von Behandlungen* auftreten. Teilnehmende, die bereits Erfahrung mit Prototypen durch ihre Tätigkeit in der Softwareentwicklung haben, könnten ihr Verhalten im Experiment entsprechend anpassen, was die Ergebnisse beeinflussen könnte. Weiterhin könnte die Teilnahme an den vorherigen zwei Evaluationen des Dashboards einen Einfluss auf die Performance dieser Teilnehmenden haben, da ihnen eine gewisse Vertrautheit mit der Anwendung unterstellt werden könnte. Dies ist

aber aufgrund der Korrelationsanalysen auszuschließen, da keine signifikanten Zusammenhänge zwischen der Vertrautheit durch Teilnahme an den Evaluationen und den jeweiligen Messwerten nachgewiesen werden konnten.

Insgesamt könnte die interne Validität des Experiments durch die Freiwilligkeit der Teilnehmenden gestärkt worden sein. Es ist jedoch wichtig, diese Bedrohungen sorgfältig zu berücksichtigen, um sicherzustellen, dass die beobachteten Effekte tatsächlich auf die experimentelle Intervention zurückzuführen sind.

### 7.7.4 Schlussfolgerungsvalidität

Die Schlussfolgerungsvalidität betrifft die Gültigkeit der Schlussfolgerungen, die aus den beobachteten Daten gezogen werden. Sie beantwortet die Frage, ob die Schlussfolgerungen, die aus den Ergebnissen einer Studie gezogen werden, tatsächlich gerechtfertigt sind und ob sie auf die getroffenen Annahmen und Interpretationen zurückzuführen sind.

Die statistischen Tests basieren alle auf einer erstmaligen Prüfung auf Normalverteilung und einer darauf folgenden Auswahl des geeigneten Tests. Demnach wird die Schlussfolgerungsvalidität damit erhöht, da keine Bedrohung durch *verletzte Annahmen der statistischen Tests* besteht.

Um die Bedrohung durch die *Fishing- und Fehlerfrequenz* zu mindern, wird nach mehrfacher Testung auf Effizienz mit verschiedenen Tests eine Bonferroni-Holm-Korrektur durchgeführt, um die Wahrscheinlichkeit für fälschlicherweise beobachtete Signifikanz zu reduzieren.

Die *Zuverlässigkeit der Messungen* wird durch die Aufzeichnung der Experimente gewährleistet. Diese Aufnahmen können erneut auf objektive Messungen überprüft werden. Daher sind objektive Messungen den subjektiven vorzuziehen. In diesem Experiment wurden jedoch beide verwendet, um möglichst viele Aspekte zu beleuchten und einen persönlichen Eindruck der Teilnehmenden zu gewinnen, da es um die Benutzerfreundlichkeit geht.

Die Bedrohungen im Rahmen der *Zuverlässigkeit der Behandlungsdurchführung* beziehen sich auf Abweichungen bei der Durchführung der Experimente mit den verschiedenen Teilnehmenden. Dabei sollte der Ablauf so standardisiert wie möglich sein. Da es sich beim Dashboard um einen Prototypen handelt, können unerwartete Fehler auftreten, die nicht ordnungsgemäß in der Implementierung behandelt wurden, sodass sich ein unterschiedliches Verhalten des Dashboards bei den Durchführungen zeigen kann. Zusätzlich habe ich darauf geachtet, allen Teilnehmenden dieselben Erklärungen zu geben und sie auf dieselbe Weise verbal zu unterstützen. Dennoch hat sich erst nach einigen durchgeführten Experimenten eine gewisse Routine entwickelt.

# Kapitel 8

## Diskussion

Die drei gewonnenen Erkenntnisse im Rahmen der Datenanalyse schließen darauf, dass die Entwicklung des Dashboards als gelungen betrachtet werden kann. Das Experiment selbst zielte darauf ab (vgl. Forschungsziel 7.1), sowohl die Bedienbarkeit des Dashboards als auch die Erfüllung der Bedürfnisse des Kuratierungs- und Gemeinschaftsbildungsteams zu untersuchen.

Die **erste Erkenntnis** 7.4 zeigt, dass die Teilnehmenden vier von fünf Aufgaben in unter fünf Minuten abschließen konnten. Die Aufgabe, die den Teilnehmenden am meisten Zeit abverlangte, war die Identifikation eines bestimmten Knotens innerhalb des Netzwerkgraphen für die Besucherstatistiken. Im Zuge der Implementierung hat sich diese Komponente zudem als die aufwändigste erwiesen. Durch den analytischen Charakter dieser Dashboardkomponente fordert sie komplexe Visualisierungen um eine tiefgehende Analyse zu ermöglichen. Die Komponente hat sich zu einer wesentlich umfangreicheren als auch komplexeren Version entwickelt, als zum Zeitpunkt der Erstellung der Hypothesen abzusehen war. Durch diesen Umstand konnte diese Aufgabe schließlich nicht in der erwarteten Zeit von den Teilnehmenden umgesetzt werden.

In der abschließenden Umsetzung der gewonnenen Lehren aus dem zweiten Zyklus erschloss sich zudem das Potenzial, konkrete Besucherpfade zu listen. Diese Funktionalität birgt auch am meisten Potenzial um weiterentwickelt zu werden indem darauf weitere Analysen durchgeführt werden können um differenzierte Besucherprofile zu identifizieren. Zudem haben viele Teilnehmende angemerkt, dass die Benutzeroberfläche im Allgemeinen zu überladen ist und vorallem für neue Nutzer:innen verbessert werden könnte. So war es im Experiment ohne Erklärung meinerseits nicht möglich, den Netzwerkgraphen auf Anhieb zu verstehen und in Einklang mit der geforderten Aufgabenstellung zu bringen. Das lässt sich durch das hohe Abstraktionslevel

---

der erhaltenen Daten von Matomo erklären, die stark modifiziert wurden (vgl. Kapitel 6.3.3) um möglichst konstruktiv dargestellt werden zu können. So war es für die Teilnehmenden nicht trivial, wofür Knoten und Kanten stehen und was farbige Übergänge bedeuten.

Die **zweite Erkenntnis** 7.5 zeigt, dass mehr als die Hälfte der Aufgaben von den Teilnehmenden erfolgreich bewältigt wurden. Die Bewertung des Erfolgs gestaltete sich jedoch nicht trivial, wie bereits in den Einschränkungen der Validität (vgl. Kapitel 7.7) erläutert wurde. Interessanterweise entschieden sich die Teilnehmer:innen gelegentlich bewusst für alternative Ressourcen, um den Aufgaben einen für sie sinnvolleren Zweck zu verleihen, da die Aufgaben lediglich dazu gedacht waren, exemplarisch das Potenzial des Dashboards darzustellen. Zudem bewerteten sie ihren persönlichen Erfolg teilweise anhand der Bereitstellung einer für sie passenden Benutzeroberfläche im Hinblick auf ihre üblichen Arbeitsabläufe. In diesem Zusammenhang kam es vor, dass sie ihren Erfolg der Aufgabe negativ bewerteten, weil sie der Meinung waren, nicht die intendierte Navigation im Dashboard gefunden zu haben. Diese Rückmeldung unterstreicht die Notwendigkeit, die Benutzeroberfläche intuitiver zu gestalten. Teilnehmende, die eine Aufgabe nicht erfolgreich abschlossen, taten dies hauptsächlich aufgrund der vermuteten Komplexität der Aufgabenstellung, was schwierig machte, das genaue Ziel zu erfassen. Daher haben diese Teilnehmenden ihr Ergebnis nicht mit der ursprünglichen Aufgabenstellung abgeglichen.

