



# netidee

PROJEKTE

## CrOSSD

Towards a Critical Open-Source Software Database

Endbericht | Call 17 | Projekt ID 6252

Lizenz CC BY-SA

# Inhalt

Einleitung.....	3
Projektbeschreibung.....	3
Verlauf der Arbeitspakete.....	5
Arbeitspaket 1: Detailplanung und Formales am Projektstart.....	5
Arbeitspaket 2: Projektmanagement, Outreach und Dissemination.....	5
Arbeitspaket 3: Metriken – Recherche, Definition und Operationalisierung.....	6
Arbeitspaket 4: Implementierung und Crawling.....	6
Arbeitspaket 5: Plattform, Website und Integration.....	8
Webinterface.....	8
API-Schnittstelle.....	11
Arbeitspaket 6: Dokumentation und Formales am Projektende.....	11
Umsetzung Förderauflagen.....	11
Liste Projektergebnisse.....	12
Verwertung der Projektergebnisse in der Praxis.....	13
Öffentlichkeitsarbeit/Vernetzung.....	13
Eigene Projektwebsite.....	13
Geplante Aktivitäten nach netidee-Projektende.....	14
Anregungen für Weiterentwicklungen durch Dritte.....	14

## Einleitung

Open-Source-Software treibt eine breite Palette von Softwareanwendungen an und bildet die Grundlage für viele Dienste im Internet. Trotz dieser wichtigen Rolle stehen OSS-Projekte oft vor Herausforderungen in Bezug auf Nachhaltigkeit und Wartung. Um diesem Problem zu begegnen, werden in CrOSSD kritische OSS-Projekte identifiziert und bewertet, um so einen umfassenden Überblick über deren „Gesundheit“ zu bieten. Dies geschieht mithilfe verschiedener geeigneter Metriken und automatisierter Analysen. Die verwendeten Metriken umfassen Stabilität, Widerstandsfähigkeit, Sicherheit und Compliance.

Das Projektziel ist der Aufbau einer Plattform, die eine große Anzahl von OSS-Projekten überwacht und bewertet. Dies ermöglicht es verschiedenen Interessengruppen, wie den Projekteigentümer:innen selbst, der gesamten OSS-Community, oder z. B. anderen Entwickler:innen, Unternehmen und Fördergebern, informierte Entscheidungen zu treffen.

In diesem Endbericht präsentieren wir die Forschungsdesign-Methodik und die Implementierungsergebnisse von CrOSSD: Recherche und Definition von Metriken (siehe Arbeitspaket 3), die Implementierung von Metriken (Arbeitspaket 4), sowie eine detaillierte Beschreibung der Plattform den Monitoring-Funktionalitäten (Arbeitspaket 5).

Obwohl es bereits bestehende Forschungsprojekte und Initiativen gibt, die ähnliche Ziele verfolgen und Metriken sowie bewährte Praktiken anbieten (einige sogar mit Bewertungen ähnlich unserer Grundidee), hebt sich CrOSSD davon ab, indem es ein umfassenderes Verständnis für die Gesundheit von Projekten bieten soll. Darüber hinaus bietet keine der existierenden Initiativen kontinuierliche Überwachung und Reporting an. Um unsere Idee zu evaluieren, stellten wir die Projektidee bereits in einem frühen Stadium in der Community zur Diskussion, indem wir unseren derzeitigen Aufbau von CrOSSD auf wissenschaftlichen Konferenzen vorstellten (siehe Arbeitspaket 2: Projektmanagement, Outreach und Dissemination).

## Projektbeschreibung

Das Ziel von CrOSSD ist es, ein umfassendes Verständnis für die Gesundheit von Open-Source-Software-Projekten (OSS) zu schaffen und gleichzeitig eine Plattform bereitzustellen, die in der Lage ist, kontinuierlich die Stabilität und Qualität von OSS-Projekten zu überwachen und bewerten. Die Zielgruppe des Projekts umfasst verschiedene Interessengruppen, darunter:

1. **Projekteigentümer:innen:** Die Plattform bietet Projektbesitzer:innen einen Einblick in die Gesundheit und Leistung ihrer OSS-Projekte, wodurch sie fundierte Entscheidungen zur Weiterentwicklung und Wartung treffen können.
2. **OSS-Community:** Die breitere OSS-Community profitiert von transparenten Informationen über die Gesundheit von OSS-Projekten, die ihnen helfen, qualitativ hochwertige und stabile Softwarekomponenten auszuwählen und zu verwenden.
3. **Entwickler:innen:** Die Plattform bietet Entwickler:innen Einblicke in die Zuverlässigkeit und Sicherheit von OSS-Projekten, was ihnen bei der Auswahl von Bibliotheken und Frameworks für ihre eigenen Projekte hilft.
4. **Unternehmen:** Unternehmen können CrOSSD nutzen, um potenzielle Abhängigkeiten und Risiken in ihren Software-Stacks zu identifizieren und die Auswahl von OSS-Komponenten zu unterstützen, die ihren Anforderungen an Qualität und Sicherheit entsprechen.

