DaTra

Endbericht | Call 17 | Projekt ID 6362

Inhalt

1 Einleitung

Das Projekt DaTra wurde im Zeitraum von Dezember 2022 bis Jänner 2024 durchgeführt. Das Hauptprojektergebnis ist eine Webplattform, welche es Nutzer:innen ermöglicht herauszufinden, welche Daten im Internet über sie gespeichert sind. Unser Ziel ist es, mit DaTra Awareness für den Umgang mit persönlichen Informationen im Internet zu schaffen, indem transparent dargestellt wird, wo welche Daten öffentlich verfügbar sind und mit welchen einfachen Methoden, Nutzer:innen selbst weitere persönliche Informationen im Internet aufspüren können.

Die Plattform ist unter https://datra.sec.univie.ac.at öffentlich – auch nach Ende der Projektlaufzeit – verfügbar.

2 Projektbeschreibung

Beschreibung der Projektziele / Zielgruppe und inhaltlicher Überblick über das Projektergebnis (max. 5 Seiten)

Viele Teile unseres Lebens finden mittlerweile digital statt – wir kommunizieren, organisieren und präsentieren uns auf verschiedensten Plattformen im Internet. Doch wie jeder öffentliche Raum bergen auch Social-Media Plattformen Gefahren. Häufig sind diese zu unspezifisch und auch Gegenmaßnahmen – beispielsweise strengere Privatssphäreeinstellungen – werden auf Plattformen nicht transparent genug vermittelt, wodurch ein nachlässiger Umgang mit den eigenen Daten ermöglicht und sogar gefördert wird. Hier wollen wir mit DaTra eine nachhaltige Möglichkeit entwickeln, welche es ermöglicht, die eigenen Social-Media Kanäle unter die Lupe zu nehmen und zu prüfen, was mit den eigenen Daten geschieht. Für unser Projekt bilden Digital Natives die primäre Zielgruppe, da sie aufgrund ihrer hohen Präsenz im Internet und auf Social-Media besonders gefährdet sind. In zweiter Linie können allerdings natürlich alle Nutzer:innen von Social-Media von unserer Web-Plattform profitieren – denn Datenspuren hinterlassen wir alle bei unserer täglichen Nutzung des Internets.

Im Laufe des Projekts hatten wir vor allem mit dem immer stärker eingeschränkten Zugriff auf die APIs der Plattformanbieter im Jahr 2023 zu kämpfen. Dass diese Problematik während der Projektlaufzeit eintreten könnte, haben wir im Projektantrag als mögliches Risiko bei der Umsetzung erwähnt. Dennoch hat uns der Menge an drastischen Einschränkungen beim Zugriff

auf Social-Media-Daten über APIs in den vergangenen 12 Monaten – ausgelöst nicht zuletzt durch den Erfolg von ChatGPT und die Diskussion um die kostenlose Nutzung von großen Datenmengen aus Social Media durch OpenAI – überrascht. Nach X (Twitter) im Jänner 2023 hat auch Reddit im August 2023 den nicht-kommerziellen API-Zugriff praktisch vollständig gesperrt und somit die Nutzung von Daten aus diesen Plattformen in DaTra verhindert. Dadurch hat sich die technische Herangehensweise an das Projekt geändert ohne jedoch das übergeordnete Ziel, Internet-Nutzer:innen eine einfache Möglichkeit zu geben, ihre bewusste und unbewusste Präsenz in Social Media überprüfen zu können, aus den Augen zu verlieren.

Der Startpunkt einer DaTra-Analyse ist immer die Anmeldung über ein bestehendes Social-Media-Profil. Wir haben uns für die Social-Login-Anbieter Google, Facebook sowie LinkedIn entschieden. Abbildung 1 zeigt die Startseite von DaTra. Für die weitere Personensuche im Internet und insbesondere auf Social-Media-Plattformen nutzen wir ausschließlich Daten, welche wir aus dem angemeldeten Account extrahieren. Dadurch wollen wir die missbräuchliche Nutzung von DaTra für Stalking-Versuche zumindest erschweren, da Namen, Accountnamen, usw. nicht direkt eingegeben werden können. Abbildung 2 zeigt die allgemeinen Account-Informationen nach dem Einloggen auf der DaTra-Plattform.