Die **dritte Erkenntnis** 7.6 zeigt, dass die Teilnehmenden eine überwiegend positive Nutzererfahrung mit dem Dashboard hatten und dieses Ergebnis in den oberen 10% im Vergleich zum Benchmark liegt. Im Rahmen des Think-Aloud-Verfahrens gaben die Teilnehmenden an, dass das Dashboard bereits von Wert ist, trotz deutlichem Verbesserungsbedarf in der Performance und Benutzerfreundlichkeit. Die Teilnehmenden interpretiert die Skalen des UEQ sehr unterschiedlich, so bezogen manche ihre Antworten auf allgemeinere Vergleiche zu gängigen Dashboards, andere wiederum auf den speziellen Kontext der Forschungswissensgraphen. Um ein umfassenderes Bild der Bedienbarkeit, auch in Bezug auf Zufriedenheit, zu bekommen, erfordert es Langzeitstudien, in der die Teilnehmenden regelmäßig mit dem Dashboard arbeiten. Dadurch wird eindeutiger, welche Interaktionsmöglichkeiten die Nutzenden vom Dashboard benötigen, damit es für sie nachhaltig nutzbar ist.

Die **Bedarfe des Teams** wurden von den Teilnehmenden größtenteils als erfüllt bewertet. Speziell die Mitglieder des Kuratierungs- und Gemeinschaftsbildungsteams haben für die überwiegende Mehrheit der Umsetzung von Bedarfen B1 bis B4 ihre vollständige Zustimmung gegeben. Der Bedarf B2, in Bezug auf die Möglichkeit, Ressourcen als inakkurat oder unvollständig zu markieren, wurde als mindestens erfüllt angesehen. Die Mitglieder des Kuratierungs- und Gemeinschaftsbildungsteams

äußerten positive Rückmeldungen und hoben hervor, dass das Dashboard einen bedeutenden Beitrag zur Unterstützung ihrer Arbeit leistet und speziell auf ihre Bedürfnisse zugeschnitten ist. Gleichzeitig wurde angemerkt, dass weitere Verbesserungen möglich sind, insbesondere durch die Integration dynamischerer Workflows.

Weiterhin konnten im Verlauf der Arbeit trotzdem nicht alle ursprünglich geforderten Bedarfe umgesetzt werden. Teilweise lag das an fehlenden Möglichkeiten, die speziellen Daten aus dem ORKG abzufragen. So werden momentan keine Informationen über die Aktivität der Nutzer:innen festgehalten. Im Rahmen der begrenzten Zeit für die Arbeit wurde sich, im Aspekt des MVPs, auf wesentliche Bedarfe konzentriert. Dadurch entstand eine Anwendung, die verdeutlicht, dass die Implementierung einer maßgeschneiderten Anwendung für einen speziellen Wissensgraphen möglich ist, wenn enger Austausch mit den zukünftigen Nutzenden stattfindet und ihre Bedürfnisse und Interessen bekannt sind.

Das Dashboard hat außerdem die Grundlage für einen modularen Ansatz gelegt, um die Qualität anderer Wissensgraphen, insbesondere DBpedia, grundlegend zu untersuchen. Jedoch stieß das Projekt auf eine Grenze in Bezug auf das ursprünglich intendierte modulare Ziel. Während der Konzeption und Entwicklung des Dashboards zeigte sich, dass verschiedene Wissensgraphen teilweise sehr spezifische Unterschiede in ihrer Struktur aufweisen, was bei einer generischen Betrachtung meist nur zu einer sehr allgemeinen und weniger hilfreichen Analyse führt. Der Mehrwert einer generischen Lösung ist dadurch deutlich limitiert. Für eine wertschöpfendere Analyse sind spezifische Anpassungen erforderlich, die nicht auf alle Wissensgraphen gleichermaßen anwendbar sind. Eine Methode des Mappings wäre erforderlich, um die Entitäten zwischen den Graphen vergleichbar zu machen, sodass ähnliche Bedeutungen übertragen werden können. Dabei könnte es sein, dass bestimmte Konzepte in einem Wissensgraphen vorhanden sind, während sie in einem anderen nicht vorkommen, oder dass sie unterschiedlich benannt oder beschrieben werden. Auch das kann durch den Aspekt des Crowdsourcing entstehen, da sich innerhalb einer Gemeinschaft Benennungen etablieren, die eventuell das Gleiche in anderen Wissensgraphen bedeuten, jedoch nicht pauschal übertragen werden können.

Bei der Analyse des Wissensgehalts eines Wissensgraphen kann beispielsweise die Anzahl der Prädikate als Maß dienen. Ein höheres Prädikatenaufkommen in einem Wissensgraphen könnte den Eindruck vermitteln, dass dort mehr Wissen vorhanden ist. Allerdings ist es entscheidend, den Prozess der Dateneingabe in den Wissensgraphen zu berücksichtigen. DBpedia extrahiert seine Daten aus den Infoboxen von Wikipedia<sup>1</sup>. Wikipedia wiederum basiert auf einer umfangreichen Wiederverwendung

---

<sup>1</sup><http://dev.dbpedia.org/Mappings>

---

bestehender Vorlagen (Templates)<sup>2</sup>. Dieses Phänomen lässt vermuten, dass im Laufe der Zeit weniger Prädikate hinzugefügt werden und verstärkt auf die Wiederverwendung von Wissen im Wissensgraphen gesetzt wird. Das ORKG bietet zwar ein Template-System an, jedoch ist dessen Verwendung nicht verpflichtend. Dies liegt daran, dass es schwierig ist, universelle Vorlagen für die verschiedenen Fachgebiete zu entwickeln. Dadurch kann der ORKG als weniger restriktiv angesehen werden, was die Dateneingabe betrifft. Dies ermöglicht es den Nutzenden, ihr fachspezifisches Wissen frei einzupflegen. Allerdings führt dies auch dazu, dass weniger Wiederverwendung stattfindet und Inhalte neu erstellt werden müssen, um ihre Bedeutung präzise zu beschreiben.

Ein Beispiel ist die unterschiedliche Strukturierung der Klassen in DBpedia im Vergleich zum ORKG. In DBpedia gibt es eine begrenzte Anzahl von Hauptklassen wie Personen, Orte oder Organisationen, die wiederum eine Vielzahl von Instanzen umfassen. Im Gegensatz dazu weist der ORKG eine vielfältigere Klassenhierarchie auf, was allerdings bedeutet, dass weniger Instanzen pro Klasse vorhanden sind. Dies erfordert eine detaillierte Untersuchung, um die Wissensgraphen in Bezug auf ihre Tiefe, ihren Kontext und ihre Struktur angemessen zu vergleichen.

