



netidee

PROJEKTE

CrOSSD2 –

Extending the Critical Open-Source Software Database

Endbericht | Call 19 | Projekt ID 7399

Lizenz CC BY-SA

Inhalt

1	Einleitung	3
2	Projektbeschreibung	4
3	Verlauf der Arbeitspakete	7
3.1	Arbeitspaket 1: Detailplanung und Formales am Projektstart	7
3.2	Arbeitspaket 2: Projektmanagement, Outreach und Dissemination	7
3.3	Arbeitspaket 3: Weiterentwicklung CrOSSD-Plattform und Metriken	9
3.4	Arbeitspaket 4: Community-Arbeit und qualitative Erhebungen	13
3.5	Arbeitspaket 5 – Dokumentation und Formales am Projektende	15
4	Umsetzung Förderauflagen	16
5	Liste Projektergebnisse	16
6	Verwertung der Projektergebnisse in der Praxis	19
7	Öffentlichkeitsarbeit/ Vernetzung	20
8	Eigene Projektwebsite	20
9	Geplante Aktivitäten nach netidee-Projektende	20
10	Anregungen für Weiterentwicklungen durch Dritte	21

1 Einleitung

Open-Source-Software (OSS) bildet das Rückgrat moderner digitaler Infrastrukturen und ist Bestandteil von über 96 % aller Softwareprodukte. Trotz dieser zentralen Rolle stehen viele OSS-Projekte vor erheblichen Herausforderungen in Bezug auf Nachhaltigkeit, Sicherheit und Wartung. Aufsehenerregende Sicherheitsvorfälle wie der XZ-Utils-Backdoor, Log4Shell oder Heartbleed haben gezeigt, dass selbst kritische Softwarekomponenten häufig von wenigen Einzelpersonen ohne ausreichende Ressourcen gepflegt werden.

Im Vorgängerprojekt „Towards a Critical Open-Source Software Database“ (kurz: CrOSSD1, netidee Call #17) wurde eine Plattform aufgebaut, die es erlaubt, kritische OSS-Projekte zu identifizieren und deren Gesundheitszustand anhand quantitativer Indikatoren – wie Anzahl aktiver Entwickler:innen, Commits oder Releases – zu bewerten. Diese Arbeit hat gezeigt, dass rein quantitative Kennzahlen nicht ausreichen, um ein verlässliches Bild der Projektgesundheit zu gewinnen. Entscheidend für die Nachhaltigkeit und Stabilität von OSS-Projekten sind vor allem die Qualität der Kommunikation, Zusammenarbeit und Governance-Strukturen innerhalb der Community.

In „Extending the Critical Open-Source Software Database“ (kurz: CrOSSD2) haben wir daher bewusst neue Schwerpunkte gesetzt:

- Einführung neuer Metriken auf Basis von NLP-/LLM-Techniken – konkret der CrOSSD Health Analyser, der qualitative Dimensionen wie Freundlichkeit der Kommunikation, Dokumentationsqualität, Entwicklungseffizienz und Projektreifegrad bewertet.
- Stärkere Community-Arbeit durch semi-strukturierte Stakeholder-Interviews sowie aktive Teilnahme an Community-Veranstaltungen, um praxisnahe Perspektiven verschiedener Stakeholder:innen zu erfassen und so eine tiefergehende Bewertung der Projektgesundheit zu ermöglichen.
- Technische Weiterentwicklung der CrOSSD-Plattform hinsichtlich Skalierbarkeit, Benutzerfreundlichkeit und Funktionsumfang.

Die CrOSSD-Plattform (<https://health.crossd.tech/>) indiziert aktuell 2.379 OSS-Projekte mit durchschnittlich 45 Scanvorgängen pro Projekt, umfasst 687.839 Dokumente und verwaltet rund 72 GB an Daten. Die Ergebnisse werden über ein Webinterface sowie über API-Schnittstellen bereitgestellt und stehen unter Open-Source- bzw. Creative-Commons-Lizenzen zur Verfügung.

In diesem Endbericht präsentieren wir die Ergebnisse aller Arbeitspakete von CrOSSD2: die technische Weiterentwicklung der Plattform und ihrer Metriken (Arbeitspaket 3), die qualitative Community-Arbeit und Stakeholder-Interviews (Arbeitspaket 4) sowie die Disseminations- und Vernetzungsaktivitäten (Arbeitspaket 2), die unter anderem zu einer Präsentation auf der

CHAOScon EU 2026 in Brüssel und zur erfolgreichen Anbahnung eines FFG-KIRAS-Folgeprojekts geführt haben.

2 Projektbeschreibung

Projektziele

Das übergeordnete Ziel von CrOSSD2 ist die Weiterentwicklung der im Vorgängerprojekt CrOSSD1 aufgebauten Plattform zur Bewertung der Gesundheit von Open-Source-Software-Projekten. Während CrOSSD1 den Grundstein legte – mit dem Aufbau einer Datenbank, der Definition quantitativer Metriken und der Implementierung einer ersten Plattform –, adressiert CrOSSD2 gezielt die Limitationen des ersten Projekts und erweitert die Plattform in drei Richtungen:

1. *Qualitative Dimensionen erschließen:* Die Erfahrung aus CrOSSD1 hat gezeigt, dass rein quantitative Kennzahlen (z. B. Commit-Frequenz, Issue-Schließzeiten, Contributor-Zahlen) zwar wichtige Indikatoren liefern, aber wesentliche Aspekte der Projektgesundheit nicht abbilden können. Die Qualität der Kommunikation innerhalb eines Projekts, der Umgangston in Issues und Pull Requests, die Willkommenskultur gegenüber neuen Beitragenden, die Dokumentationsqualität oder der Reifegrad der Governance-Strukturen lassen sich nicht allein durch Zählen erfassen. CrOSSD2 setzt daher auf NLP- und LLM-basierte Analysetechniken, um diese qualitativen Dimensionen messbar zu machen. Der eigens entwickelte CrOSSD Health Analyser nutzt lokale Large Language Models, um Repository-Daten automatisiert hinsichtlich Friendliness, Documentation Quality, Development Efficiency und Project Maturity zu bewerten – teilweise unter Einsatz innovativer Multi-Agent-Debate-Verfahren, die robustere und nachvollziehbarere Ergebnisse liefern als einfache LLM-Aufrufe.
2. *Community-Perspektiven integrieren:* Ein zentrales Defizit vieler bestehender Bewertungsansätze ist die fehlende Rückkopplung mit den tatsächlichen Nutzer:innen und Stakeholder:innen von OSS. Welche Kriterien verwenden Entwickler:innen in der Praxis, wenn sie sich für oder gegen ein OSS-Paket entscheiden? Was verstehen verschiedene Akteur:innen unter „Kritikalität“? Und wie wird der potenzielle Nutzen einer Plattform wie CrOSSD eingeschätzt? Um diese Fragen empirisch zu beantworten, wurden in CrOSSD2 über zwanzig semi-strukturierte Stakeholder-Interviews durchgeführt und induktiv analysiert. Die Ergebnisse fließen direkt in die Auswahl, Gewichtung und Interpretation der Metriken ein und stellen sicher, dass die Plattform an den tatsächlichen Bedürfnissen der Praxis ausgerichtet ist.
3. *Technische Plattform ausbauen:* Die CrOSSD-Plattform wurde in CrOSSD2 umfassend weiterentwickelt. Der Crawler wurde auf die GitHub GraphQL API umgestellt, um Scans zu beschleunigen und die API-Last zu reduzieren. Das Webinterface wurde auf aktuelle Svelte/SvelteKit-Versionen migriert und um neue Funktionalitäten ergänzt – insbesondere eine Gruppenansicht, die aggregierte Metriken über Projektgruppen hinweg darstellt, sowie eine überarbeitete Projektansicht mit Benchmarking gegen Durchschnittswerte. Der Projektbestand wurde systematisch erweitert, unter anderem um sämtliche netidee-geförderten OSS-Projekte.