Abbildung 1: Die Startseite von DaTra

Die Verarbeitung der Accountdaten erfolgt auf zwei unterschiedliche Weisen. Zum einen werden die Daten direkt nach der Anmeldung an das OSINT-Tool Sherlock übergeben, was dazu dient,

weitere Social-Media-Accounts zu identifizieren, die entweder denselben Benutzernamen wie der angemeldete Account besitzen oder Kombinationen aus Vor- und Nachnamen (z.B. mmustermann, max-mustermann, usw.) aus dem Account verwenden. Dies ermöglicht das Auffinden von alten, möglicherweise nicht mehr aktiv genutzten Accounts, die gelöscht werden können, und hilft bei der Erkennung von Fällen von Identitätsdiebstahl, also der unerlaubten Nachahmung einer Person durch Fake-Profile. Wir suchen insgesamt auf fast 400 Webplattformen nach Accounts, wobei NSWF-Plattformen aus Jugendschutzgründen ausgenommen wurden. Abbildung 3 zeigt die Erkennung und Darstellung identifizierter Social-Media-Accounts.

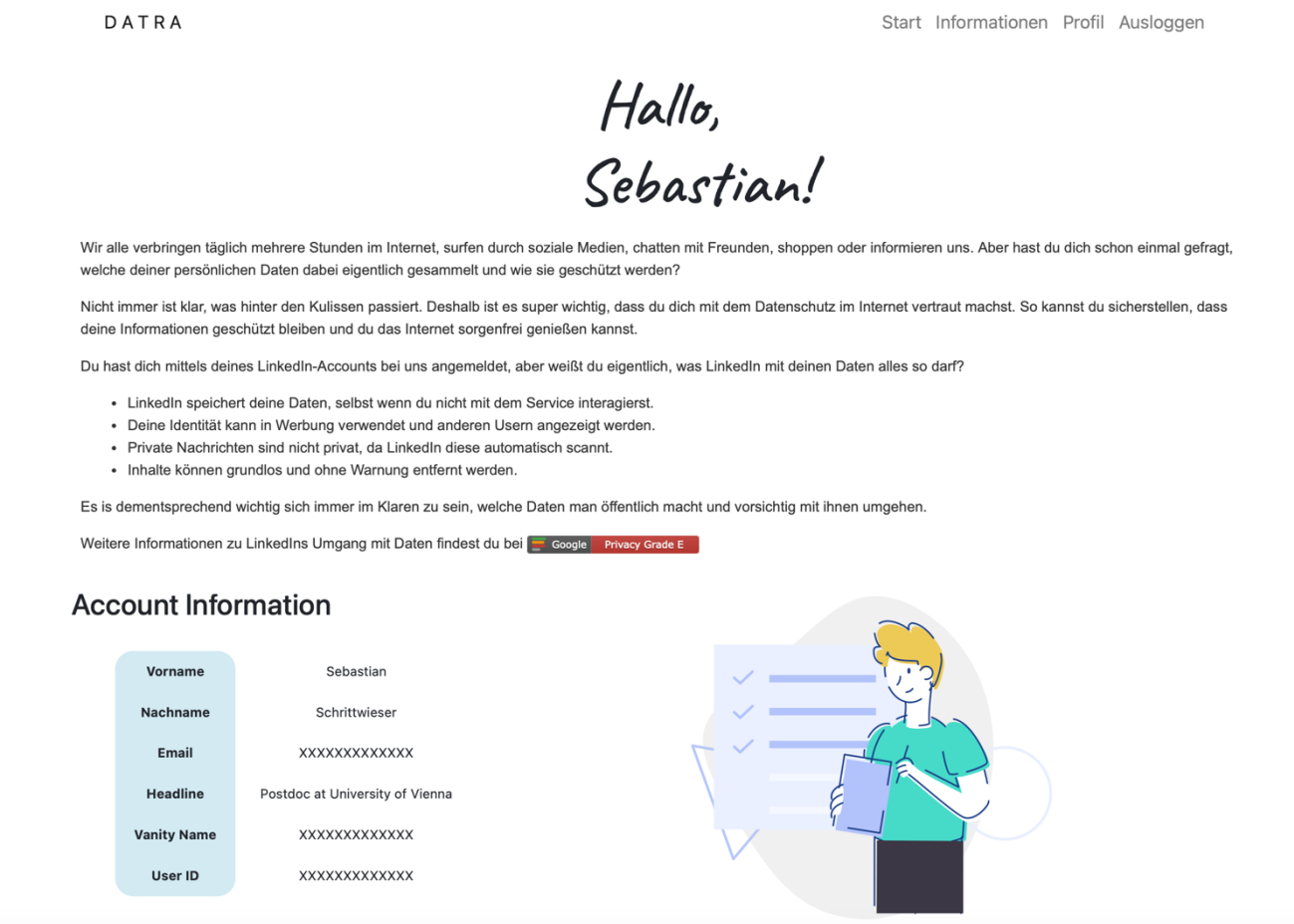

Abbildung 2: Accountinformationen und grundlegender Datenschutzinformationen einer Social-Media-Plattform