CrOSSD liefert ein ganzheitliches System zur Überwachung und Bewertung von OSS-Projekten, das auf folgenden Säulen basiert:

1. *Forschungsdesign und Methodik:* Das Projekt umfasst eine gründliche Untersuchung der OSS-Gesundheit und -Qualität sowie die Definition und Operationalisierung von Metriken zur Bewertung dieser Aspekte.
2. *Implementierung und Crawling:* CrOSSD implementiert ein robustes System zur regelmäßigen Datenerfassung aus OSS-Repositorys und wendet definierte Metriken auf diese Daten an, um die Gesundheit der Projekte zu bewerten.
3. *Plattform und Website:* Die Plattform bietet ein benutzer:innenfreundliches Webinterface, über das Benutzer:innen die Metriken und Bewertungen der OSS-Projekte abrufen, vergleichen und analysieren können. Eine API-Schnittstelle ermöglicht den maschinellen Zugriff auf die Daten.

Die Ergebnisse von CrOSSD tragen dazu bei, die Transparenz und Verlässlichkeit von OSS-Projekten zu verbessern und unterstützen verschiedene Akteur:innen dabei, informierte Entscheidungen im Zusammenhang mit Open-Source-Software zu treffen.

Das netidee-Projekt soll dabei nur der Start unserer Bemühungen sein: Die CrOSSD-Plattform soll kontinuierlich verbessert werden, um Funktionalität, Reichweite und Relevanz zu erhöhen. Wir werden in Zukunft weiter eng mit der OSS-Community zusammenarbeiten, um Feedback zu den verwendeten Metriken zu sammeln und damit den Bedürfnissen der Nutzer:innen gerecht zu werden. Im nächsten Schritt werden wir an der Einbindung von weiteren Datenquellen und der Integration von fortgeschritteneren/automatisierten Analysetechniken zur Berechnung der Metriken arbeiten.

## Verlauf der Arbeitspakete

### Arbeitspaket 1: Detailplanung und Formales am Projektstart

Arbeitspaket 1 wurde bereits mit April 2023 erfolgreich abgeschlossen. Die folgenden Arbeiten wurden in diesem initialen Arbeitspaket durchgeführt:

- der Vertrag wurde unterschrieben,
- der Detailprojektplan wurde erstellt und abgenommen,
- eine detaillierte Liste der geplanten Projektergebnisse mit Lizenz und Ort der öffentlichen Bereitstellung wurde erstellt und abgenommen,
- die Projekt-Website ging in Betrieb und ein erster Blogeintrag wurde erstellt, und
- die erste Förderrate wurde beantragt.

### Arbeitspaket 2: Projektmanagement, Outreach und Dissemination

Das Projektmanagement verlief erwartungsgemäß und ohne größere Hindernisse, sodass hier keine Umplanungen oder Ähnliches notwendig waren.

Im Blog unter <https://www.netidee.at/crossd> haben wir plangemäß Beiträge veröffentlicht:

- Im Feber 2023 haben wir unser Projekt zunächst einmal vorgestellt, sowohl bzgl. der Motivation und grundlegenden Idee als auch bzgl. der Systemarchitektur und Konzeption unserer Infrastruktur: <https://www.netidee.at/crossd/introducing-crossd>
- Im Mai 2023 haben wir unser Projektteam vorgestellt. Zusätzlich zu den bereits bei der Antragstellung eingeplanten drei Projektmitarbeitern (Sebastian Neumaier, Lukas Daniel Klausner und Tobias Dam) sind auch zwei FHStP-Studierende (Matthias Kopeinig und Jacqueline Schmatz) in unser Team hinzugekommen, die ihre Abschlussarbeiten im Rahmen von CrOSSD geschrieben und sich mit Metriken und Analysen befasst haben (mehr dazu im nächsten Abschnitt): <https://www.netidee.at/crossd/our-team>
- Im August 2023 schließlich haben wir die Arbeit eines der zwei Student:innen vorgestellt. Matthias Kopeinig hat im Rahmen seiner Abschlussarbeit untersucht, welche Erkenntnisse eine Analyse der Firmenzugehörigkeiten von Contributors in einem Sample von 100 GitHub-Projekten offenbaren kann: <https://www.netidee.at/crossd/impact-companies-open-source>
- Im Frühjahr 2024 haben wir die finale Architektur sowie das Webinterface der CrOSSD Plattform vorgestellt.

Weiters wurden wissenschaftliche Publikationen zu den Projektergebnissen veröffentlicht:

- Tobias Dam und Sebastian Neumaier haben sich in „Towards Measuring Vulnerabilities and Exposures in Open-Source Packages“ mit Sicherheitsaspekten im Kontext unseres Projekts befasst. Das Paper wurde Anfang Mai von Sebastian Neumaier auf der 5th International Data Science Conference an der UWK in Krems präsentiert und erscheint demnächst in den

Konferenzproceedings (siehe <https://idsc.at/proceedings/>); ein Preprint ist unter <https://arxiv.org/abs/2206.14527> verfügbar.