Die Plattform berechnet aktuell 24 quantitative Metriken über die Dimensionen Entwicklungsaktivität, Community, Codequalität, Sicherheit, Issue-Management und Repository-Gesundheit, ergänzt um die vier neuen LLM-basierten Metriken.

Zielgruppe

CrOSSD2 richtet sich an dieselben Zielgruppen wie das Vorgängerprojekt, erweitert den Fokus jedoch insbesondere auf qualitative Aspekte und Community-Bedürfnisse:

- *Projekteigentümer:innen und Maintainer:innen*: Die Plattform bietet einen umfassenden Einblick in die Gesundheit des eigenen Projekts – nicht nur hinsichtlich quantitativer Aktivitätsmetriken, sondern auch bezüglich der wahrgenommenen Freundlichkeit, Dokumentationsqualität und Reife. Dies ermöglicht es Maintainer:innen, gezielt Verbesserungsmaßnahmen zu ergreifen.
- *OSS-Community und Entwickler:innen*: Entwickler:innen, die vor der Entscheidung stehen, welche Bibliotheken oder Frameworks sie in ihren Projekten einsetzen, können mithilfe der CrOSSD-Plattform fundierte Vergleiche auf Basis von Gesundheitsmetriken anstellen. Die Gruppenansicht ermöglicht dabei auch thematische oder fördergeberbezogene Vergleiche (z. B. aller netidee-Projekte).
- *Unternehmen*: Softwareentwicklungsunternehmen und Betreiber kritischer Infrastrukturen können die Plattform nutzen, um Abhängigkeiten in ihren Software-Stacks systematisch auf Gesundheit und Sicherheitsrisiken zu prüfen. Die API-Schnittstellen ermöglichen die Integration in bestehende CI/CD-Prozesse und Software-Supply-Chain-Workflows.
- *Fördergeber:innen und politische Entscheidungsträger:innen*: Die aggregierten Analysen und Gruppenansichten ermöglichen es Fördergebenden wie netidee, die Gesundheit der von ihnen unterstützten Projekte systematisch zu monitoren. Die in CrOSSD2 erarbeiteten Grundlagen – insbesondere die Stakeholder-Interviews und die Vernetzung mit der internationalen CHAOSS-Community – bilden zudem die Basis für das FFG-KIRAS-Folgeprojekt DigiSov-CrOSSD, das die Plattform gezielt in Richtung digitale Souveränität und Sicherheit für die öffentliche Verwaltung weiterentwickeln soll.
- *Wissenschaftliche Community*: Die Open-Source-Veröffentlichung aller Ergebnisse, die dokumentierte Metrik-Pipeline und die API ermöglichen es Forschenden, die CrOSSD-Daten für eigene Studien zu nutzen und die Metriken weiterzuentwickeln. Zudem sind wissenschaftliche Publikationen für hochrangige Journale/Konferenzen in Vorbereitung.

Inhaltlicher Überblick über das Projektergebnis

Das zentrale Projektergebnis von CrOSSD2 ist die weiterentwickelte CrOSSD-Plattform (<https://health.crossd.tech/>), die aktuell 2.379 OSS-Projekte indiziert und deren Gesundheit kontinuierlich überwacht. Die Plattform besteht aus folgenden Kernkomponenten:

Datenerfassung und Crawling: Ein Kubernetes-basiertes System sammelt regelmäßig Metadaten zu OSS-Repositoryn über die GitHub GraphQL API. Die gesammelten Daten werden in einer verteilten ArangoDB-Datenbank gespeichert. Eine Celery Task Queue verwaltet die Scan-Aufgaben und sorgt für deren regelmäßige Ausführung und Wiederholung im Fehlerfall. Die Plattform hat bisher durchschnittlich 45 Scans pro Projekt durchgeführt und verwaltet 687.839 Dokumente (ca. 72 GB).

Metriken-Engine: Die Plattform berechnet 28 Metriken (24 quantitative + 4 LLM-basierte) über mehrere Dimensionen. Die quantitativen Metriken umfassen unter anderem Project Velocity, Code Churn, Contribution Distribution (Bus Factor), Elephant Factor, Security Advisories, Code Dependencies, Criticality Score und Support Rate. Die vier neuen LLM-Metriken – Friendliness, Documentation Quality, Development Efficiency und Project Maturity – werden vom CrOSSD Health Analyser berechnet, einer LangGraph-basierten Pipeline, die lokale LLMs (Ollama/GPT-OSS 20B) nutzt und über innovative Multi-Agent-Debate-Verfahren robuste, evidenzbasierte Bewertungen liefert.

Webinterface: Das mit Svelte/SvelteKit entwickelte Webinterface bietet Suchfunktionalität, detaillierte Projektansichten mit historischen Metrikverläufen, eine Gruppenansicht für aggregierte Analysen (z. B. aller netidee-Projekte unter <https://health.crossd.tech/groups/netidee>) sowie eine umfassende Metrikdokumentation (<https://health.crossd.tech/doc>). Die überarbeitete Projektansicht zeigt die wichtigsten Metriken im Vergleich zu den Durchschnittswerten aller indizierten Projekte.

API-Schnittstellen: Für den maschinellen Zugriff stehen API-Endpunkte zur Verfügung, über die die Liste aller Projekte, die Zeitstempel der Scanvorgänge und die Metriken einzelner Scans abgerufen werden können.

Qualitative Forschungsergebnisse: Die über zwanzig Stakeholder-Interviews liefern empirisch fundierte Erkenntnisse zu den Entscheidungskriterien, Kritikalitätsverständnissen und Nutzungserwartungen verschiedener OSS-Akteur:innen. Diese Ergebnisse werden derzeit in wissenschaftlichen Publikationen aufbereitet und fließen in ein Konzept zur stärkeren Einbindung qualitativer Aspekte in die CrOSSD-Metriken ein.

Community-Vernetzung und Nachhaltigkeit: Durch die Präsentation auf der CHAOSScon EU 2026, die Posterpräsentation auf der SAINT 2025 und die aktive Vernetzung mit der internationalen CHAOSS-Community wurde CrOSSD in der relevanten Fachcommunity sichtbar positioniert. Die erfolgreiche Einreichung des FFG-KIRAS-Folgeprojekts DigiSov-CrOSSD – mit dem Bundeskanzleramt als Bedarfsträger und fünf Industriepartnern (Robert Bosch GmbH, A1 Digital/Exoscale, MBIT Solutions, Solutions4Energy, maglo GmbH) – belegt die Praxisrelevanz und Anschlussfähigkeit der Projektergebnisse und sichert die Weiterentwicklung der Plattform über die netidee-Förderung hinaus.

Alle Ergebnisse sind unter offenen Lizenzen (CC BY-SA 4.0 bzw. Open-Source-Lizenzen) veröffentlicht und stehen der Öffentlichkeit auf der Plattform, auf GitHub (<https://github.com/FH-CrOSSD>) und in der technischen Dokumentation (<https://fh-crossd.github.io/>) zur Verfügung.