Zum anderen erfolgt eine Websuche, die in einer anschließenden Kategorisierung und Darstellung der Ergebnisse resultiert. Diese Ergebnisse können Bilder, Erwähnungen in Social Media, Social-Media-Profile mit öffentlichem Profil sowie weitere Webergebnisse umfassen. Zusätzlich wird Wissen vermittelt, wie man nach Informationen über sich selbst im Internet suchen kann. Ein wichtiger Bestandteil dabei sind die sogenannten "Google Dorks", Suchanfragen, die spezielle Suchmaschinen-Funktionen nutzen, um gezielt Informationen über sich selbst zu finden. Wir stellen die Dorks automatisiert zusammen und ermöglichen eine

Websuche mit einem Klick. Weiters stellen wir den Dork auch dar und erklären ihn, um ein Bewusstsein dafür zu schaffen, wie man effektiv nach eigenen Daten im Internet recherchiert.

In der Accountansicht von DaTra werden außerdem die Privatsphäreneinstellungen der eigenen Inhalte überprüft und angezeigt, wo dies möglich und sinnvoll ist. Diese Funktion ist besonders bei Facebook relevant, wo wir die Sichtbarkeitseinstellungen der einzelnen Posts über die API automatisiert überprüfen können. Bei den anderen Plattformen, wo keine API-Zugriffe möglich sind, stellen wir textuelle Informationen dar.