- Tobias Dam, Lukas Daniel Klausner und Sebastian Neumaier haben weiters im Short Paper „Towards a Critical Open-Source Software Database“ das Projekt vorgestellt, um auch in der wissenschaftlichen Community erstes Networking zu betreiben. Das Paper wurde Anfang Mai von Tobias Dam auf der ACM Web Conference 2023 in Austin, Texas, vorgestellt und ist bereits in den Konferenzproceedings erschienen: <https://doi.org/10.1145/3543873.3587336>
- Eine dritte Publikation, die den Stand der Forschung zu quantitativen Metriken ausführlich analysiert, ist derzeit noch in Ausarbeitung. Eine Einreichung in einer wissenschaftlichen Konferenz ist in Vorbereitung.

### **Arbeitspaket 3: Metriken – Recherche, Definition und Operationalisierung**

In diesem Arbeitspaket haben wir uns gemäß unserem ursprünglichen Plan zunächst auf quantitative Metriken fokussiert.

In einer systematischen Literaturrecherche haben wir 26 relevante Papers gefunden, die sich mit OSS-Gesundheit oder -Qualität und Metriken oder Kriterien befassen. Wir haben zunächst die unterschiedlichen Gesundheits- und Qualitätsbegriffe sowie die Metriken und Analyseverfahren aus diesen Papers extrahiert und strukturiert aufgeschlüsselt.

Im nächsten Schritt haben wir die verschiedenen Konzeptionen von Gesundheit und Qualität gesichtet, geclustert und codiert, um so eine umfassende Taxonomie zentraler Aspekte von OSS-Gesundheit zu erhalten.

In der Folge haben wir aus den erfassten Metriken jene selektiert, die erstens gewissen Nützlichkeits- und Pragmatismusheuristiken genügen, insbesondere aber die zuvor erfassten Aspekte möglichst gut abdecken. Diese Metriken wurden in der Folge in unserem System implementiert und erfasst.

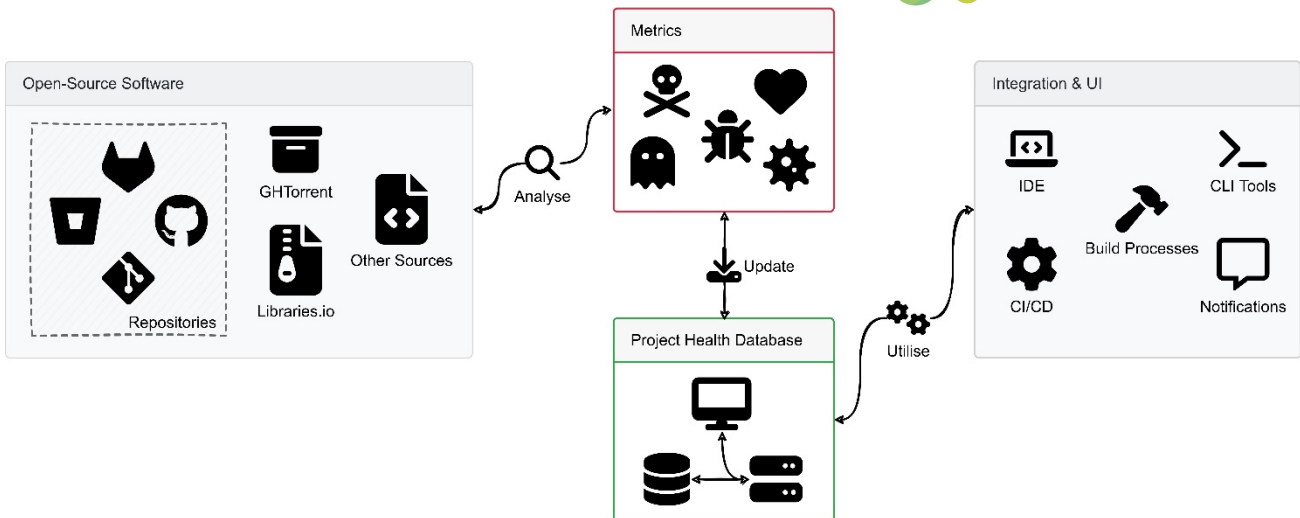
Die Resultate dieses Prozesses wurden zu einer wissenschaftlichen Publikation aufbereitet. Auf der Projektwebseite findet sich eine detaillierte Darstellung.<sup>1</sup>

### **Arbeitspaket 4: Implementierung und Crawling**

Im Rahmen der Planungen wurde ein abstraktes Architekturdiagramm erstellt:

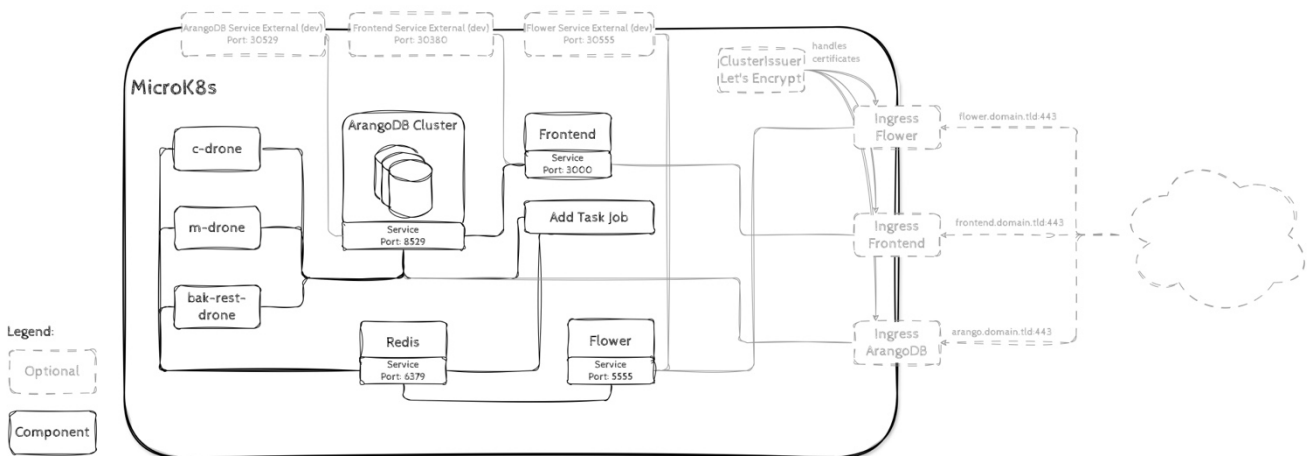
---

<sup>1</sup> <https://health.crossd.tech/doc>



Gemäß unserem ursprünglich geplanten Aufbau werden Informationen zu Open-Source-Repositoryys von verschiedenen Quellen abgerufen und gespeichert. Auf diese Informationen werden unsere Metriken angewandt und die Ergebnisse anschließend wiederum in unserer Datenbank gespeichert. Diese Vorgänge erfolgen regelmäßig, um laufend aktuelle Ergebnisse präsentieren zu können. Diese Ergebnisse sind anschließend über ein Web-Interface beziehungsweise über andere Integrationen abrufbar.

In diesem Arbeitspaket wurde die ursprüngliche Architektur konkretisiert und die einzelnen Komponenten implementiert. Die Architektur sieht nun wie folgt aus:



Unser System wurde aufgrund benötigter Skalierbarkeit und Portabilität als Kubernetes-Cluster konzipiert. Die einzelnen Komponenten, insbesondere die Crawler- und die Metrik-Container, können je nach Bedarf in beliebiger Anzahl eingesetzt werden. Weiters kann das System über mehrere Knoten (Rechner) verteilt werden. Das Sammeln der Informationen zu Repositoryys wird mit Hilfe einer Celery Task Queue verwaltet. Diese kümmert sich darum, die Scans regelmäßig auszuführen (und im Fehlerfall zu wiederholen) und weist die Aufgabe einem Crawler-Container zu. Es können beliebige Container für das Sammeln von Informationen eingesetzt werden. Aktuell existieren zwei verschiedene Typen: Type 1 stützt sich hauptsächlich auf die GitHub REST API und verwendet die Softwarebibliothek, die von Jacqueline Schmatz im Rahmen ihrer Abschlussarbeit programmiert wurde. Type 2 verwendet großteils die GraphQL-Schnittstelle von GitHub. Die gesammelten Daten werden in einer verteilten ArangoDB-Datenbank gespeichert.

Analog zu den verschiedenen Typen der Crawler können auch verschiedene Typen der Metrik-Berechnung eingesetzt werden. Die Ergebnisse der Metrik-Container werden wiederum in der Datenbank gespeichert. Das Web-Interface lädt die entsprechenden Metriken und stellt diese dann visuell aufbereitet dar. Das Web-Interface wird in AP5 entwickelt werden.

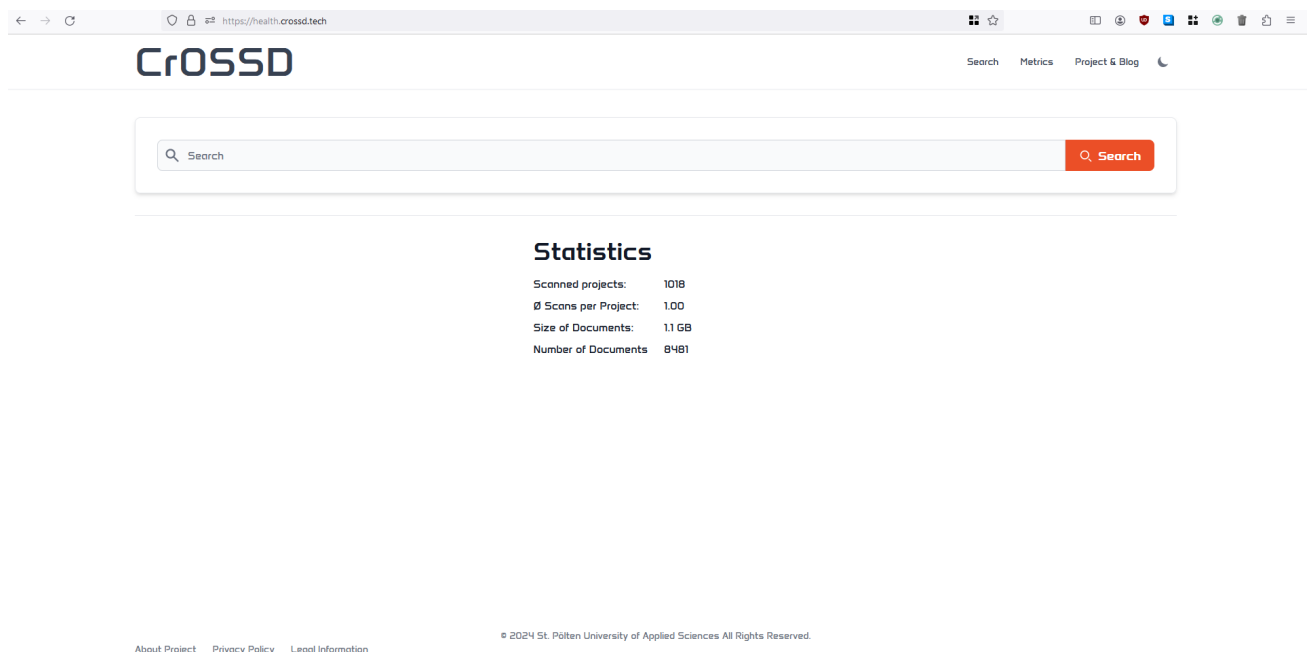
## Arbeitspaket 5: Plattform, Website und Integration