3 Verlauf der Arbeitspakete

3.1 Arbeitspaket 1: Detailplanung und Formales am Projektstart

Arbeitspaket 1 wurde bereits mit April 2025 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.

3.2 Arbeitspaket 2: Projektmanagement, Outreach und Dissemination

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

Blogbeiträge

Im Blog unter <https://www.netidee.at/crossd2> haben wir plangemäß Beiträge veröffentlicht:

- Im Februar 2025 haben wir die Ziele unseres Folgeprojekts vorgestellt, sowohl bzgl. der geplanten Gesundheitsanalysen durch KI-Metriken als auch bzgl. der geplanten stärkeren Community-Einbindung: <https://www.netidee.at/crossd2/wie-geht-es-weiter-crossd2>
- Im April 2025 haben wir eine Posterpräsentation auf der Social Artificial Intelligence Night an der FH St. Pölten vorgestellt: <https://www.netidee.at/crossd2/crossd2-auf-der-saint-2025>
- Im August 2025 haben wir unsere Interviewstudie zur Arbeit mit Open-Source-Software vorgestellt sowie zur Mitwirkung bei dieser aufgerufen: <https://www.netidee.at/crossd2/call-participation-oss-interview-study>
- Im Oktober 2025 haben wir einen technischen Beitrag über den Einsatz von LLMs zur automatisierten Bewertung der Gesundheit von Open-Source-Software veröffentlicht. Darin wurde dargestellt, wie KI-gestützte Analysen – etwa die automatische Klassifikation von Commits als Bugfix, Feature oder Refactoring – die Aussagekraft klassischer Metriken

verbessern und kontextualisierte Health-Reports ermöglichen können:
<https://www.netidee.at/crossd2/llms-zur-automatisierten-bewertung-der-gesundheit-von-open-source-software>

- Im März 2026 haben wir über unsere Teilnahme und Präsentation auf der CHAOSScon EU 2026 in Brüssel berichtet: <https://www.netidee.at/crossd2/chaoscon-eu-2026>
- Im Mai 2026 haben wir die zentralen Analyseergebnisse aus den Stakeholder-Interviews (AP4) vorgestellt und deren Implikationen für die Weiterentwicklung der CrOSSD-Metriken diskutiert: <https://www.netidee.at/crossd2/erkenntnisse-aus-22-interviews-mit-der-oss-community>

Insgesamt wurden damit sechs Blogbeiträge im ca. zweimonatlichen Rhythmus veröffentlicht, die den Projektfortschritt von der Zielsetzung über die technische Umsetzung bis hin zu den Ergebnissen der qualitativen Forschung dokumentieren.

Wissenschaftliche Publikationen und Präsentationen

- Im April 2025 wurde das CrOSSD2-Projekt auf der Social Artificial Intelligence Night (SAINT) 2025 an der FH St. Pölten mit einer Posterpräsentation vorgestellt. Das Poster präsentierte die technische Architektur von CrOSSD2, die sich in vier zentrale Komponenten gliedert: Datensammlung, Metriken & Bewertung, Health-Datenbank sowie UI & Integration. Die Präsentation stieß auf großes Interesse bei den Besucher:innen.
- Im Februar 2026 hatten Sebastian Neumaier und Tobias Dam die Möglichkeit, CrOSSD auf der CHAOSScon EU 2026 in Brüssel zu präsentieren. Die Konferenz, organisiert vom CHAOSS Project der Linux Foundation, brachte Forscher:innen, Praktiker:innen und Open-Source-Community-Mitglieder zusammen, um Gesundheit, Nachhaltigkeit und Wirkung von Open-Source-Ökosystemen zu diskutieren. Die Präsentation umfasste die Metriken und Forschungsgrundlagen des Projekts, Erkenntnisse aus den Stakeholder-Interviews, die technische Architektur der Plattform sowie eine Live-Demo. Eine Aufzeichnung des Vortrags ist online verfügbar: <https://youtu.be/X-kKgoN87Ds?list=PL60k37cxl-HT4eJrJd1LJcLfhjeQ8zT-o&t=9486>

Die CHAOSScon EU 2026 war stark Workshop-artig gestaltet: Neben klassischen Vorträgen umfasste das Programm geführte Unconference-Sessions und offene Diskussionsrunden zu drängenden Themen der OSS-Community – darunter die Abhängigkeit von Open-Source-Projekten von proprietären Kommunikationskanälen, Strategien zur nachhaltigen Finanzierung von OSS sowie digitale Souveränität im europäischen Kontext. Die aktive Teilnahme an diesen Workshop-Formaten ermöglichte es, die CrOSSD-Ergebnisse direkt mit der internationalen Community zu diskutieren, Feedback einzuholen und konkrete Anknüpfungspunkte für die Weiterentwicklung zu identifizieren. Die dort geknüpften Kontakte und gewonnenen Erkenntnisse haben maßgeblich zur Vorbereitung des FFG-KIRAS-Folgeantrags beigetragen.

Wissenschaftliche Publikation

Eine wissenschaftliche Publikation zu den Ergebnissen der qualitativen Interviewstudie (AP4) – einschließlich eines Konzepts zur stärkeren Einbindung qualitativer Aspekte in die CrOSSD-Metriken – ist derzeit in Ausarbeitung und wird bei einer einschlägigen Konferenz eingereicht.

Weitere Öffentlichkeitsarbeit umfasste Social-Media-Aktivitäten auf den Accounts der Projektmitarbeiter sowie auf den Kanälen der FH St. Pölten.

3.3 Arbeitspaket 3: Weiterentwicklung CrOSSD-Plattform und Metriken

In diesem Arbeitspaket wurde die CrOSSD-Plattform sowohl technisch als auch funktional wesentlich weiterentwickelt.

Technische Modernisierung und Optimierung

- Sämtliche Softwareabhängigkeiten der einzelnen Komponenten des CrOSSD-Clusters wurden auf aktuelle Versionen aktualisiert, um die Sicherheit und Stabilität zu gewährleisten.
- Der Crawler, welcher die benötigten Daten und Metadaten der GitHub-Repositorys sammelt, wurde grundlegend überarbeitet. Das Akquirieren der Daten erfolgt nun soweit möglich über die GitHub GraphQL API, da hierbei weniger Requests nötig sind als bei der Verwendung der entsprechenden REST APIs. Dies reduziert die verursachte Last auf Seiten GitHubs und beschleunigt die Scans erheblich.
- Das Webinterface wurde auf die aktuelle Version von Svelte und SvelteKit aktualisiert und das Ladeverhalten optimiert.

Neue Funktionalitäten

- *Gruppenansicht*: Eine zentrale Neuerung ist die Gruppenansicht, die es ermöglicht zu sehen, welche Repositorys zu einer bestimmten Gruppe gehören und welche Ausprägungen (Durchschnitt, Minimum, Maximum) die wichtigsten Metriken in dieser Gruppe haben. Dies

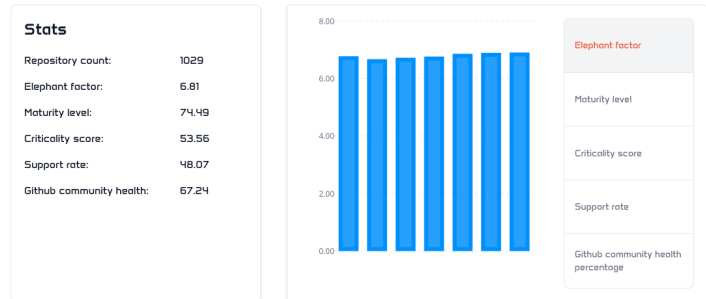
erlaubt erstmals gruppenübergreifende Vergleiche und aggregierte Analysen.