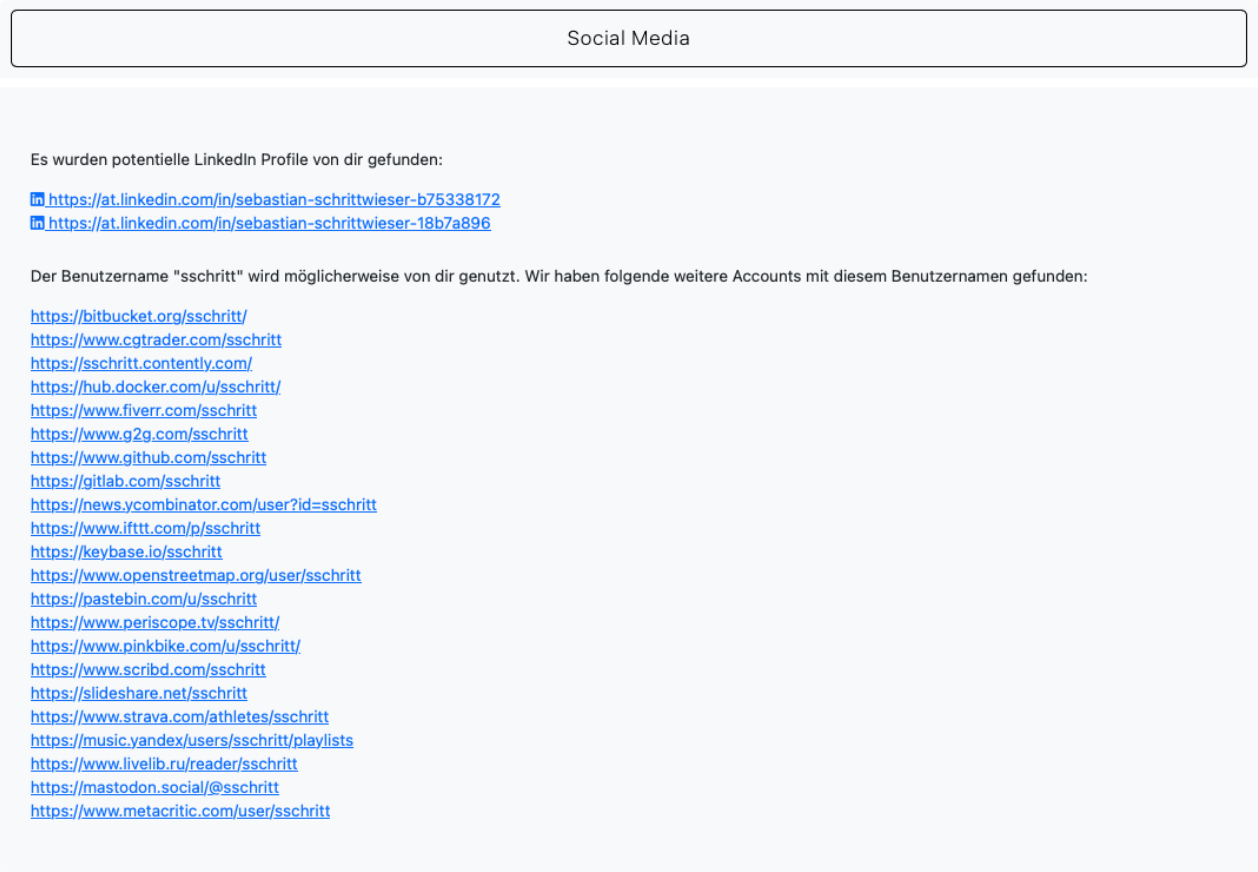

Abbildung 3: Erkennung und Darstellung identifizierter Social-Media-Accounts

3 Verlauf der Arbeitspakete

3.1 Arbeitspaket 1 - *Detailplanung und Formales am Projektstart*

Arbeitspaket 1 wurde im Dezember 2022 mit der Abgabe und Abnahme des Detailprojektplans, dem ersten Blogeintrag sowie dem ersten Förderratenabruf abgeschlossen.

3.2 Arbeitspaket 2 - *Datenquellen identifizieren*

Innerhalb dieses Arbeitspakets wurden verschiedene Datenquellen identifiziert, welche im Rahmen der DaTra-Webplattform verwendet werden könnten. Unter anderem wurden dabei verschiedenste OSINT-Tools (Spiderfoot, sn0int, Sherlock, etc.) recherchiert und getestet. Ergebnisse wurden in einem Dokument tabellarisch zusammengefasst und für die Verwendung in DaTra bewertet. Die jeweilige Lizenzierung der Tools wurde dabei ebenfalls analysiert. Weiters wurden auch APIs von verschiedenen Social-Media-Plattformen (Facebook, Instagram, LinkedIn, Google, Reddit und Twitter) für die Eignung in DaTra betrachtet. Als Ergebnis dieser Analyse wurden Facebook, LinkedIn und Google ausgewählt, um in einem ersten Schritt in DaTra eingebunden zu werden. Instagram kann in weiterer Folge über die Facebook-API integriert werden, Twitter und Reddit haben den kostenlosen API-Zugriff nach dem Projektstart eingestellt. Aus diesem Grund ist eine Einbindung dieser beiden Plattformen nicht mehr möglich.