Um unsere Daten nutzbar zu machen, wurden in diesem Arbeitspaket das Webinterface sowie eine API entwickelt.

Webinterface

Das Webinterface wurde mithilfe von Svelte und SvelteKit erstellt, um möglichst einfach einen öffentlichen Zugang zu unseren Daten bieten.

Die Startseite bietet die Möglichkeit nach Projekten zu suchen und zeigt allgemeine Statistiken der indizierten Daten:



Die Suchfunktion liefert eine Liste an Projekten, welche in ihrem Bezeichner den Suchbegriff beinhalten:



# CROSSD

 Search

## Search results:

<b>kelseyhightower/kubernetes-the-hard-way</b> Last Crawled: Sun, 14 Apr 2024 12:21:03 GMT	Bootstrap Kubernetes the hard way. No scripts.
<b>kubernetes/kubernetes</b> Last Crawled: Sun, 14 Apr 2024 12:21:01 GMT	Production-Grade Container Scheduling and Management
<b>kubernetes/minikube</b> Last Crawled: Sun, 14 Apr 2024 12:21:05 GMT	Run Kubernetes locally

Nach einem Klick auf das gewünschte Projekt werden die berechneten Metriken des letzten Scanvorgangs angezeigt. Im rechten, oberen Bereich befindet sich eine Dropdownliste mit allen Scanvorgängen dieses Projekts. Es können so die berechneten Metriken zu unterschiedlichen Zeitpunkten dargestellt und verglichen werden:

← → ↻ <https://health.crossd.tech/project/kubernetes%2Fkubernetes> Search Metrics Project & Blog

# CROSSD

## kubernetes/kubernetes Sun, 14 Apr 2024 12:21:01 GMT

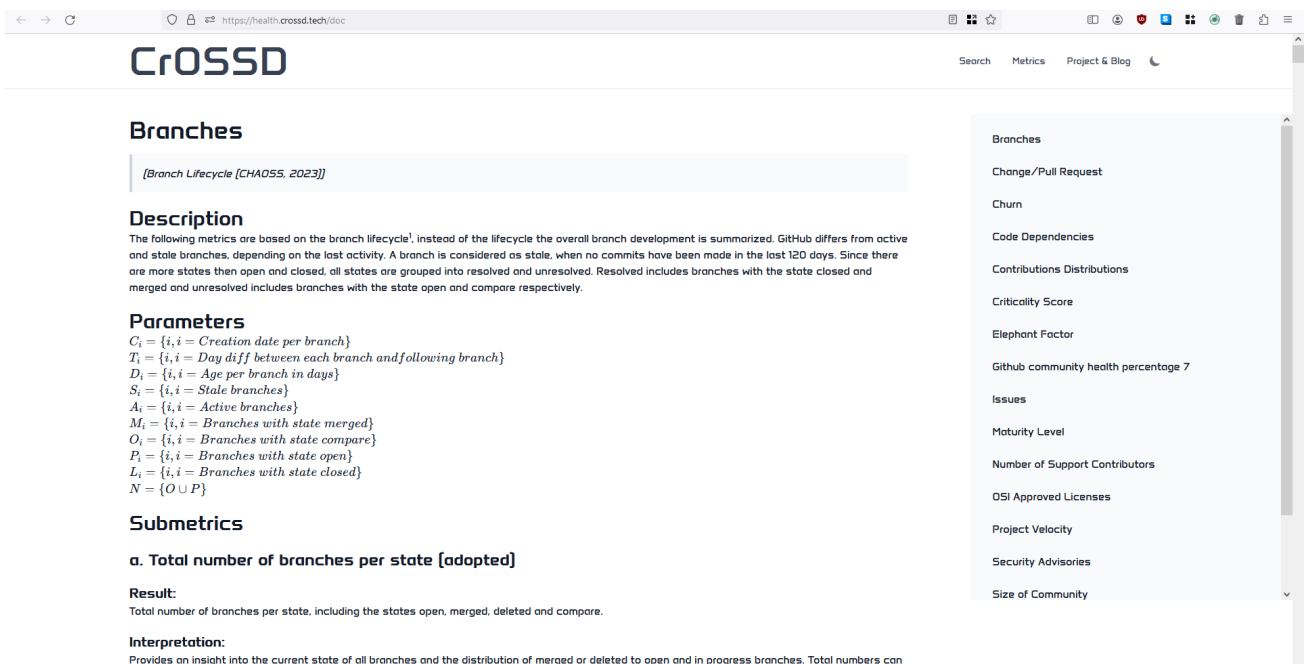
<b>Pull Requests (mean):</b> 2 wks, 19 hr, 43 min, 53 sec, 654.911 ms <b>Pull Requests (median):</b> 3 days, 2 hr, 1 min, 11 sec <b>Dependents:</b> 1 <b>Dependencies:</b> 2386 <b>Feature Requests:</b> 0 <b>Closed Feature Requests:</b> 0 <b>Security Policy:</b> <span style="color: green;">●</span> <b>Contribution Policy:</b> <span style="color: green;">●</span> <b>Collaboration Platform:</b> <span style="color: red;">●</span> <b>Fundable:</b> <span style="color: red;">●</span> <b>Workflows:</b> <span style="color: red;">●</span>	<b>Elephant factor:</b> 6 <b>Churn:</b> 55.52 <b>Community size:</b> 100 <b>Support contributors:</b> 100 <b>Maturity level:</b> 73.33 <b>OSI approved license:</b> <span style="color: blue;">👍</span> <b>Criticality score:</b> 81 <b>Support rate:</b> 2.36 <b>Code dependency upstream:</b> 410 <b>Code dependency downstream:</b> 6699 <b>Total forks:</b> 1689 <b>Average forks per week:</b> 65	<b>Pull requests [PR]</b> <b>Total PR:</b> 3317 <b>Average days for closing PR:</b> 28.30 <b>Ratio open PR:</b> 22.07 <b>Ratio closed PR:</b> 77.93 <b>Ratio merged PR:</b> 46.61
---	---	--