CROSSD

Search Groups Metrics Docs Project & Blog

Group most-starred

Avg Min Max Agg Stddev



Group repositories

Search by name

REPOSITORY 1029

- VOICE/interview_internal_reference
- DXAX/linux-insides
- 1c7/chinese-independent-developer
- 2dust/v2rayN
- 2dust/v2rayNG
- 3blb/manim
- 52lxuewei/helloGitHub
- 996icu/996.ICU
- obi/screenshot-to-code

© 2025 University of Applied Sciences St. Pölten All Rights Reserved.

About Project Privacy Policy Legal Information netidee

- **Überarbeitete Projektansicht:** Die Projektseite der einzelnen Repositories wurde überarbeitet und bietet nun einen verbesserten Überblick über die wichtigsten Metriken und deren Relation zu den Durchschnittswerten aller indizierten Projekte.

CROSSD

Search Groups Metrics Docs Project & Blog

microsoft/vscode



Snapshots

- Tue, 31 Mar 2025 09:59:10 GMT
- Mon, 23 Mar 2025 20:10:02 GMT
- Mon, 02 Mar 2025 20:10:03 GMT
- Mon, 23 Feb 2025 20:10:02 GMT
- Mon, 02 Feb 2025 21:11:48 GMT
- Tue, 27 Jan 2025 13:06:30 GMT
- Tue, 09 Dec 2025 21:10:46 GMT
- Fri, 05 Dec 2025 17:14:15 GMT
- Mon, 22 Sep 2025 20:10:04 GMT
- Mon, 15 Sep 2025 20:10:05 GMT
- Mon, 01 Sep 2025 20:10:03 GMT
- Mon, 25 Aug 2025 20:10:05 GMT

© 2025 University of Applied Sciences St. Pölten All Rights Reserved.

About Project Privacy Policy Legal Information netidee

Erweiterung des Projektbestands – netidee-Projekte

Ein wesentlicher Schwerpunkt war die systematische Erweiterung des Projektbestands. Insbesondere wurden die von netidee geförderten OSS-Projekte seit Beginn des Förderprogramms über die netidee-Website gesammelt und deren GitHub-Links in die Plattform aufgenommen. Alle netidee-geförderten Projekte sind als eigene Gruppe auf der Plattform abrufbar: <https://health.crossd.tech/groups/netidee>. Dadurch können Fördergebende und die OSS-Community die Gesundheit netidee-geförderter Projekte gezielt vergleichen und analysieren.

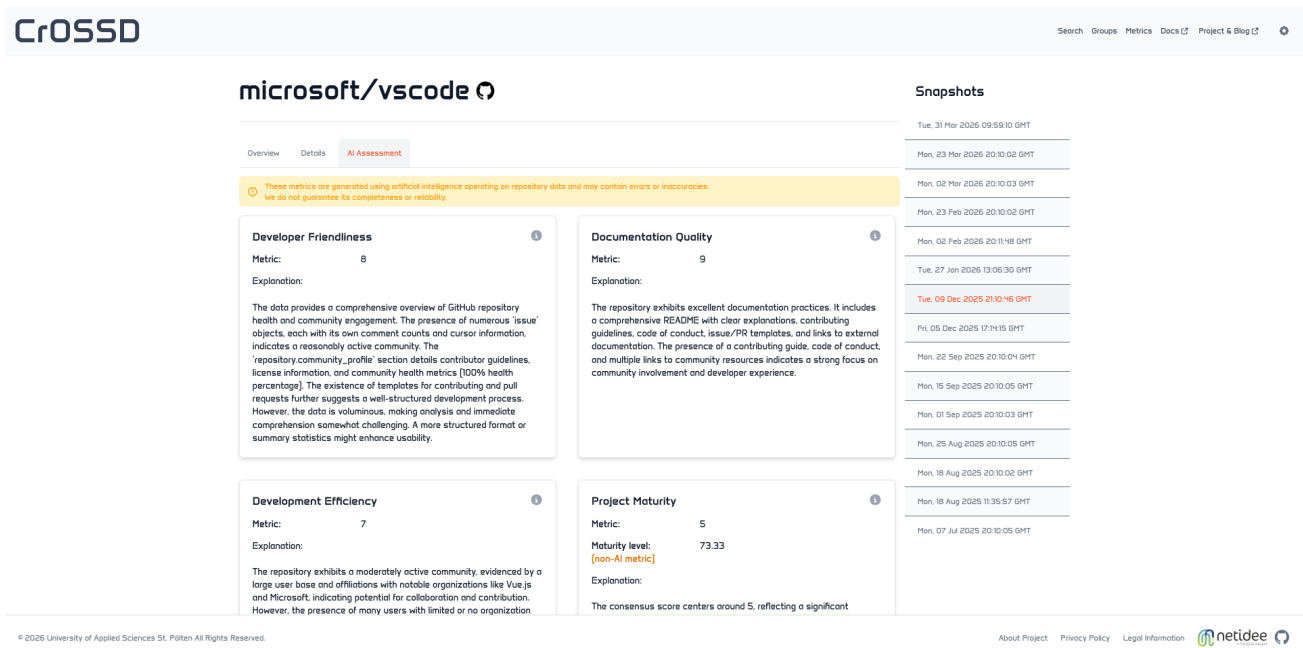
Quantitative Metriken

Die CrOSSD-Plattform berechnet 24 quantitative Metriken über mehrere Dimensionen, die auf der Plattform dokumentiert sind (<https://health.crossd.tech/doc>). Diese umfassen:

- Entwicklungsaktivität: Project Velocity, Churn, Branches, Change/Pull Requests
- Community und Beiträge: Contribution Distribution (inkl. Bus Factor), Number of Support Contributors, Size of Community, Elephant Factor
- Codequalität und Sicherheit: Security Advisories, Code Dependencies, Criticality Score, Maturity Level
- Issue-Management: Issues (New/Closed/Active), Support Rate
- Repository-Gesundheit: GitHub Community Health Percentage, OSI Approved Licenses, Technical Forks

NLP-/LLM-basierte Metriken – CrOSSD Health Analyser

Ein zentrales neues Ergebnis von CrOSSD2 ist der CrOSSD Health Analyser (<https://github.com/FH-CrOSSD/llm-metrics/tree/main>), eine LLM-gestützte Pipeline zur automatisierten Bewertung der Gesundheit von Open-Source-Repositorys. Die Komponente wurde vollständig implementiert und in die CrOSSD-Plattform integriert.



CrOSSD Search Groups Metrics Docs Project & Blog

microsoft/vscode

Overview Details **AI Assessment**

⚠️ These metrics are generated using artificial intelligence operating on repository data and may contain errors or inaccuracies. We do not guarantee its completeness or reliability.

Developer Friendliness 8

Metric: 8

Explanation:

The data provides a comprehensive overview of GitHub repository health and community engagement. The presence of numerous 'issue' objects, each with its own comment counts and cursor information, indicates a reasonably active community. The 'repository.community_profile' section details contributor guidelines, license information, and community health metrics (100% health percentage). The existence of templates for contributing and pull requests further suggests a well-structured development process. However, the data is voluminous, making analysis and immediate comprehension somewhat challenging. A more structured format or summary statistics might enhance usability.