3.3 Arbeitspaket 3 - *Festlegung empfohlener Privatsphäre-Einstellungen*

Dieses Arbeitspaket beinhaltet die grundsätzliche Analyse der Privatsphäre-Einstellungen verschiedenster Sozialer Medien. Wir haben verfügbare Account-Einstellungen für das eigene Profil sowie Posts bei Facebook, LinkedIn und Google analysiert und Empfehlungen für sichere Einstellungen zusammengefasst. Diese werden in der Webplattform DaTra zur Verfügung gestellt, sowie – wo möglich – direkt in die Account-Analyse eingebunden.

3.4 Arbeitspaket 4 - *Einbindung Sozialer Netzwerke*

In diesem Arbeitspaket wurden die Schnittstellen zu Sozialen Netzwerken implementiert. Die APIs von zuvor ausgewählten Plattformen – Facebook, LinkedIn und Google – wurden dabei verwendet. Die konkrete Anbindung erfolgt über ein Social-Login, wobei sich User:innen mit Accounts der schon erwähnten Plattformen bei DaTra anmelden können. Dadurch ist zusätzlich

gewährleistet, dass nur die Besitzer:innen der jeweiligen Accounts Zugriff auf die Datenanalyse von DaTra haben. Da die Nutzung der APIs auf den meisten Plattformen in den letzten Monaten stark eingeschränkt wurde, ist leider darüber oft nur eine rudimentäre Datenanalyse möglich. Wir haben diesem Umstand, welcher auch als mögliches Projektrisiko bei der Antragsstellung erwähnt wurde, durch eine vermehrte Nutzung von Suchmaschinen-basierter Datensammlung Rechnung getragen. Nach dem initialen Login über eine der drei Plattformen, werden die Accountdaten verwendet, um weitere Profile auf anderen Plattformen zu finden. Dazu werden automatisiert Suchmaschinen-Anfragen gestellt und aus den Ergebnissen Social-Media-Accounts extrahiert. Zusätzlich haben wir das OSINT-Tool Sherlock (https://github.com/sherlock-project) eingebunden, um Social-Media-Accounts mit dem gleichen Usernamen bzw. einer Kombination aus Vorname und Nachname auf fast 400 Web-Plattformen zu identifizieren.

3.5 Arbeitspaket 5 – *Webapplikation*

Dieses Arbeitspaket beinhaltet die konkrete Implementierung der Webapplikation DaTra. Nutzer:innen sind dabei in der Lage sich über eine der drei Sozialen Medien – Facebook, LinkedIn, Google – anzumelden, um so eine auf sie zugeschnittene Aufschlüsselung ihrer Daten zu erhalten. Da die Dienste jeweils unterschiedliche Limitationen ihrer APIs aufweisen, ist auch die Aufschlüsselung auf die verschiedenen Plattformen zugeschnitten. Zusätzlich erhalten Nutzer:innen Informationen zu Privatsphäre-Einstellungen und Empfehlungen unsererseits, um ihre Daten besser schützen zu können. Nach dem Anmelden über einen der drei Social-Logins werden öffentlich zugängliche Daten auf den Profilen der Nutzer:innen dargestellt und unter anderem grafisch aufgeschlüsselt. Beispielsweise sind die Anteile der unterschiedlichen Sichtbarkeitseinstellungen der Posts auf dem eigenen Facebook-Profil mittels eines Tortendiagramms dargestellt. Weiters ist eine Seite mit Informationen zu Privatsphäre-Einstellungen aufrufbar.