<b>Project velocity</b> <b>Total issues:</b> 5680 <b>Closed issues:</b> 3795	<b>Github community health</b> <b>Community Health score:</b> 100 <b>Custom Health score:</b> 87.5	<b>Issues</b> <b>Total issues:</b> 1773 <b>Open issues:</b> 1007	<b>Distribution of contributions</b> <b>Ratio of files tail:</b> 85.71 <b>Ratio of files dominant:</b> 14.29
--	--	--	--

Klickt man auf eine unterstrichene Bezeichnung der Metrik, wird die zugehörige Beschreibung und Definition der Metrik sowie ein Diagramm über die historischen Werte der Metrik über aller Scanvorgänge angezeigt:



Der Menüpunkt „Metrics“ in der Navigationsleiste führt zur Dokumentation aller Metriken:



Weiters gibt eine Projektbeschreibung und eine Anwender:innendokumentation unter „About Project“ in der Fußleiste im unteren Bereich:

# CrOSSD

## About CrOSSD

The Critical Open-Source Software Database [CrOSSD] is an open-source software [OSS] project meant to assist a variety of stakeholders to make informed decisions about which OSS projects to use, assist or support.

By collecting and continuously monitoring a growing catalogue of OSS projects according to a selection of curated metrics, CrOSSD gives a comprehensive overview of the state of the observed projects: How stable are they? How often are they updated? How many active contributors are there? Do they come from different organisations? The answers to these questions (and more) allow lay users and software developers to decide which projects they want to include in their own codebase and which projects might benefit from their involvement and assistance as well as funding agencies or companies to allocate funding or support to critically important OSS projects in need of help.

The current iteration of CrOSSD is based on a Kubernetes cluster consisting of various containers built upon Docker images. We utilise different Python crawlers to gather information about OSS repositories and calculate our set of metrics. Those results are stored inside an ArangoDB as JSON documents. The crawlers retrieve their tasks from a Celery task queue using Redis as a broker. The public interface for our cluster is web frontend built with Sveite and SveiteKit, which provides a search functionality and presents our metrics.

We are regularly crawling selected sets of projects and recording their metrics. Currently, we are limiting our scope to projects publicly listed on GitHub, but we are actively developing it and continuously improving it. If you have suggestions for desirable features or wish to contribute yourself, please let us know!

## API-Schnittstelle

Um die Daten auch maschinell zugänglich zu machen, wurden API-Schnittstellen eingerichtet. Es gibt die Möglichkeit, folgende Daten abzurufen:

- eine Liste aller in der Datenbank vorhandenen Projekte;
- eine Liste aller Zeitstempel der Scanvorgänge zu einem Projekt; und
- die Metriken eines bestimmten Scanvorgangs zu einem Projekt.

Eine genaue Beschreibung der API-Endpunkte befindet sich in der Github-Dokumentation:

<https://github.com/FH-CrOSSD/crossd?tab=readme-ov-file#api>

## Arbeitspaket 6: Dokumentation und Formales am Projektende

Die Dokumentation der Projektergebnisse wurde abgeschlossen, die Projekt-Website wurde aktualisiert und Projektergebnisse sind unter Angabe der entsprechenden Open-Source- bzw. Creative-Commons-Lizenz der Öffentlichkeit zur Verfügung gestellt worden:

- die Entwickler:innendokumentation ist auf Github hinterlegt;
- die Anwender:innendokumentation ist auf der Projekt-Website hinterlegt;
- der Katalog von OSS-Gesundheitsmetriken wurde auf der Projekt-Website sowie in einem wissenschaftlichen Paper publiziert; und
- der Software-Code wurde auf Github veröffentlicht.