Documentation Quality 9

Metric: 9

Explanation:

The repository exhibits excellent documentation practices. It includes a comprehensive README with clear explanations, contributing guidelines, code of conduct, issue/PR templates, and links to external documentation. The presence of a contributing guide, code of conduct, and multiple links to community resources indicates a strong focus on community involvement and developer experience.

Development Efficiency 7

Metric: 7

Explanation:

The repository exhibits a moderately active community, evidenced by a large user base and affiliations with notable organizations like Vue.js and Microsoft, indicating potential for collaboration and contribution. However, the presence of many users with limited or no organization

Project Maturity 5

Metric: 5

Maturity level: 73.33 (non-AI metric)

Explanation:

The consensus score centers around 5, reflecting a significant

© 2026 University of Applied Sciences St. Pölten All Rights Reserved. About Project Privacy Policy Legal Information netidee

Snapshots

- Tue, 31 Mar 2026 09:59:10 GMT
- Mon, 23 Mar 2026 20:10:02 GMT
- Mon, 02 Mar 2026 20:10:03 GMT
- Mon, 23 Feb 2026 20:10:02 GMT
- Mon, 02 Feb 2026 20:11:48 GMT
- Tue, 27 Jan 2026 13:06:30 GMT
- Tue, 09 Dec 2025 21:10:16 GMT**
- Fri, 05 Dec 2025 17:14:15 GMT
- Mon, 22 Sep 2025 20:10:04 GMT
- Mon, 15 Sep 2025 20:10:05 GMT
- Mon, 01 Sep 2025 20:10:03 GMT
- Mon, 25 Aug 2025 20:10:05 GMT
- Mon, 18 Aug 2025 20:10:02 GMT
- Mon, 18 Aug 2025 11:35:57 GMT
- Mon, 07 Jul 2025 20:10:05 GMT

Die Pipeline basiert auf LangGraph (LangChain) und nutzt lokale LLMs über Ollama (Standard: Gemma 3, 27B Parameter). Repository-Daten werden über die CrOSSD-API abgerufen und über ein konfigurierbares Set an LLM-Metriken bewertet. Es stehen zwei Bewertungsstrategien zur Verfügung:

- *Single-Agent-Pipeline*: Ein einzelner LLM-Aufruf bewertet die Metrik direkt und liefert ein strukturiertes JSON-Ergebnis.

- *Multi-Agent-Debate-Pipeline*: Mehrere konfigurierbare Agenten (z. B. Analyst, Reviewer, User Advocate, Developer Advocate, Security Auditor) diskutieren in mehreren Runden im Round-Robin-Verfahren. Jeder Agent sieht dabei nur eine score-freie Zusammenfassung der Argumente der anderen Agenten, um soziale Kontagion und Herdenverhalten zu verhindern. Ein Synthesizer liest anschließend das vollständige Transkript und gibt den finalen Score aus, verankert am Python-berechneten Median aller genannten Einschätzungen.

Die folgenden LLM-Metriken wurden implementiert:

- *Friendliness* (Single-Agent): Bewertet den Ton und die Willkommenskultur in Entwickler:innen-Interaktionen (Issues, Pull Requests, Diskussionen).
- *Documentation Quality* (Single-Agent): Bewertet README, Contributing Guides, Issue-/PR-Templates und weitere Dokumentation.
- *Development Efficiency* (Debate, 2 Runden, 2 Agenten): Evaluiert Issue-/PR-Durchlaufzeiten, Branching-Strategien und Release-Kadenz.
- *Project Maturity* (Debate, 3 Runden, 5 Agenten): Holistische Reifegradbewertung über Governance, Community, CI/CD und Sicherheit.

Zur Sicherstellung der Score-Integrität wurden mehrere Mechanismen implementiert: Peer-Summaries enthalten keine numerischen Scores (Vermeidung von Herdeneffekten), der Transcript-Median wird als expliziter Anker in den Synthesizer-Prompt injiziert, ein Hard-Clamp begrenzt den finalen Score auf den deklarierten Skalenbereich, und eine JSON-Fallback-Chain fällt bei fehlerhafter LLM-Ausgabe auf den Transcript-Median zurück.

Das System ist modular und erweiterbar: Neue Metriken können durch einfache Konfiguration in `config.py` hinzugefügt werden, ohne bestehenden Code ändern zu müssen. Vollständige Debate-Transkripte werden als strukturierte JSON-Dateien gespeichert und ermöglichen die nachträgliche Analyse und Validierung der LLM-Bewertungen.

Stakeholder-Feedback

Die Ergebnisse der Stakeholder-Interviews aus AP4 fließen ebenfalls in die Priorisierung und Gewichtung der Metriken ein: Auf Basis des qualitativen Feedbacks der Befragten werden die Metriken iterativ angepasst und um jene Dimensionen ergänzt, die von den Stakeholder:innen als besonders relevant identifiziert wurden.

Eine interaktive Demo der Plattform wurde als Jupyter-Notebook erstellt und ist online verfügbar: <https://fh-crossd.github.io/demo/lab/index.html?path=CrOSSD.ipynb>

Dokumentation

Die technische Dokumentation wurde auf <https://fh-crossd.github.io/> konsolidiert und umfasst Installationsanleitungen, Komponentenbeschreibungen (ArangoDB, Redis, Flower, Crawler, Metriken, Frontend, API, Job-Systeme), Konfigurationshinweise sowie bekannte Probleme.

3.4 Arbeitspaket 4: Community-Arbeit und qualitative Erhebungen

Dieses Arbeitspaket stellte einen der zentralen neuen Schwerpunkte von CrOSSD2 dar und ging deutlich über die rein technische Arbeit des Vorgängerprojekts hinaus.

Methodische Entscheidung: Einzelinterviews statt Fokusgruppen

Im ursprünglichen Projektplan waren neben Einzelinterviews auch Fokusgruppen vorgesehen. Im Laufe der Erhebungsphase wurde bewusst zugunsten ausschließlich semi-strukturierter Einzelinterviews entschieden. Diese methodische Anpassung war aus mehreren Gründen geboten:

- Die *heterogene Zusammensetzung* der Zielgruppe – von individuellen Open-Source-Entwickler:innen über Projektmaintainer:innen bis hin zu Entscheidungsträger:innen in Unternehmen – machte die Bildung homogener Fokusgruppen schwierig. Die Befragten sind weltweit verteilt und in unterschiedlichsten Organisationskontexten tätig, was die zeitliche und logistische Koordination von Gruppensitzungen erheblich erschwerte.
- Die *methodische Tiefe* der Einzelinterviews (30–85 Minuten pro Gespräch) erwies sich als besonders geeignet, um die individuelle Bandbreite an Perspektiven, Erfahrungen und Entscheidungskriterien differenziert zu erfassen.
- Das Workshop-artige Format der *CHAOSCon EU 2026* (siehe AP2) hat die Funktion einer Fokusgruppe teilweise komplementiert: In den Unconference-Sessions und Diskussionsrunden wurden die CrOSSD-Ergebnisse mit einem breiteren Fachpublikum diskutiert und validiert.

Die theoretische Sättigung wurde trotz des Verzichts auf Fokusgruppen zuverlässig erreicht; die gewählte Methodik erwies sich als effizienter und ergiebiger für die Forschungsfragen des Projekts.