Da auch eine tiefergreifende Analyse über OSINT-Tools implementiert wird, werden die Daten der Nutzer:innen temporär in einer SQLite-Datenbank abgespeichert. Diese Daten werden nach dem Ausloggen, Schließen des Browserfensters bzw. nach 24 Stunden automatisch gelöscht. Eine zwischenzeitlich umgesetzte Funktionalität, bei der das Ergebnis der OSINT-Analyse nach dem asynchronen Aufruf von Sherlock, per Mail an die Nutzer:innen verschickt wurde, haben wir wieder verworfen und durch eine direkte Einbindung in die Plattform ersetzt. Im Zuge der Tests hat sich gezeigt, dass die benötigte Zeit für den kompletten Durchlauf von Sherlock nur wenige Minuten beträgt und somit eine direkte Darstellung umsetzbar ist. Dadurch müssen wir auch keine Mailadresse mehr abfragen, wodurch das Konzept der Datensparsamkeit der Plattform besser umgesetzt werden kann.

Anders als ursprünglich geplant, verwenden wir für unsere Plattform anstatt Django das Flask-Framework (ebenfalls in Python implementiert). Im Zuge der Umsetzung hat sich rasch

herausgestellt, dass der Overhead von Django dessen Einsatz nicht rechtfertigt und dessen Authentication-Modul, welches wir ursprünglich verwenden wollten, für DaTra nicht geeignet ist. Den genauen technischen Aufbau von DaTra beschreiben wir in Arbeitspaket 7.

3.6 Arbeitspaket 6 – *Implementierung Privatsphäreneinstellungs-Check*

Die Arbeiten in diesem Arbeitspaket waren leider am meisten von den API-Änderungen der Plattform-Anbieter betroffen. Konzepte und auch Code mussten mehrfach verworfen und erneut erstellt werden. Der Fokus wurde mehr in Richtung Websuche (in Arbeitspaket 4) verschoben, welche unabhängig von den verfügbaren APIs der Plattform-Anbieter funktioniert. Die Facebook-API erlaubt eine feingranulare Analyse der Sichtbarkeitseinstellungen aller bisher in einem Profil veröffentlichten Beiträge. Wir nutzen diese Funktionalität, um Beiträge zu identifizieren, die mit erweiterten Sichtbarkeitseinstellungen veröffentlicht wurden. Die Ergebnisse visualisieren wir in einer verständlichen Tortengrafik für welche wir chart.js einsetzen.

3.7 Arbeitspaket 7 – *Testing und Bereitstellung*

Die Applikation hosten wir auf einem Server der Universität Wien und nutzen eine Vielzahl von Open-Source-Projekten. Die einzelnen Komponenten von DaTra laufen dabei in unterschiedlichen Docker-Containern, welche wir mittels Docker Compose verwalten. Die eigentlich Webapplikation ist in Flask geschrieben. Für die Produktionsumgebung nutzen wir gunicorn, einen Python WSGI Webserver, welchen wir hinter einen Reverse Proxy auf Basis von nginx gestellt haben. Für die verschlüsselte HTTPS-Verbindung zum Server nutzen wir Let's Encrypt, wobei certbot sich um die automatische Beantragung der Zertifikate kümmert. Das Suchen von weiteren Social-Media-Account mit dem OSINT-Tool Sherlock kann mehrere Minuten in Anspruch nehmen und wird direkt nach dem Einloggen in die DaTra-Plattform über die asynchrone Aufgabenwarteschlange celery gestartet und die Ergebnisse nach Abschluss der Analyse direkt in der Auswertungsseite von DaTra dargestellt. Als Nachrichtenbroker verwenden wir dazu RabbitMQ. Ergebnisse werden für maximal 24 Stunden in einer SQLite-Datenbank gespeichert und danach automatisch gelöscht (bzw. auch nach dem Ausloggen oder Schließen des Browserfensters).

In diesem Arbeitspaket wurden sowohl eine Entwicklungs- als auch eine Produktivumgebung umgesetzt, intensiv getestet und auf den Servern der Universität Wien zur öffentlichen Nutzung zur Verfügung gestellt.