## Umsetzung Förderauflagen

*(In der Fördervereinbarung sind keine Förderauflagen vorgesehen.)*

## Liste Projektergebnisse

1	Projektzwischenbericht	CC BY-SA 4.0	<a href="https://www.netidee.at/crossd">https://www.netidee.at/crossd</a>
2	Projektendbericht	CC BY-SA 4.0	<a href="https://www.netidee.at/crossd">https://www.netidee.at/crossd</a>
3	Entwickler_innen-Dokumentation des Projektergebnisses für andere Entwickler_innen ("Dritte"), die das Projektergebnis nach Projektende nutzen/weiterentwickeln wollen	CC BY-SA 4.0	<a href="https://github.com/FH-CrOSSD/crossd">https://github.com/FH-CrOSSD/crossd</a>
4	Anwender_innen-Dokumentation des Projektergebnisses für Anwender_innen, die das Projektergebnis nach Projektende nutzen wollen	CC BY-SA 4.0	<a href="https://health.crossd.tech/about">https://health.crossd.tech/about</a>
5	Veröffentlichungsfähiger Einseiter / Zusammenfassung	CC BY-SA 4.0	<a href="https://www.netidee.at/crossd">https://www.netidee.at/crossd</a>
6	Dokumentation Externkommunikation zur Erreichung Sichtbarkeit /Nachhaltigkeit (als Teil des Endberichtes)	CC BY-SA 4.0	<a href="https://www.netidee.at/crossd">https://www.netidee.at/crossd</a>
7	Bericht über Literatur zu OSS-Gesundheitsmetriken sowie Bericht und Definition von quantitativen und qualitativen Metriken, die in die Projektplattform aufgenommen wurden.	Metriken: CC BY-SA 4.0  Wiss. Publikation in Arbeit	<a href="https://health.crossd.tech/doc">https://health.crossd.tech/doc</a>
8	Die Metriken werden auf ausgewählten OSS-Repositorys angewendet und ausgewertet. Die Daten sind als Open Data – sowohl in maschinenverarbeitbarer als auch in für Menschen lesbarer Form via Web-UI – zugänglich.	CC BY-SA 4.0	<a href="https://health.crossd.tech">https://health.crossd.tech</a>  API Doc: <a href="https://github.com/FH-CrOSSD/crossd?tab=readme-ov-file#api">https://github.com/FH-CrOSSD/crossd?tab=readme-ov-file#api</a>

9	Software-Code, zur Berechnung der Metriken und zur Darstellung auf der Projektseite.	CC-BY-SA 4.0	<a href="https://github.com/FH-CrOSSD">https://github.com/FH-CrOSSD</a>
---	--	--------------	---

## Verwertung der Projektergebnisse in der Praxis

Die Projektergebnisse von CrOSSD werden bereits konkret in der Praxis verwertet, insbesondere durch die Anbahnung eines Folgeprojekts (siehe geplante Aktivitäten) und die Nutzung der Plattform durch Firmenpartner: Die Weitergabe der Projektergebnisse an die Konsortialpartner hat zur Anbahnung eines Folgeprojekts geführt. Dies zeigt das Interesse der Partner an den Erkenntnissen und der Plattform von CrOSSD und verdeutlicht deren Relevanz für die Praxis. Konsortialpartner sind u. a. auch Softwareentwicklungsunternehmen, die vor der Entscheidung stehen, welche Frameworks sie für ihre Anwendungen verwenden sollen. Indem sie verschiedene Open-Source-Frameworks anhand ihrer Gesundheitsmetriken vergleichen können, erhalten sie wertvolle Einblicke in die Qualität, Stabilität und Sicherheit der einzelnen Frameworks.

## Öffentlichkeitsarbeit/Vernetzung

Die Aktivitäten zur Öffentlichkeitsarbeit und Vernetzung sind teilweise bereits weiter oben in Arbeitspaket 2 beschrieben, hier aber noch einmal eine kurze Aufstellung:

- Blogposts im netidee-Blog;
- drei wissenschaftliche Publikationen; damit zusammenhängend
- zwei Präsentationen auf Konferenzen;
- Social-Media-Aktivitäten auf den Accounts der Projektmitarbeiter; sowie
- Berichte im Newsroom der FH St. Pölten und über die Austria Presse Agentur:
  - <https://www.fhstp.ac.at/de/newsroom/news/projektfoerderung-fuer-netidee-projekt-crossd>
  - <https://www.fhstp.ac.at/de/newsroom/news/eine-gesunde-open-source-landschaft-schaffen>
  - <https://science.apa.at/power-search/15750441504919008005>