Stakeholder-Interviews

Es wurden über zwanzig semi-strukturierte Interviews mit verschiedenen Stakeholder:innen durchgeführt. Die Befragten umfassten hauptsächlich Softwareentwickler:innen und Programmierer:innen, aber auch Personen in anderen Funktionen im OSS-Ökosystem – darunter Projektmaintainer:innen, Entscheidungsträger:innen in Unternehmen und Community-Manager:innen. Die Interviews dauerten in der Regel zwischen 30 und 60 Minuten, mit einigen wenigen Ausreißern von 75–85 Minuten. Mittelwert sowie Median der Interviewlänge war 54 Minuten.

Der Interviewleitfaden fokussierte auf drei Kernpunkte:

1. Die benutzten Kriterien für Entscheidungen rund um die Verwendung und/oder Unterstützung von OSS-Paketen.
2. Die Verständnisse und Konzeptionen von Kritikalität im Kontext von OSS.

3. Grundlegende Fragen zum Ziel, der erwarteten Nützlichkeit und der grundlegenden Zugangsweise von CrOSSD als Bewertungsplattform.

Community-Veranstaltungen und Vernetzung

Die Community-Arbeit wurde ergänzend zu den Einzelinterviews durch die aktive Teilnahme an Community-Veranstaltungen umgesetzt:

- Auf der Social Artificial Intelligence Night (SAINT) 2025 an der USTP wurde CrOSSD2 mit einer Posterpräsentation vorgestellt und mit Teilnehmer:innen aus Wissenschaft und Industrie diskutiert.
- Die CHAOSScon EU 2026 in Brüssel bot ein stark Workshop-artiges Format mit Unconference-Sessions und offenen Diskussionsrunden. Die aktive Teilnahme an diesen Formaten ermöglichte es, die CrOSSD-Ergebnisse direkt mit der internationalen CHAOSS-Community zu diskutieren, Feedback zu den verwendeten Metriken einzuholen und Partnerschaften mit anderen OSS-Initiativen zu pflegen.

Auswertung und Ergebnisse

Die Interviews wurden unter Nutzung von Whisper (konkret noScribe) transkribiert und in der Folge induktiv analysiert, um die qualitativen Daten strukturiert aufzubereiten. Eine Kurzfassung der Interviewresultate ist als Deliverable auf der netidee-Projekte verfügbar (und wird hier kurz exzerpiert); die detaillierten Resultate fließen in wissenschaftliche Publikationen ein, die derzeit noch in Vorbereitung sind.

In den Interviews zeichnete sich ein vielschichtiges Bild bzgl. der Qualitätsverständnisse von Open-Source-Projekten ab. Zwar wurde die Funktionalität als grundlegend angesehen, doch reichte sie selten aus. Insgesamt waren die am häufigsten genannten Faktoren für Projektgesundheit die von uns vorab erwarteten: Beliebtheit/Verbreitung der Nutzung, Dokumentationsqualität, Anzahl der Contributors und Aktivität. Es gab darüber hinaus aber einige Varianz und einige interessante Outlier in den Daten, die es noch näher zu analysieren gilt.

Die Definition von Kritikalität erwies sich als überraschend komplex; wenige der Interviewten hatten vorab eine klare Vorstellung davon und viele stellten Rückfragen, was wir denn damit meinten. Dennoch zeichneten sich gewisse Gemeinsamkeiten in den Interviews ab. Ein roter Faden war insbesondere die Frage der Dependencys und der Auswirkungen im Falle eines Ausfalls. Kritische Open-Source-Software wurde als tief in andere Systeme, Infrastrukturen oder Arbeitsabläufe eingebettet beschrieben, was bedeutet, dass ein Ausfall oder eine Kompromittierung erhebliche Folgewirkungen nach sich ziehen würde.

Die Interviews zeigen, dass CrOSSD als potenziell sehr wertvoll angesehen wird und eine klare Lücke in der Open-Source-Sicherheitslandschaft schließt. Es herrschte die überwältigende Meinung vor, dass ein Tool, das die automatisierte Erkennung von Schwachstellen und die

Abhängigkeitsanalyse speziell für Open-Source-Projekte erleichtert, dringend benötigt wird, insbesondere angesichts der zunehmenden Komplexität von Software-Lieferketten. Es wurden jedoch auch einige wichtige Vorbehalte und Vorschläge geäußert; einige Interviewte betrachteten ein derartiges Tool für sich selbst sogar für kaum oder gar nicht nützlich. Insgesamt lieferten die Interviews damit eine große Bandbreite an Input für die Weiterentwicklung von CrOSSD, die wir in den kommenden Projektphasen weiterführen und einfließen lassen werden.

Die Erkenntnisse aus den Interviews fließen direkt in die Weiterentwicklung der Metrikauswahl und der Plattformfunktionalitäten ein. Insbesondere wurden die Ergebnisse auch bei der Präsentation auf der CHAOSScon EU 2026 vorgestellt und diskutiert, was zu positivem Feedback aus der internationalen Community führte.

Konzept zur Einbindung qualitativer Aspekte

Auf Basis der Interviewergebnisse wird ein Konzept zur stärkeren Einbindung qualitativer Aspekte in die CrOSSD-Metriken erarbeitet. Dieses Konzept dokumentiert, welche qualitativen Dimensionen (z. B. Entscheidungskriterien, Konzeptionen von Kritikalität, wahrgenommene Nützlichkeit) in operationalisierbare Metriken oder Metrik-Anpassungen überführt werden können. Das Konzept wird als Teil einer wissenschaftlichen Publikation veröffentlicht, die derzeit in Ausarbeitung ist und bei einer einschlägigen Konferenz eingereicht wird.

Sonstiges

Zuletzt ist noch zu erwähnen, dass während der Projektlaufzeit drei Masterarbeiten betreut wurden; eine davon (zum Thema »Sicherheit von Open-Source-Software – Code-Scanning-Tools zur Absicherung von Open-Source-Software: Auswahl, Einsatz und Grenzen«) wurde Anfang Mai abgeschlossen und ist (allerdings erst nach Berichtslegung) auf <https://phaidra.fhstp.ac.at/> abrufbar. Die anderen beiden Masterarbeiten werden nach Projektende weiter- und fertigbetreut.

3.5 Arbeitspaket 5 – Dokumentation und Formales am Projektende

Die Dokumentation der Projektergebnisse wurde abgeschlossen, die Projekt-Website wurde aktualisiert und alle 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: <https://fh-crosssd.github.io/>
- Die Anwender:innendokumentation ist auf in der Plattformdokumentation hinterlegt: <https://fh-crosssd.github.io/intro/intro.html>
- Der erweiterte Katalog von OSS-Gesundheitsmetriken wurde auf der Plattform dokumentiert: <https://health.crosssd.tech/doc>
- Der gesamte Software-Code wurde auf GitHub veröffentlicht: <https://github.com/FH-CrOSSD>
- Der Endbericht und alle Begleitdokumente wurden fristgerecht erstellt.

4 Umsetzung Förderauflagen

(In der Fördervereinbarung sind keine speziellen Förderauflagen vorgesehen.)

5 Liste Projektergebnisse

Kurzbeschreibung der erreichten Projektergebnisse jeweils mit Open Source Lizenz und Webadresse (netidee Vorgaben beachten!)