3.8 Arbeitspaket 8 – *Dokumentation und Formales am Projektende*

Dieses Arbeitspaket beinhaltet die abschließende Prüfung und Dokumentation des Projekts. Es wurde ein letzter Blogeintrag verfasst. Der Projektendbericht, Zusammenfassung,

Anwender:innen-Dokumentation und Entwickler:innen-Dokumentation wurden erstellt und an netidee übermittelt sowie auf die Projektwebsite hochgeladen. Die Endabrechnung inkl. aller Originalbelege wurde dokumentiert und an netidee übermittelt. Die Projektwebsite wurde aktualisiert und alle Ergebnisse unter Angaben der Lizenzen der Öffentlichkeit zur Verfügung gestellt. Das Projekt wurde somit vollständig dokumentiert abgeschlossen.

4 Umsetzung Förderauflagen

Dieses Kapitel ist nur relevant, wenn in der Fördervereinbarung spezielle Förderauflagen festgelegt wurden. In diesem Fall soll in diesem Kapitel dargestellt werden, wie diese berücksichtigt werden.

Das Projekt hat keine Förderauflagen.

5 Liste Projektendergebnisse

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

1	*Projektzwischenbericht*	*CC BY-SA 4.0*	*https://www.netidee.at/datra/*
2	*Projektendbericht*	*CC BY-SA 4.0*	*https://www.netidee.at/datra/*
3	*Entwickler_innen-DOKUMENTATION*	*CC BY-SA 4.0*	*https://www.netidee.at/datra/*
4	*Anwender_innen-DOKUMENTATION*	*CC BY-SA 4.0*	*https://www.netidee.at/datra/*
5	*Veröffentlichungsfähige Einseiter / Zusammenfassung*	*CC BY-SA 4.0*	*https://www.netidee.at/datra/*
6	*Dokumentation Externkommunikation zur Erreichung Sichtbarkeit /Nachhaltigkeit (als Teil des Endberichtes)*	*CC BY-SA 4.0*	*https://www.netidee.at/datra/*
7	*Webapplikation „DaTra“*	*GPL 3.0*	*https://github.com/sschritt/netidee-datra* *https://www.netidee.at/datra/*
8	*Privatsphäreneinstellungscheck*	*GPL 3.0*	*https://github.com/sschritt/netidee-datra* *https://www.netidee.at/datra/*

6 Verwertung der Projektergebnisse in der Praxis

Angaben zur Verwertung der Projektergebnisse in der Praxis

Wir haben uns viele Gedanken zur Verwertung der Projektergebnisse in der Praxis gemacht und wie wir diese Erkenntnisse effektiv nutzen können, um einen positiven und nachhaltigen Einfluss zu erzielen. Ein zentrales Element unserer Strategie ist es, die im Rahmen unseres Projektes entwickelte Plattform der Öffentlichkeit unentgeltlich zur Verfügung zu stellen, wodurch wir eine breite Nutzerbasis erreichen und einen wesentlichen Beitrag zur Verbesserung der digitalen Self-Awareness in der Zielgruppe leisten können.

Darüber hinaus planen wir, die Plattform aktiv für die Security-Awarenessbildung einzusetzen, wie wir in Kapitel 7 näher beschreiben. Dies umfasst sowohl die Sensibilisierung für Datenschutzrisiken als auch die Vermittlung von Kenntnissen und Fähigkeiten, um diese Risiken zu minimieren.

Wir streben mit der Plattform keine Kommerzialisierung des Konzepts an. Wir sind uns jedoch der Herausforderungen bewusst, die mit der Abhängigkeit von externen APIs einhergehen. Änderungen oder Einschränkungen dieser APIs könnten potenziell die Funktionalität unserer Plattform beeinträchtigen. Auch wenn keine weitere Finanzierung des Projekts existiert, wollen wir versuchen bei Bedarf rasche Anpassungen vorzunehmen, um die Kontinuität der Plattform auch in Zukunft bestmöglich zu gewährleisten.