Es existiert eine definierte Ebene, bis zu der eine Vergleichbarkeit von Wissensgraphen möglich ist, beispielsweise durch strukturelle Merkmale wie die Vernetztheit einzelner Knoten zur Untersuchung der Wiederverwendbarkeit. Diese Qualitätsmetriken [16, 24] setzen eine bestimmte Überschneidung in der Struktur voraus, um einen Vergleich der Wissensgraphen zu ermöglichen. Auf einer bestimmten Stufe erfordert dies jedoch eine Abwägung zwischen einem generischen und einem spezifischen Ansatz. Das Dashboard wurde am Beispiel des ORKG mit dem Hintergrund des modularen Vorgehens konzipiert, wobei der Fokus auf dem spezifischen Anwendungsfall und individuellen Bedürfnissen lag. Somit stellt es einen deutlichen Mehrwert für das Team dar und verdeutlicht zugleich das Potenzial jedoch auch die Grenzen der Modularisierung.

Die Idee eines Dashboards zur Messung der Datenqualität stieß auch bei Mitgliedern der DBpedia-Community auf Interesse, wie aus einem Beitrag im Forum hervorgeht<sup>3</sup>. Dies unterstreicht den Bedarf, einen Überblick über die Datenqualität von DBpedia zu erhalten. Gleichzeitig wird jedoch darauf hingewiesen, dass die Bewertung der Datenqualität ein *fitness for use*-Ansatz ist, der eine semantische Interpretation erfordert, die ausschließlich von Menschen vorgenommen werden kann und damit stark vom Kontext abhängt.

---

<sup>2</sup>[https://en.wikipedia.org/wiki/Wikipedia:Manual\\_of\\_Style/Infoboxes](https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Infoboxes)

<sup>3</sup><https://forum.dbpedia.org/t/data-quality-dashboard-gsoc2020/359>

# Kapitel 9

## Zusammenfassung

### 9.1 Fazit

Die vorliegende Herausforderung bestand darin, ein maßgeschneidertes Dashboard für das Kuratierungs- und Gemeinschaftsbildungsteam zu konzipieren, das ihre individuellen Bedürfnisse in der täglichen Arbeit effektiv unterstützt. Das Ziel war es, ein Instrument zu schaffen, das es dem Team ermöglicht, die Qualität ihres Wissensgraphen zu beurteilen, indem es spezifische Anforderungen und Anwendungsfälle berücksichtigt. In Anbetracht der Vielfalt an Aufgaben und Daten, mit denen das Team konfrontiert ist, war es entscheidend, eine Lösung zu finden, die nicht nur funktional für den spezifischen Anwendungsfall mit seinen konkreten Bedarfen ausgerichtet ist, sondern auch die Flexibilität und Modularität aufweist, um für den ORKG, zusätzlich ebenso auf andere Wissensgraphen leicht erweitert werden zu können. Es sollte eine Plattform bieten, die es ermöglicht, neue Komponenten und Funktionen problemlos einzubinden, um das Dashboard im Sinne eines Rahmenwerks an unterschiedliche Kontexte anzupassen.

Das Vorgehen bei der Entwicklung des Dashboards war geprägt von einem iterativen Prozess im Zuge der Aktionsforschung, der eng mit dem Team zusammenarbeitete. Durch intensive Feedbackschleifen wurden die wichtigsten Bedürfnisse identifiziert und in das Design integriert. Die Implementierung erfolgte in Form eines Minimal Viable Products (MVP), das die Kernfunktionen des Dashboards abbildete. Dies ermöglichte eine kontinuierliche Anpassung an die sich ändernden Anforderungen und eine enge Ausrichtung an den tatsächlichen Arbeitsabläufen.

Das Experiment mit dem entwickelten Dashboard zeigte vielversprechende Ergebnisse. Die zukünftigen Nutzenden bewerteten das Dashboard als einen deutlichen Mehrwert für ihre Arbeit. Das Dashboard eröffnete neue Wege für die Arbeitsweise

des Teams, indem es Indikatoren bereitstellte, die auf die Qualität hinwiesen. Die Implementierung des MVPs war ein wichtiger Schritt in Richtung der angestrebten Lösung. Es demonstrierte das Potenzial eines maßgeschneiderten Dashboards für das Team und bestätigte die Relevanz einer engen Zusammenarbeit zwischen Entwickler:innen und Endbenutzer:innen.

Neben den positiven Ergebnissen stieß das Projekt an eine Grenze bei der Umsetzung des modularen Ansatzes. Die Balance zwischen generischen und spezifischen Lösungen erwies sich als Herausforderung. Die Vielfalt der Wissensgraphen und ihrer unterschiedlichen Strukturen unterstreicht die Notwendigkeit weiterer Forschung. Diese Erfahrung verdeutlicht die Komplexität bei der Entwicklung modularer Systeme, insbesondere in einem Umfeld mit spezifischen Anforderungen wie dem Kuratierungs- und Gemeinschaftsbildungsteam. Es zeigt die Dringlichkeit zusätzlicher Forschung und Innovation, um flexible Lösungen zu gestalten, die gleichermaßen anpassungsfähig und benutzerfreundlich sind.

## 9.2 Ausblick

Die vorliegende Arbeit hat deutlich gemacht, dass durch die Entwicklung des Minimal Viable Products (MVP) klargestellt wurde, welche strukturellen Anpassungen im Open Research Knowledge Graph (ORKG) erforderlich sind, um ihn besser mit den Anforderungen der Kuratierungsarbeit in Einklang zu bringen. Es wurde erkannt, dass viele Anfragen, die das Dashboard generiert, effizienter ins Backend des ORKG übertragen werden müssen, um die Performance zu steigern. Des Weiteren könnte die Implementierung eines Monitorings in Betracht gezogen werden, um die Qualität des ORKG im Laufe der Zeit zu verfolgen. Ähnlich wie bei DBpedia, wo regelmäßige Dumps ermöglichen, Änderungen nachzuvollziehen, könnte ein vergleichbares Verfahren angewendet werden, um die Entwicklung und Aktualisierungen des ORKG zu dokumentieren. Dies würde eine systematische Überwachung der Datenqualität und -integrität ermöglichen und die Transparenz über die Entwicklungen im ORKG verbessern.

Der Aktionsforschungsansatz hat gezeigt, dass iterative Feedbackschleifen kurzfristige Vorteile bieten, da das Team direkten Einfluss auf die Entwicklung des Dashboards nehmen konnte. Gleichzeitig wurde jedoch auch deutlich, dass das Dashboard einen potenziell langfristigen Einfluss auf die Architektur und die Funktionalität des ORKGs haben kann, indem es neue Werkzeuge bereitstellt und Optimierungen für bestehende Prozesse ermöglicht. Diese Wechselwirkung zwischen Praxis und Forschung ist vielversprechend und legt den Grundstein für weitere Entwicklungen in diesem Bereich.

Zukünftige Forschungsarbeiten könnten sich darauf konzentrieren, die Analyse von Wissensgraphen durch die Untersuchung ihrer Vergleichbarkeit zu vertiefen. Dies könnte durch die Entwicklung von Metriken erfolgen, die über rein strukturelle Aspekte hinausgehen und eine Schnittmenge von deckungsgleichen Metriken zwischen verschiedenen Wissensgraphen identifizieren. Es ist wichtig, die spezifischen Schwerpunkte bei der Erstellung von Wissensgraphen zu verstehen und zu analysieren, sowie ihre Bedeutung für den Kuratierungsprozess zu untersuchen.