1	<i>Projektwischenbericht</i>	CC BY-SA 4.0	https://www.netidee.at/crossd2
2	<i>Projektendbericht</i>	CC BY-SA 4.0	https://www.netidee.at/crossd2
3	<p><i>Entwickler_innen-DOKUMENTATION des Projektergebnisses für andere Entwickler_innen ("Dritte"), die das Projektergebnis nach Projektende nutzen/weiterentwickeln wollen</i></p> <p><i>Für Entwickler_innen (Systemkonzept, ggf. Grobspezifikationen):</i></p> <p><i>a. WAS IST ES</i></p> <p><i>b. FÜR WEN IST ES /WEM HILFT ES WODURCH</i></p> <p><i>c. WIE FUNKTIONIERT ES (für Entwickler_innen: Übersicht und detailliertes Systemkonzept, SW-Struktur)</i></p>	CC BY-SA 4.0	https://fh-crossd.github.io/
4	<i>Anwender_innen-DOKUMENTATION des Projektergebnisses für Anwender_innen, die das</i>	CC BY-SA 4.0	https://fh-crossd.github.io/intro/intro.html

	<p>Projektergebnis nach Projektende nutzen wollen</p> <p>Für Anwender_innen ("Bedienungsanleitung") :</p> <p>a. WAS IST ES</p> <p>b. FÜR WEN IST ES /WEM HILFT ES WODURCH</p> <p>c. WIE FUNKTIONIERT ES</p>		
5	<p>Veröffentlichungsfähiger Einseiter / Zusammenfassung</p> <p>* Kurzfassung WAS FÜR WEN WIE</p> <p>* Liste Projektergebnisse - also diese Liste, ggf. kompromiert</p> <p>* mit Angabe Open Source Lizenz/Webadresse</p> <p>* wo finden Dritte die Projektergebnisse (inkl. Dokumentation Anwender_innen bzw. Entwickler_innen)</p> <p>* mögliche Weiterentwicklungen/ weitere Einsatz-/ Nutzungsmöglichkeiten</p>	CC BY-SA 4.0	https://www.netidee.at/crossd2
6	<p>Dokumentation Externkommunikation zur Erreichung Sichtbarkeit /Nachhaltigkeit (separates Dokument oder als Teil des Endberichtes)</p> <p>* Welche Maßnahmen wurden in welchem Umfang gesetzt</p> <p>* Jeweils Bewertung Aufwand / Nutzen</p>	CC BY-SA 4.0	https://www.netidee.at/crossd2

	<i>* Lessons Learned / Empfehlungen für andere Projekte</i>		
7	<i>Erweiterter und verbesserter Software-Code, basierend auf dem bisherigen CrOSSD-Code. Enthält den Code für den CrOSSD Health Analyser (LangGraph-basierte LLM-Pipeline mit Single-Agent- und Multi-Agent-Debate-Strategien für 4 qualitative Metriken), die technische Weiterentwicklung der Plattform (Crawler, Webinterface, Gruppenansicht, 24 quantitative + 4 LLM-basierte Metriken) sowie die Anpassungen der Metriken aufgrund der Stakeholder-Interviews.</i>	<i>Apache-2.0 license</i>	https://github.com/FH-CrOSSD
8	<i>Konzept zur stärkeren Einbindung qualitativer Aspekte in die CrOSSD-Metriken anhand der durchgeführten Stakeholder-Interviews. Das Konzept wird als Teil einer wissenschaftlichen Publikation veröffentlicht, die derzeit in Ausarbeitung ist.</i>	<i>CC BY-SA 4.0</i>	https://www.netidee.at/crossd2
9	<i>Interaktive Demo der Plattform als Jupyter-Notebook.</i>	<i>Apache-2.0 license</i>	https://fh-crossd.github.io/demo/lab/index.html
10	<i>Daten-Projektergebnisse als Open Data: Gesammelte Metadaten, sowie historische Metrikberechnungen der Repositorys indiziert auf der Plattform.</i>	<i>CC BY-SA 4.0</i>	https://github.com/FH-CrOSSD bzw. https://health.crossd.tech/
11	<i>App</i>	<i>Apache-2.0 license</i>	https://health.crossd.tech/

12	CrOSSD-Website	CC BY-SA 4.0	https://crossd.tech/
----	----------------	--------------	---

6 Verwertung der Projektergebnisse in der Praxis

Die Projektergebnisse von CrOSSD2 werden bereits konkret in der Praxis verwertet und haben zu einer signifikanten Ausweitung der Projektreichweite geführt:

FFG-KIRAS-Folgeprojekt (DigiSov-CrOSSD)

Aufbauend auf den Ergebnissen von CrOSSD1 und CrOSSD2 wurde ein umfangreicher Forschungsantrag bei der Österreichischen Forschungsförderungsgesellschaft (FFG) im Rahmen der KIRAS-Ausschreibung „K-PASS 2025 – F&E Dienstleistungen“ eingereicht. Das Projekt „DigiSov-CrOSSD“ (Digitale Souveränität durch die Critical Open-Source Software Database) adressiert die Bewertung von OSS hinsichtlich Sicherheit und digitaler Souveränität für die öffentliche Verwaltung und Betreiber kritischer Infrastrukturen. Das Bundeskanzleramt fungiert als Bedarfsträger, und Industriepartner wie Robert Bosch GmbH, A1 Digital/Exoscale, MBIT Solutions, Solutions4Energy und maglo GmbH haben Letters of Interest beigelegt. DigiSov-CrOSSD erweitert die CrOSSD-Plattform um souveränitäts- und sicherheitsrelevante Metriken, integriert CVE-Datenbanken und OpenSSF-Scorecard-Indikatoren und adressiert auch neue Risiken durch KI-generierten Code („Vibe-Coding“). Die erfolgreiche Einreichung dieses Folgeantrags zeigt die Relevanz und Anschlussfähigkeit der in CrOSSD2 erarbeiteten Ergebnisse.

Darüber hinaus ist noch ein zweites Folgeprojekt zur Einreichung im ESS26-Call des WWTF (<https://wiki.wwtf.at/books/ess26-resilience-and-crises>) in Vorbereitung, das zum Berichtslegungszeitpunkt allerdings noch nicht spruchreif ist und daher hier nicht im Detail beschrieben wird.

Nutzung durch die OSS-Community und Unternehmen

Die CrOSSD-Plattform wird von verschiedenen Stakeholder:innen genutzt: Softwareentwicklungsunternehmen, die vor der Entscheidung stehen, welche Frameworks sie für ihre Anwendungen verwenden sollen, können verschiedene OSS-Projekte anhand ihrer Gesundheitsmetriken vergleichen. Die neue Gruppenansicht ermöglicht dabei erstmals auch aggregierte Analysen über Projektgruppen hinweg. Durch die Stakeholder-Interviews in AP4 konnten wir sicherstellen, dass die Plattform an den tatsächlichen Bedürfnissen der Praxis ausgerichtet wird.

Vernetzung mit der internationalen CHAOSS-Community

Durch die Präsentation auf der CHAOSScon EU 2026 wurde CrOSSD in der internationalen Community für OSS-Gesundheitsmetriken sichtbar positioniert. Die dort geknüpften Kontakte und gewonnenen Erkenntnisse fließen unmittelbar in die Weiterentwicklung der Plattform ein.