Die Verwertung der Projektergebnisse in der Praxis zielt darauf ab, einen offenen, kostenlosen und effektiven Zugang zu Methoden für digitale Self-Awareness zu ermöglichen. Wir sind bestrebt, einen nachhaltigen und positiven Einfluss auf die digitale Sicherheit und Bildung in der Zielgruppe der Digital Natives zu haben.

7 Öffentlichkeitsarbeit/ Vernetzung

Beschreibung der im Rahmen Ihres netidee-Projektes bereits erfolgten bzw. noch geplanten Öffentlichkeitsarbeit oder Vernetzung

Wir legen großen Wert auf effektive Öffentlichkeitsarbeit und Vernetzung, um unsere Ziele erfolgreich zu erreichen und eine breite Wirkung zu erzielen. Ein zentraler Aspekt unserer Strategie ist die enge Zusammenarbeit mit verschiedenen Projektpartnern im Bereich der IT-Security-Awarenessbildung. Durch diese Kooperationen erweitern wir nicht nur unsere Reichweite, sondern profitieren auch von der Expertise und den Netzwerken unserer Partner.

Unser zentraler Partner in der IT-Security-Awarenessbildung ist das Österreichische Institut für angewandte Telekommunikation (ÖIAT), das uns mit seiner Expertise im Bereich der digitalen Bildung und Medienkompetenz unterstützt. Durch die Zusammenarbeit mit dem ÖIAT können wir

sicherstellen, dass unsere Inhalte nicht nur effektiv verbreitet werden – etwa in Workshops für Schüler:innen – , sondern auch fachlich fundiert und zielgruppengerecht aufbereitet werden.

Ein weiterer wichtiger Partner in diesem Bereich ist die Arbeiterkammer Niederösterreich, die im Konsumentenschutz eine breite Zielgruppe erreicht.

Zudem planen wir, unser Projekt im Rahmen des Women4Cyber Austria Chapter aktiv zu präsentieren und zu diskutieren. Dieses Netzwerk bietet eine hervorragende Plattform, um Frauen in der Cybersecurity-Branche zu erreichen und zu fördern.

Neben diesen spezifischen Partnerschaften setzen wir auch auf allgemeinere Maßnahmen zur Öffentlichkeitsarbeit, bei denen wir durch die Universität Wien unterstützt werden (z.B. auf den Social-Media-Kanälen). Durch die Kombination aus strategischen Partnerschaften und allgemeinen Maßnahmen streben wir danach, unser Projekt nachhaltig in der Öffentlichkeit zu verankern und einen Beitrag zur digitalen Bildung und Sicherheit zu leisten.

8 Eigene Projektwebsite

Wird zusätzlich zur netidee-Projektwebsite noch eine eigene Website betrieben, so ist hier die Adresse anzugeben.

Die DaTra-Plattform ist unter https://datra.sec.univie.ac.at erreichbar.

9 Geplante Aktivitäten nach netidee-Projektende

Sind weiterführende Aktivitäten nach dem netidee-Projektende geplant?

Die Plattform (https://datra.sec.univie.ac.at) von uns weiterbetrieben und unentgeltlich der Öffentlich als Awarenessmaßnahme zur Verfügung gestellt. Wir planen wir, die Plattform aktiv in der Awarenessarbeit mit Jugendlichen einzusetzen.

10 Anregungen für Weiterentwicklungen durch Dritte

Welche Nutzungs- und Weiterentwicklungsmöglichkeiten für Dritte ergeben sich durch Ihr netidee-Projekt bzw. empfehlen Sie?

DaTra lebt von der Anbindung an Social-Media-Plattformen. Durch unseren modularen Ansatz ist es für Entwickler:innen einfach möglich, weitere Plattformen anzubinden und somit die Möglichkeiten von DaTra auszubauen.