Die Entwicklung solcher Metriken erfordert zunächst eine gründliche Untersuchung ausgewählter Graphen unter Berücksichtigung ihres jeweiligen Kontextes. Anschließend könnten Interviewstudien mit Verantwortlichen durchgeführt werden, um Einblicke in die Kuratierung ihrer Wissensgraphen zu erhalten und mögliche Überschneidungen mit bestehenden Systemen wie dem Open Research Knowledge Graph (ORKG) zu identifizieren. Diese Erkenntnisse könnten dann genutzt werden, um die Implementierung des Dashboards in ein Framework umzubauen, das es ermöglicht, verschiedene Wissensgraphen nahtlos zu integrieren.

Um die Modularität des Dashboards weiter zu steigern und den Nutzer:innen mehr Kontrolle zu geben, könnte die Integration eines visuellen SPARQL-Abfragetools in Betracht gezogen werden. Ein Beispiel hierfür ist Sparnatural<sup>1</sup>, ein visueller Baukasten für verschiedene Wissensgraphen und ihre Endpunkte, der Benutzer:innen Vorschläge macht, sie durch den Abfrageprozess führt und schließlich die SPARQL-Abfrage generiert, ohne dass spezifisches Wissen über die Graphenstruktur erforderlich ist. Darüber hinaus könnte die Einbettung von Redash<sup>2</sup> in Betracht gezogen werden, um die Erstellung relevanter Diagramme direkt innerhalb der Anwendung zu ermöglichen und so ein individuelles Dashboard zusammen zu stellen.

Diese Arbeit legt den Grundstein für weitere Entwicklungen und Forschungen im Bereich der Wissensgraph-Kuratierung, indem sie praxisnahe Ansätze und innovative Werkzeuge vorstellt, die vielversprechende Potenziale für weitere Fortschritte aufzeigen.

---

<sup>1</sup><https://sparnatural.eu/index.html>

<sup>2</sup><https://redash.io/>



# Anhang A

## Screenshots des MVP - SciKG Dashboard

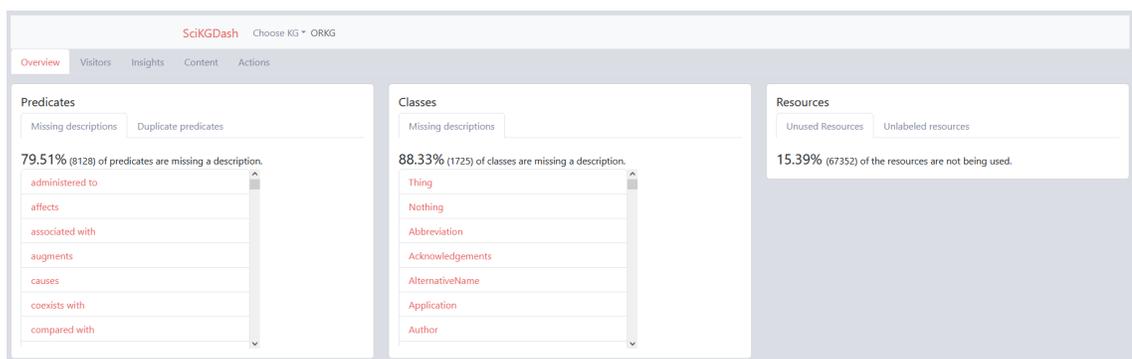


Abbildung A.1: Overview-Reiter

Predicates

Missing descriptions Duplicate predicates

Predicates without description

There are 185 duplicate predicates.

Label	Occurrences
Phenotypic plasticity form	115
Material	62
Assay measurement type	45
Has assay footprint	45
Has assay format	45
Has assay method	45