7 Öffentlichkeitsarbeit/ Vernetzung

Die Aktivitäten zur Öffentlichkeitsarbeit und Vernetzung in CrOSSD2 umfassten folgende Maßnahmen:

- Sechs Blogposts im netidee-Blog (<https://www.netidee.at/crossd2>), zu den Themen Projektvorstellung, SAINT-Posterpräsentation, Interviewstudie, LLM-basierte Metriken, CHAOSScon EU 2026 und Analyseergebnisse der Interviews;
- Posterpräsentation auf der Social Artificial Intelligence Night (SAINT) 2025 an der FH St. Pölten;
- Vortrag und Live-Demo auf der CHAOSScon EU 2026 in Brüssel, organisiert vom CHAOSS Project der Linux Foundation (Aufzeichnung: <https://youtu.be/X-kKgoN87Ds?list=PL60k37cxl-HT4eJrJd1LJcLfjheQ8zT-o&t=9486>);
- Social-Media-Aktivitäten auf den Accounts der Projektmitarbeiter sowie auf den Kanälen der FH St. Pölten (z. B. LinkedIn: https://www.linkedin.com/posts/lukas-daniel-klausner-4aa484249_do-you-regularly-work-with-or-develop-open-source-activity-7363161315902664705-l2bZ);
- Aufruf zur Teilnahme an der OSS-Interviewstudie über verschiedene Kanäle.

8 Eigene Projektwebsite

- <https://crossd.tech> – Projekt-Website für Dissemination, Outreach und Networking
- <https://health.crossd.tech> – CrOSSD-Plattform mit Benutzeroberfläche für indizierte Projekte und Gesundheitsmetriken (2.379 Projekte, 687.839 Dokumente)
- <https://fh-crossd.github.io/> – Technische Dokumentation und interaktive Demo

Die Projekt-Website und die CrOSSD-Plattform, die bereits in CrOSSD1 aufgebaut wurden, werden weiterverwendet und wurden in CrOSSD2 wesentlich weiterentwickelt. Die Plattform bietet Suchfunktionalität, Gruppenansichten, detaillierte Metrikdarstellungen mit historischen Verläufen sowie API-Schnittstellen für den maschinellen Zugriff.

9 Geplante Aktivitäten nach netidee-Projektende

Nach Abschluss von CrOSSD2 sind folgende weiterführende Aktivitäten geplant:

1) FFG-KIRAS-Projekt DigiSov-CrOSSD

Der wichtigste geplante Folgeschritt ist das bei der FFG eingereichte Forschungsprojekt DigiSov-CrOSSD. Dieses Projekt erweitert CrOSSD gezielt in Richtung digitale Souveränität und Sicherheit und adressiert vier zentrale Projektziele:

- Ziel 1 – Weiterentwicklung der CrOSSD-Plattform zu einer belastbaren Bewertungsinfrastruktur für Sicherheit und digitale Souveränität von OSS, mit Integration zusätzlicher Datenquellen und Schnittstellen für öffentliche Einrichtungen.
- Ziel 2 – Messbarmachung digitaler Souveränität durch Entwicklung souveränitätsbezogener Metriken auf Basis von Stakeholder-Interviews, darunter SBOM-basierte Abhängigkeitsanalyse, Kritikalitätsbewertung, europäische Verankerung und lizenzrechtliche Rahmenbedingungen.
- Ziel 3 – Integration von Vulnerability- und Exposure-Daten durch Anbindung von CVE-Datenbanken (inkl. der European Vulnerability Database der ENISA), OpenSSF-Scorecard-Metriken und wissenschaftlichen Sicherheitsmetriken.
- Ziel 4 – Validierung und KI-gestützte Qualitätsanalyse, einschließlich Backcasting auf reale Sicherheitsvorfälle und Erkennung von Risiken durch KI-generierten Code (Vibe-Coding).

Das Projekt soll in Zusammenarbeit mit dem Bundeskanzleramt als Bedarfsträger sowie mit Industriepartnern (Robert Bosch GmbH, A1 Digital/Exoscale, MBIT Solutions, Solutions4Energy, maglo GmbH) durchgeführt werden.

2) Wissenschaftliche Publikationen

Die Ergebnisse der qualitativen Interviewstudie (AP4) werden in wissenschaftlichen Publikationen aufbereitet und bei einer einschlägigen Konferenz oder einem hochrangigen Journal eingereicht. Ebenso ist eine Publikation zur LLM-basierten Metrikanreicherung in Vorbereitung.

3) Weiterentwicklung der Plattform

Die CrOSSD-Plattform wird unabhängig von der Projektförderung weiter betrieben und kontinuierlich verbessert. Die in CrOSSD2 gewonnenen Community-Erkenntnisse und die internationale Vernetzung (insbesondere mit der CHAOSS-Community) bilden eine solide Basis für die nachhaltige Weiterentwicklung.

10 Anregungen für Weiterentwicklungen durch Dritte

Das CrOSSD-Projekt bietet Dritten verschiedene Möglichkeiten zur Nutzung und Weiterentwicklung:

Integration in CI/CD-Pipelines und Entwicklungsworkflows:

- Die API-Schnittstellen der CrOSSD-Plattform ermöglichen die Integration von Gesundheitsmetriken in bestehende Entwicklungs- und Deployment-Prozesse.

- Unternehmen können automatisierte Checks auf Basis der CrOSSD-Metriken in ihre Software-Supply-Chain-Prozesse einbauen, um Abhängigkeiten regelmäßig auf Gesundheit und Sicherheit zu prüfen.

Erweiterung der Metriken und KI-basierte Analysen:

- Die in CrOSSD2 begonnene Arbeit zur LLM-basierten Metrikanreicherung kann weiter ausgebaut werden, um fortgeschrittenere Analysen wie automatische Sentimentanalysen von Kommunikation in Issues und Pull Requests, prädiktive Modelle für Projektgesundheit oder die Erkennung von KI-generiertem Code umzusetzen.
- Die Integration weiterer Sicherheitsmetriken (z. B. OpenSSF Scorecard, CVE-Datenbanken) bietet Potenzial für umfassendere Sicherheitsbewertungen.

Digitale Souveränität und öffentliche Verwaltung:

- Die im FFG-KIRAS-Antrag DigiSov-CrOSSD adressierten Themen – insbesondere die Messbarmachung digitaler Souveränität und die Bewertung von OSS für kritische Infrastrukturen – bieten Anknüpfungspunkte für öffentliche Einrichtungen und Betreiber kritischer Infrastrukturen in ganz Europa.
- Die CrOSSD-Plattform kann als Grundlage für Beschaffungsentscheidungen im öffentlichen Sektor dienen.

Open Source abseits von GitHub:

- Die Erweiterung der Datenquellen auf Plattformen wie GitLab, Codeberg und anderen Open-Source-Hosting-Diensten kann dazu beitragen, eine breitere Palette von Projekten abzudecken und ein umfassenderes Bild der OSS-Community zu erhalten.
- Im Kontext digitaler Souveränität ist die Verringerung der Abhängigkeit von US-amerikanischen Hosting-Plattformen ein besonders relevantes Thema.

Ausweitung auf andere kollaborative Ökosysteme:

- Das CrOSSD-Konzept kann auf andere kollaborative Communitys ausgeweitet werden, etwa auf kollaborative Knowledge Graphs (z. B. Wikidata), Datenprojekte (z. B. OpenStreetMap, Open Food Facts) oder andere digitale Ökosysteme.
- Durch die Anwendung ähnlicher Metriken und Bewertungskriterien können auch Plattformen, die auf kollaborativer Datenkuratierung basieren, von einem solchen Framework profitieren.