## Eigene Projektwebsite

- <https://crossd.tech>
- <https://health.crossd.tech>

Wie bereits weiter oben in den Aktivitäten zu Arbeitspaket 2 angegeben, wird zu Disseminations-, Outreach- und Networking-Zwecken eine eigene Website unter <https://crossd.tech> benutzt. Sie ist auch die Grundlage für die CrOSSD-Plattform, d. h. die Benutzeroberfläche für die indizierten Projekte und erfassten Metriken: <https://health.crossd.tech>

## Geplante Aktivitäten nach netidee-Projektende

Nach Abschluss des CrOSSD-Projekts ist eine Fortsetzung durch zwei Folgeprojekte vorgesehen:

1) Geplant ist eine netidee-Projektanschlussförderung um die Nachhaltigkeit von CrOSSD zu stärken. Es soll die Auswahl an Metriken basierend auf Community-Feedback überarbeitet werden, die Integration in bestehende CI/CD-Pipelines verbessert werden, sowie die (lokalen) OSS-Communities stärker angesprochen werden. Geplant ist dazu u.a. die Organisation von Veranstaltungen zur Vernetzung, um die Sichtbarkeit des Projekts erhöhen.

2) Ein Antrag zur Finanzierung des Projekts CHAoSDEC wurde bei der Österreichischen Forschungsförderungsgesellschaft (FFG) eingereicht. CHAoSDEC steht für „Comprehensive Health and Quality Assessment of Collaborative Digital Ecosystems“ und verfolgt das Ziel, das Gesundheits-Monitoring von kollaborativen Software-Plattformen auch auf Datenökosysteme auszuweiten. Es sollen Daten aus verschiedenen digitalen kollaborativen Ökosystemen analysiert werden, wobei ein besonderer Fokus dabei auf der Vernetzung verschiedener Communities liegt, um verallgemeinerbare Erkenntnisse abzuleiten und diese auf konkrete Anwendungsfälle der Unternehmenspartner anzuwenden. Im Zusammenhang mit der kollaborativen Erstellung und Bearbeitung digitaler Inhalte durch Communities sind die wichtigsten Plattformen MediaWiki, das als Grundlage für Wikipedia dient, die Datenplattformen Semantic MediaWiki und Wikibase (die Wikidata zugrunde liegende Plattform) sowie Open-Source-Entwicklungsplattformen wie GitHub und Gitlab. Jede dieser Plattformen bietet unterschiedliche Funktionen, die auf die jeweiligen Zwecke der Zusammenarbeit zugeschnitten sind, darunter die Erstellung und Bearbeitung von Inhalten, die Zusammenarbeit bei strukturierten Daten und die gemeinsame Softwareentwicklung.

CHAoSDEC integriert die Metriken zur Community-Gesundheit aus dem CrOSSD-Projekt mit Metriken zur Datenqualität für eine ganzheitliche Bewertung der Gesundheit und Qualität kollaborativer digitaler Ökosysteme. Dabei soll ein interdisziplinäres Forschungsteam bestehend aus zwei Hochschulen (FH St. Pölten und WU Wien), einem Partner für sozialwissenschaftliche und geisteswissenschaftliche Forschung (VICESSE) sowie zwei KMUs (nexyo und KMA) zusammenarbeiten.

## Anregungen für Weiterentwicklungen durch Dritte

Das CrOSSD-Projekt bietet Dritten verschiedene Möglichkeiten zur Weiterentwicklung:

### **Metriken weiter ausbauen, AI/LLMs zur Metrikberechnung verwenden:**

- Die Metriken zur Bewertung der Gesundheit von OSS-Projekten können weiterentwickelt und verfeinert werden, indem zusätzliche Faktoren und Kriterien berücksichtigt werden.
- Die Integration von künstlicher Intelligenz und Large Language Models kann dazu beitragen, fortgeschrittene Analysetechniken einzusetzen und prädiktive Modelle für die Bewertung von OSS-Gesundheit zu entwickeln.
- Die Nutzung von AI und LLMs kann auch dazu beitragen, die Automatisierung von Metrikberechnungen zu verbessern und den Prozess effizienter zu gestalten.

**Open Source abseits von Github stärker miteinbeziehen:**

- Um eine umfassende Bewertung der OSS-Gesundheit zu gewährleisten, ist es wichtig, Open-Source-Projekte abseits von GitHub stärker einzubeziehen.
- Die Erweiterung der Datenquellen auf Plattformen wie GitLab, Bitbucket und anderen Open-Source-Hosting-Diensten kann dazu beitragen, eine breitere Palette von Projekten abzudecken und ein umfassenderes Bild der OSS-Community zu erhalten.

**Ausweitung auf andere Communities, z. B. kollaborative Knowledge Graphs/Wikidata:**

- Neben dem Gesundheits-Monitoring von OSS-Projekten kann das CrOSSD-Konzept auch auf andere kollaborative Communities ausgeweitet werden, wie z. B. kollaborative Knowledge Graphs und Datenprojekte wie Wikidata.
- Durch die Anwendung ähnlicher Metriken und Bewertungskriterien können Plattformen, die auf kollaborativer Datenkuratierung basieren (etwa OpenStreetMap oder Open Food Facts), ebenfalls von einem solchen Framework profitieren, um ihre Gesundheit und Qualität zu überwachen und zu verbessern.