Abbildung A.2: Duplikate

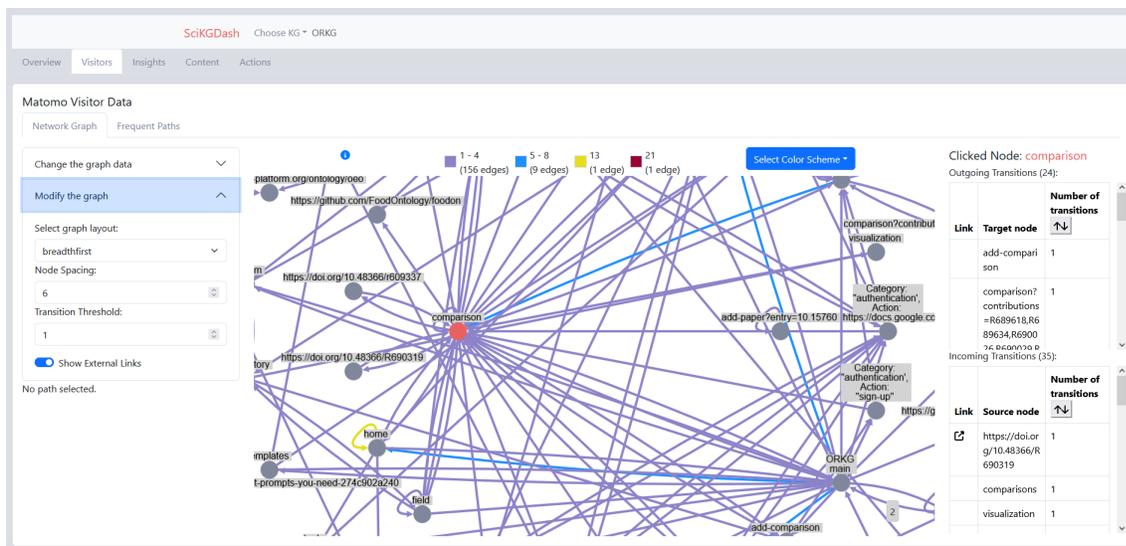


Abbildung A.3: Matomo Netzwerkgraph

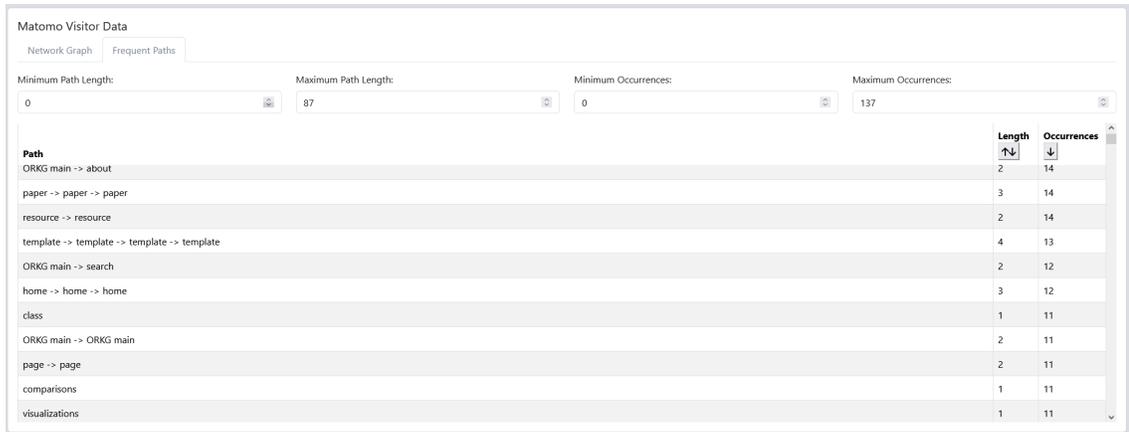


Abbildung A.4: Matomo Besucherpfade

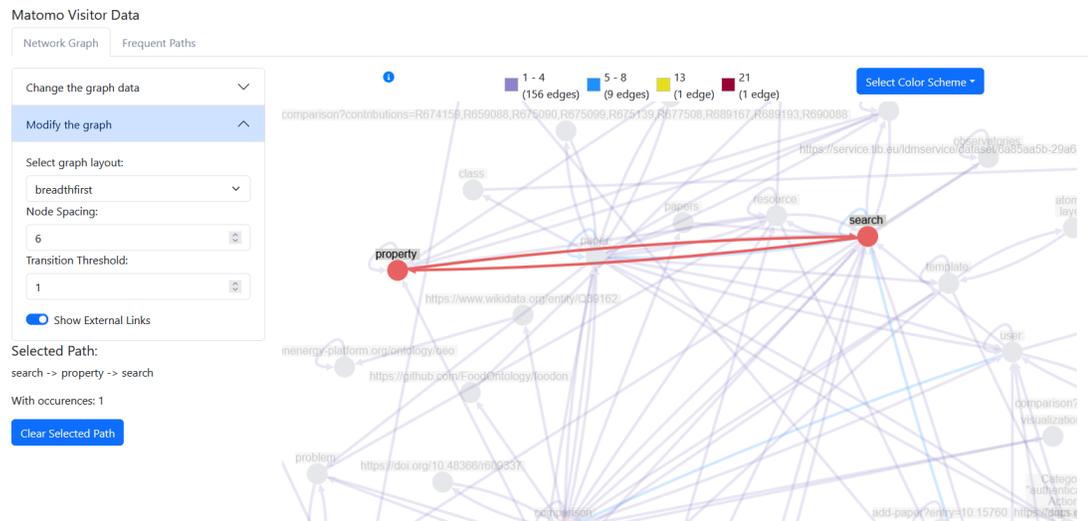


Abbildung A.5: Matomo hervorgehobener Pfad

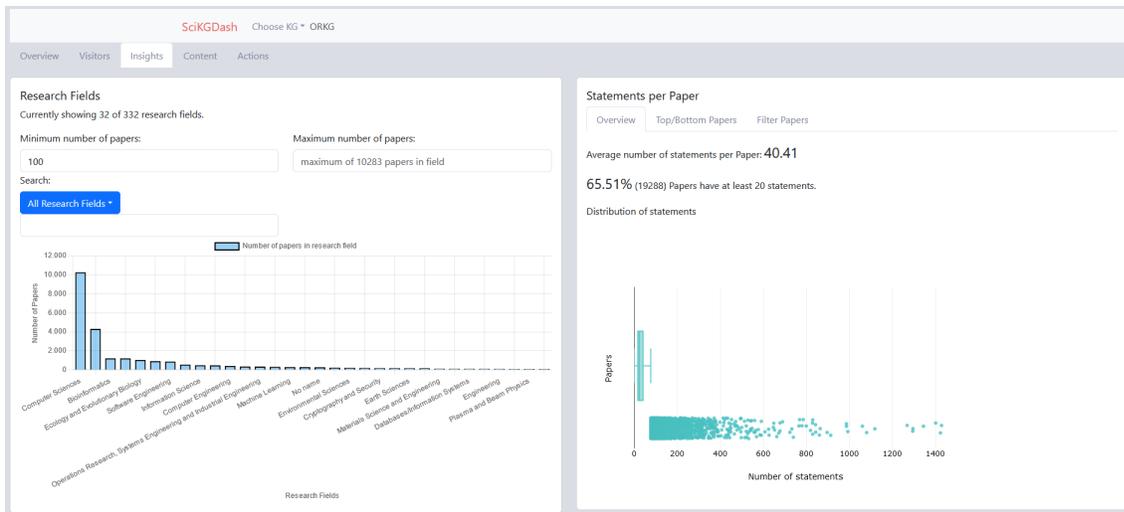


Abbildung A.6: Insights-Reiter

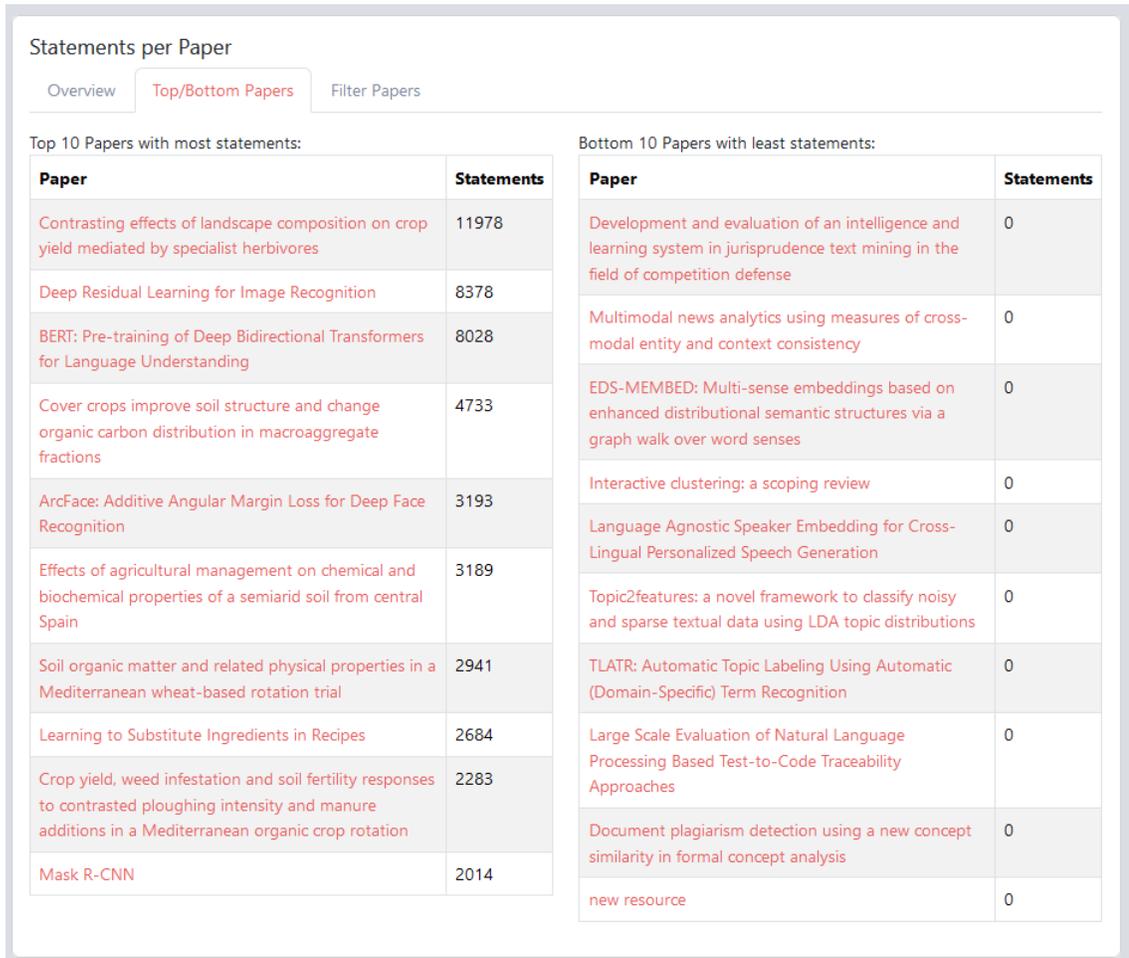


Abbildung A.7: Top und Bottom Papers

**Statements per Paper**

Overview   Top/Bottom Papers   **Filter Papers**

Showing 28.17% (8294) Papers.

Minimum number of statements:    Maximum number of statements:

Paper	Statements
The scientific events ontology of the OpenResearch.org curation platform	21
Towards a Knowledge Graph for Science	26
Investigating Correlations of Inter-coder Agreement and Machine Annotation Performance for Historical Video Data	27
Applying machine learning tools on web vacancies for labour market and skill analysis	25
A two-stage short-term load forecasting approach using temperature daily profiles estimation	26
Applying machine learning tools on web vacancies for labour market and skill analysis	26

Abbildung A.8: Paper Filter

SciKGDash Choose KG - ORKG

Overview Visitors Insights **Content** Actions

**Comparisons**  
Here you can see statistics about the comparison modeling. All comparisons: 1516.

Comparison label	Number of properties	Number of contributions	Number of all cells	Number of empty cells	Percentage of empty cells
Summary of the academic software for facial detection	6	5	30	0	0.00%
Comparison of SemTab 2022 datasets	11	14	154	52	33.77%
A Catalog of DeepMind's LLMs including their seminal Chinchilla model	19	7	133	12	9.02%
NLP Datasets for Scientific Concept and Relation Extraction from Biomedicine Scholarly Articles	18	16	288	118	40.97%
Overview of countries with a middle income trap	6	10	60	0	0.00%
Ontologies for smart city: levels and key classes ver. 2	18	6	108	10	9.26%
Comparison of unequal-sized clustering protocols	14	14	196	1	0.51%
Comparative Analysis of ALD/ALE-Deposited Films Across Varied Process Parameters and Material Properties	11	7	77	8	10.39%
Comparison of Scholarly Identifier Systems	15	4	60	27	45.00%
Teaching routes in classes with diverse students.	17	5	85	16	18.82%
Types of nanothermometers depending on the read-out signal	9	6	54	8	14.81%

**Templates**  
Here you can see who created which template.

Created by	Templates created
ORKG internal	32
Kheir Eddine Farfar	29
Gábor	1
Golsa	2
Anette	15
Jennifer D'Souza	15
Kamel	4
zhao zhao	1
Sören Auer	1

Abbildung A.9: Content-Reiter

**Templates**  
Here you can see who created which template.

Created by	Templates created
Oliver Karras	58
Markus Stocker	42
Azanzi Jiomekong	39
Anna Zieleniewska	34
ORKG internal	32
Kheir Eddine Farfar	29
Felix Wernlein	21
Karim Shalaby	19
Jan Göpfert	18

Templates by **Oliver Karras** (58)

Template Label	Created at
software engineering task	19-01-2024
SE task	19-01-2024
LLM4SE	19-01-2024
Contribution	03-08-2023
amount template	03-07-2023
EEA data template	08-05-2023
EEA comparison	05-05-2023
Trend template	05-05-2023
Measurement value Template	04-05-2023

Abbildung A.10: Templates

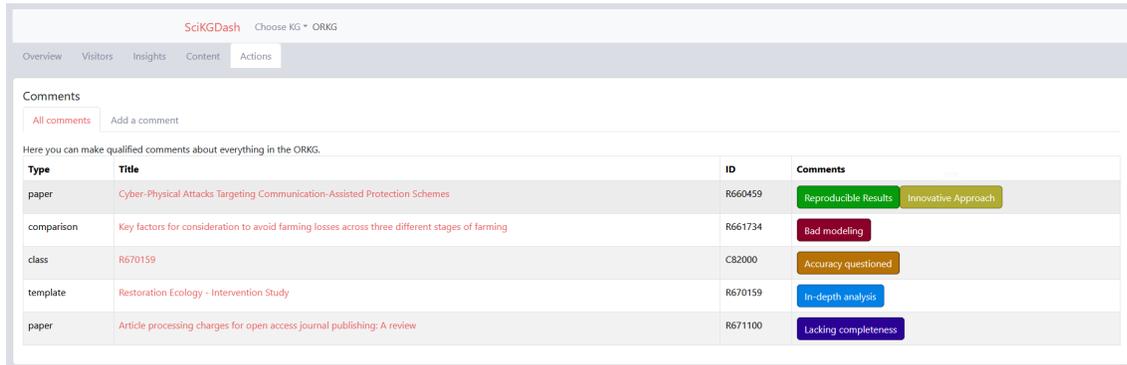


Abbildung A.11: Actions-Reiter

Comments

All comments Add a comment

URL

Enter URL of a ORKG page

Type of comment

Accuracy questioned

Description

Add the comment

Abbildung A.12: Kommentar hinzufügen

SciKGDash Choose KG • DBpedia

Properties 98.77 % <small>of the properties are missing a description.</small>	Classes 58.88 % <small>of the classes are missing a description.</small>
--	--

Abbildung A.13: DBpedia

**Anhang B**

**Begleitdokument Experiment**

**Experiment for Master Thesis "SciKGDash" - Lena John**

Thank you for your participation! Please first fill in the fields for demographic information.  
The data will be anonymized.

Name: \_\_\_\_\_

Age: \_\_\_\_\_

No answer

Job description: \_\_\_\_\_

No answer

Working at ORKG since: \_\_\_\_\_

No answer

To which team do you belong?

Curation & Community Building Team

General ORKG Team

External

In how many of the previous evaluations of the dashboard with the Curation & Community Building Team were you present? \_\_\_\_\_

- I understand, that my data is only stored for statistical purposes and is processed anonymously.
- I hereby consent, that the screen and sound is recorded while i participate. The data is not published and only necessary for analytic purposes.
- I understand, that I can always withdraw my participation and my data gets deleted.
- I confirm, that I voluntarily take part in the experiment.

Date, Signature: \_\_\_\_\_

You have 4 tasks in total for the sections of the dashboard. Please take your time and read the instructions carefully. Please feel encouraged to express your thoughts at any time and don't hesitate to ask questions. Your input and feedback are highly valued throughout the experiment.

# 1 Overview

Navigate to the Overview tab.

## Duplicate Predicates

You are in the role of a curator. Your task is to merge redundant Predicates as this will improve the overall quality of the knowledge graph. Start with merging Predicates with the fewest duplicates. To avoid merging Predicates that have the same label but describe different concepts, your focus will be on Predicates that lack a description.

1. Identify the Predicate with the least amount of duplicates which is lacking a description.
2. Open the respective pages in the ORKG.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I finished the task fast.	<input type="radio"/>				
I finished the task successfully.	<input type="radio"/>				

# 2 Visitors

Navigate to the Visitors tab.

You are in the role of a community building expert. You specialized in user behaviour analysis and your tasks involve studying visitor data within the ORKG. Focusing on a specific time span (March 6, 2024 to March 7, 2024), you aim to identify the most traversed paths and areas of interest among visitors. By evaluating this observational data, you can provide suggestions to improve the overall structure of the ORKG for optimal user experience and community engagement.

## Network graph

1. You want to see visitor statistics from March 6, 2024 to March 7, 2024. Set the respective dates and fetch the data.
2. Modify the graph by changing the parameters on the left side according to your approval for best visualization.
3. Identify the most frequently used edge in the network.
4. For the target node of this edge: Identify to which node most of the visitors would traverse after this.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I finished the task fast.	<input type="radio"/>				
I finished the task successfully.	<input type="radio"/>				

## Frequent Paths

Now you want to enhance your understanding of visitor navigation patterns. With a focus on longer paths, your aim is to identify the most frequently used route within the specified time span.

1. Identify the most used path of minimum length 3. Determine the number of occurrences.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I finished the task fast.	<input type="radio"/>				
I finished the task successfully.	<input type="radio"/>				

### 3 Insights

Navigate to the Insights tab.

#### Statements per Paper

You are in the role of a curator. Your job is to keep track of the quality and comprehensiveness of the ORKG's content. To ensure that all Papers meet the established standards of completeness, your task is to identify Papers with minimal Statements.

1. Identify the Paper with the least amount of Statements and open it in the ORKG.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I finished the task fast.	<input type="radio"/>				
I finished the task successfully.	<input type="radio"/>				

### 4 Actions

Navigate to the Actions tab.

You are in the role of a curator. You are responsible for maintaining the quality and relevance of resources. By pointing at negative aspects of improvement as well as positive examples of modeling, this practice maintains principles of transparency and responsibility. Your goal is to provide thoughtful feedback on a specific resource.

1. Make a comment for the Paper with the least amount of Statements. Choose the type of comment which seems applicable to you and add a description of your choice.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I finished the task fast.	<input type="radio"/>				
I finished the task successfully.	<input type="radio"/>				

Thank you for taking your time with the tasks! I would greatly appreciate it if you could take a moment to provide some concluding thoughts and feedback.

## User Experience Questionnaire

Please fill in the following table to rate your experience with the dashboard.

In my opinion, the dashboard is ...

obstructive	<input type="radio"/>	supportive						
complicated	<input type="radio"/>	easy						
inefficient	<input type="radio"/>	efficient						
confusing	<input type="radio"/>	clear						
boring	<input type="radio"/>	exciting						
not interesting	<input type="radio"/>	interesting						
conventional	<input type="radio"/>	inventive						
usual	<input type="radio"/>	leading edge						

## Addressed needs

Please fill in the following table to rate if the needs of the Curation & Community Building Team are addressed by the dashboard.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The dashboard allows the tracking of visitor paths.	<input type="radio"/>				
The dashboard makes it possible to indicate if a resource is inaccurate or incomplete.	<input type="radio"/>				
The dashboard provides tools for monitoring the quality of the ORKG.	<input type="radio"/>				
The dashboard provides an overview of areas for improvement.	<input type="radio"/>				
The dashboard lists how active the internal (ORKG members) and external users are.	<input type="radio"/>				

You have now the opportunity to browse through the dashboard if you like to and give me some feedback.

## Feedback

Do you have any feedback? What are suggestions for improvement? Your input is highly appreciated.

# Literatur

- [1] Albin Ahmeti u. a. “Crowdsourcing Updates of Large Knowledge Graphs”. In: *HumL@ISWC* (2018).
- [2] Yahia Atig, Ahmed Zahaf und Nadri Khiati. “Temporal Dimensions of Quality in Knowledge Graph Evolution: A Comprehensive Review”. In: *International Information and Engineering Technology Association* (2023). DOI: <https://doi.org/10.18280/isi.280428>.
- [3] Sören Auer. “Towards an Open Research Knowledge Graph”. In: (2018).
- [4] Sören Auer u. a. “Improving Access to Scientific Literature with Knowledge Graphs”. In: *Bibliothek Forschung und Praxis* (2020). DOI: <https://doi.org/10.1515/bfp-2020-2042>.
- [5] Sören Auer u. a. *Open Research Knowledge Graph*. Cuvillier Verlag, 2024.
- [6] Mrs. Rupali M. Chopade und Mr. Nikhil S. Dhavase. “Agile Software Development: Positive and Negative User Stories”. In: *International Conference for Convergence in Technology* (2017).
- [7] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.
- [8] Mike Cohn. *User Stories Applied: For Agile Software Development*. Addison Wesley Professional, 2004.
- [9] Danilo Dessì u. a. “Generating knowledge graphs by employing Natural Language Processing and Machine Learning techniques within the scholarly domain”. In: *Future Generation Computer Systems* 116 (2021). URL: <https://www.frontiersin.org/articles/10.3389/frma.2022.934930/full>.
- [10] Yadolah Dodge. “Spearman Rank Correlation Coefficient”. In: *The Concise Encyclopedia of Statistics* (2008). DOI: [https://doi.org/10.1007/978-0-387-32833-1\\_379](https://doi.org/10.1007/978-0-387-32833-1_379).
- [11] Daniel Domingo-Fernández u. a. “COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology”. In: *Bioinformatics* (2021). DOI: [10.1093/bioinformatics/btaa834](https://doi.org/10.1093/bioinformatics/btaa834).
- [12] Faruk Lawal Ibrahim Dutsinma u. a. “A Systematic Review of Voice Assistant Usability: An ISO 9241–11 Approach”. In: *SN Computer Science* (2022). DOI: <https://doi.org/10.1007/s42979-022-01172-3>.
- [13] Lisa Ehrlinger und Wolfram Wöß. “Towards a definition of Knowledge Graphs”. In: *International Conference on Semantic Systems* (2016). URL: <https://ceur-ws.org/Vol-1695/paper4.pdf>.

- 
- [14] Michael Färber. “The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data”. In: *International Semantic Web Conference* (2019).
- [15] Michael Färber u. a. “Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO”. In: *Semantic Web 9* (2018). URL: <https://www.semantic-web-journal.net/system/files/swj1465.pdf>.
- [16] Dieter Fensel u. a. *Knowledge Graphs Methodology, Tools and Selected Use Cases*. Springer, 2020.
- [17] Stephen Few. *Information Dashboard Design: The Effective Visual Communication Of Data*. O'Reilly & Associates Inc, 2005.
- [18] Michael Fire und Carlos Guestrin. “Over-Optimization of Academic Publishing Metrics: Observing Goodhart’s Law in Action”. In: *Gigascience* (2019). DOI: 10.1093/gigascience/giz053.
- [19] Erik Frøkjær, Morten Hertzum und Kasper Hornbæk. “Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?” In: *Conference on Human Factors in Computing Systems* (2000). DOI: <https://doi.org/10.1145/332040.332455>.
- [20] Miguel A. García-Pérez. “Use and misuse of corrections for multiple testing”. In: *Methods in Psychology* 8 (2023).
- [21] *History of the ORKG*. Zuletzt besucht: 28.10.2023. URL: <https://orkg.org/about/11/History>.
- [22] Aidan Hogan u. a. “Knowledge Graphs”. In: *ACM Computing Surveys* 54 (2021). URL: <https://doi.org/10.1145/3447772>.
- [23] Tasha Hollingsed und David G. Novick. “Usability Inspection Methods after 15 Years of Research and Practice”. In: *Proceedings of the 25th annual ACM international conference on Design of communication* (2007).
- [24] Elwin Huaman. “Steps to Knowledge Graphs Quality Assessment”. In: (2022). DOI: <https://doi.org/10.48550/arXiv.2208.07779>.
- [25] Elwin Huaman und Dieter Fensel. “Knowledge Graph Curation: A Practical Framework”. In: *IJCKG ’21: Proceedings of the 10th International Joint Conference on Knowledge Graphs* (2021). URL: <https://doi.org/10.48550/arXiv.2208.08130>.
- [26] Mohamad Yaser Jaradeh u. a. “Triple Classification for Scholarly Knowledge Graph Completion”. In: *Knowledge Capture Conference* (2021).
- [27] Lena John. *SciKGDash Code*. Zenodo. 2024. DOI: <https://doi.org/10.5281/zenodo.11214242>.
- [28] Lena John. *SciKGDash Experiment Handout and Results [Data set]*. Zenodo. 2024. DOI: <https://doi.org/10.5281/zenodo.11214197>.
- [29] Oliver Karras u. a. “Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering”. In: *International Symposium on Empirical Software Engineering and Measurement* (2023).

- [30] Oliver Karras u. a. “Researcher or Crowd Member? Why not both! The Open Research Knowledge Graph for Applying and Communicating CrowdRE Research”. In: *IEEE 29th International Requirements Engineering Conference Workshops* (2021).
- [31] Franki Y.H. Kung, Navio Kwok und Douglas J. Brown. “Are Attention Check Questions a Threat to Scale Validity?” In: *Applied Psychology* (2017). DOI: <https://doi.org/10.1111/apps.12108>.
- [32] Valentina Lenarduzzi und Davide Taibi. “MVP Explained: A Systematic Mapping Study on the Definitions of Minimal Viable Product”. In: *Euromicro Conference on Software Engineering and Advanced Applications* 42 (2016).
- [33] Olga Lezhnina u. a. “A Scholarly Knowledge Graph-Powered Dashboard: Implementation and User Evaluation”. In: *Frontiers in Research Metrics and Analytics* 7 (2022).
- [34] Garm Lucassen u. a. “The Use and Effectiveness of User Stories in Practice”. In: *Requirements Engineering: Foundation for Software Quality* (2016).
- [35] Khalid Mahmood Malik u. a. “Automated domain-specific healthcare knowledge graph curation framework: Subarachnoid hemorrhage as phenotype”. In: *Expert Systems with Applications* 145 (2020). DOI: <https://doi.org/10.1016/j.eswa.2019.113120>.
- [36] Paolo Manghi u. a. “The OpenAIRE Research Graph Data Model”. In: *Zenodo* (2019).
- [37] Giulio Mori, Fabio PaternoÁ und Carmen Santoro. “CTTE: support for developing and analyzing task models for interactive system design”. In: *IEEE Transactions on Software Engineering* 28 (2002). DOI: [10.1109/TSE.2002.1027801](https://doi.org/10.1109/TSE.2002.1027801).
- [38] Heiko Paulheim. “Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods”. In: *Semantic Web* (2016).
- [39] Prabhat Pokharel und Pramesh Vaidya. “A Study of User Story in Practice”. In: *International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy* (2020).
- [40] Jonny B. Pornel und Giabelle A. Saldaña. “Four Common Misuses of the Likert Scale”. In: *Philippine Journal of Social Sciences and Humanities* (2013).
- [41] Jason Priem, Heather Piwowar und Richard Orr. “OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts”. In: (2022).
- [42] Denise Rey und Markus Neuhäuser. “Wilcoxon-Signed-Rank Test”. In: *International Encyclopedia of Statistical Science* (2014). DOI: [https://doi.org/10.1007/978-3-642-04898-2\\_616](https://doi.org/10.1007/978-3-642-04898-2_616).
- [43] Amanda Ross und Victor L. Willson. “One-Sample T-Test”. In: *Basic and Advanced Statistical Tests* (2017). DOI: [https://doi.org/10.1007/978-94-6351-086-8\\_2](https://doi.org/10.1007/978-94-6351-086-8_2).
- [44] Per Runeson, Emelie Engström und Margaret-Anne Storey. “The Design Science Paradigm as a Frame for Empirical Software Engineering”. In: *Contemporary Empirical Methods in Software Engineering* (2020).
- [45] Kara Schatz u. a. “Workflow for Domain- and Task-Sensitive Curation of Knowledge Graphs, with Use Case of DRKG”. In: (2022). DOI: [10.1109/BigData55660.2022.10020536](https://doi.org/10.1109/BigData55660.2022.10020536).

- 
- [46] David Schindler, Benjamin Zopilko und Frank Krüger. “Investigating Software Usage in the Social Sciences: A Knowledge Graph Approach”. In: *The Semantic Web* (2020).
- [47] Martin Schrepp, Andreas Hinderks und Jörg Thomaschewski. “Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)”. In: *International Journal of Interactive Multimedia and Artificial Intelligence* (2017). DOI: [10.9781/ijimai.2017.09.001](https://doi.org/10.9781/ijimai.2017.09.001).
- [48] Sumin Seo u. a. “Structural Quality Metrics to Evaluate Knowledge Graph Quality”. In: (2022). DOI: <https://doi.org/10.48550/arXiv.2211.10011>.
- [49] Mahsa Shamsabadi, Jennifer D’Souza und Sören Auer. “Large Language Models for Scientific Information Extraction: An Empirical Study for Virology”. In: *European Chapter of the Association for Computational Linguistics* (2024). URL: <https://doi.org/10.48550/arXiv.2401.10040>.
- [50] S. S. Shapiro und M. B. Wilk. “An Analysis of Variance Test for Normality (Complete Samples)”. In: (1965). DOI: <https://doi.org/10.2307/2333709>.
- [51] Maarten W. van Someren, Yvonne F. Barnard und Jacobijn A.C. Sandberg. *The think aloud method: a practical approach to modelling cognitive processes*. Academic Press, 1994.
- [52] Herbert van de Sompel und Carl Lagoze. “All Aboard: Toward a Machine-Friendly Scholarly Communication System”. In: *The Fourth Paradigm* (2009).
- [53] Miroslaw Staron. *Action Research in Software Engineering*. Springer, 2020.
- [54] Markus Stocker u. a. “SKG4EOSC - Scholarly Knowledge Graphs for EOSC: Establishing a backbone of knowledge graphs for FAIR Scholarly Information in EOSC”. In: *Research Ideas and Outcomes* 8 (2022). URL: <https://doi.org/10.3897/rio.8.e83789>.
- [55] Michael Uschold. “Where Are the Semantics in the Semantic Web?” In: *AI Magazine* 24 (2003). DOI: <https://doi.org/10.1609/aimag.v24i3.1716>.
- [56] Xiangyu Wang u. a. “Knowledge graph quality control: A survey”. In: *Fundamental Research* (2021). DOI: [10.1016/j.fmre.2021.09.003](https://doi.org/10.1016/j.fmre.2021.09.003).
- [57] Claes Wohlin u. a. *Experimentation in Software Engineering*. Springer, 2012.