

Abschlussbericht

IVOD – Interaktive Visualisierung von Open Data

im Rahmen des Förderprogramms

**Digitale Modellregion Aachen, Wuppertal,
Gelsenkirchen und Gebiete de KDN. mit
der Leitkommune Gelsenkirchen**

Stand: 15. August 2022

Projektbeginn: 01.04.2019
Projektabschluss: 31.12.2021

Projektkoordinator/in:
Prof. Dr. Norbert Pohlmann
Institut für Internet-Sicherheit – if(is)
Neidenburger Straße 43
45877 Gelsenkirchen
0209 9596 515
pohlmann@internet-sicherheit.de



Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Zusammenfassung (Management Summary) | 3 |
| 2 | Beteiligte Projektpartner | 3 |
| 3 | Ausgangssituation | 4 |
| 4 | Zielsetzung und Vorgehensweise | 4 |
| | 4.1 Zielsetzung..... | 4 |
| | 4.2 Vorgehensweise | 5 |
| | 4.2.1 Arbeitspaket 1: Evaluierung der Open Data Portale..... | 5 |
| | 4.2.2 Arbeitspaket 2: Erstellen der Spezifikation | 6 |
| | 4.2.3 Arbeitspaket 3: Evaluierung Softwarekomponenten..... | 6 |
| | 4.2.4 Arbeitspaket 4; Umsetzung der evaluierten Softwarekomponenten..... | 6 |
| | 4.2.5 Arbeitspaket 5: Integration der Softwarekomponenten und abschließende Test . | 7 |
| | 4.3 Zeitplan | 8 |
| 5 | Ergebnisse | 9 |
| | 5.1 Projektergebnisse | 9 |
| | 5.2 Lehre..... | 9 |
| | 5.3 Verwertung und Anschlussfähigkeit..... | 10 |
| 6 | Unterschrift Zuwendungsempfänger – Institut für Internet-Sicherheit – if(is) | 11 |

1 Zusammenfassung (Management Summary)

Das Förderprogramm „Digitale Modellregionen“ verfolgt das Ziel, mit übertragbaren Lösungen im Bereich „E-Government“ und innovativen Projekten mit der Wirtschaft im Bereich „digitale Stadtentwicklung“ die Digitalisierung in den Regionen und Kommunen in NRW zu beschleunigen. Im Mittelpunkt der Programmumsetzung stehen neben der Entwicklung und Umsetzung digitaler Pilotprojekte auch der kontinuierliche Erfahrungsaustausch mit den Kommunen in NRW und das Zurverfügungstellen von übertragbaren Lösungen und nachnutzbaren Elementen.

„**Interaktive Visualisierung von Open Data - IVOD**“ hat sich zur Aufgabe erklärt, die Landschaft der offenen Daten durch die Erstellung einer Schnittstellen Spezifikation mitzugestalten, die zukünftig die Erstellung und Verbreitung offener Daten verbessert.

Das Projekt fokussiert sich auf die Aufgabe, Interaktive Visualisierung regionaler Open Data anhand der Modellstädte Aachen, Wuppertal und Gelsenkirchen verfügbar zu machen. Ein automatisiertes Erkennen von Darstellungsmöglichkeiten anhand der verfügbaren Datensätze, sowie automatisches Generieren dieser Darstellungsform zur Einbindung in Webdokumente ist durch die Spezifizierung einer Softwareschnittstelle sowie der demonstrativen Umsetzung anhand einer Webapplikation durchgeführt worden. Dies beinhaltet das Formulieren von Richtlinien zur Bereitstellung von Open Data Datensätzen zur schnellen, maschinellen Weiterverarbeitung, ohne vorherige manuelle Anpassung (Formatierung) dieser Datensätze.

2 Beteiligte Projektpartner

Angaben zum Projektkoordinator

Firma/Organisation: Institut für Internet-Sicherheit – if(is)

Neidenburger Straße 43

45877 Gelsenkirchen

Ansprechpartner: Prof. Dr. Norbert Pohlmann

0209 9596 515

pohlmann@internet-sicherheit.de

3 Ausgangssituation

Aufgrund der stark zunehmenden Digitalisierung in allen Branchen- und Geschäftsbereichen, insbesondere in den Bereichen e-Government und Smart City in den Regionen entwickeln sich die Möglichkeiten der Datenverarbeitung rasant weiter. Die Anforderungen an die Umsetzung des Digitalisierungsprozesses steigen. Gesetzliche Open-Data-Regelungen (§ 12 a EGovG) verpflichten Behörden der unmittelbaren Bundesverwaltung die von ihnen erhobenen unbearbeiteten Rohdaten zu veröffentlichen. Diese Daten können von jedermann uneingeschränkt genutzt werden. So lassen sich die erhobenen Daten in weiteren Prozessen durch Dritte verarbeiten. Das Gesetz umfasst vor allem die entgeltfreie Bereitstellung der Daten und schreibt einen freien, uneingeschränkten Zugang sowie die Möglichkeit der Maschinenlesbarkeit vor. Eine zeitgemäße Veröffentlichung von offenen Daten stellt neue Anforderungen an die Bundesverwaltung, denn die Anpassung von Strukturen und Prozesse ist unabdingbar.

Offene Daten sind jedoch nicht nur auf Behörden beschränkt. So steht es jeder Privatperson und Unternehmen frei selbst Datensätze unter den Open Data Kriterien zu veröffentlichen. Um BürgerInnen sowie MitarbeiterInnen von KMUs, Behörden und Organisationen eine direkte Hilfestellung bei der Freigabe solcher Daten zu geben, werden Regelungen zur Struktur bei Veröffentlichung der Rohdaten ermittelt. Denn Zugriff und Freigabe offener Daten geschieht bereits umfänglich, die Maschinenlesbarkeit stellt die Bearbeiter noch immer vor Herausforderungen. Viele Datensätze offener Daten enthalten zusätzliche, nur von Menschen lesbare Informationen direkt als Teil dieses Datensatzes, sind unstrukturiert oder liegen in Dateiformaten vor, die eine Maschinenlesbarkeit stark erschweren. Eine Regelung zur Formatierung offener Daten, sowie die Art ihrer Veröffentlichung und Präsentation kann hier Abhilfe schaffen.

4 Zielsetzung und Vorgehensweise

4.1 Zielsetzung

Um den Behörden einen erleichterten Umgang mit ihren Rohdaten zu ermöglichen, wird daher erforscht, inwieweit sich eine allgemein verständliche Struktur und Regelung für die Freigabe offener Daten etablieren lässt. Zu diesem Zweck werden die Open Data Portale der Modellstädte Aachen, Gelsenkirchen und Wuppertal auf ihre Datensätze hin untersucht. Das Ergebnis der Untersuchung gibt Aufschluss über die Tauglichkeit der bestehenden Daten und liefert Eindrücke zur Formulierung der Regelung. Das Ergebnis wird in einer Spezifikation zur Bereitstellung offener Daten festgehalten, die als Grundlage für eine Standardisierung dient. So ließen sich die noch immer unbearbeiteten Rohdaten in einer reinen, definierten Datenstruktur veröffentlichen, während die Erläuterungen zu den Daten separat festgehalten und veröffentlicht werden. Werden die Daten in einer wohlgeformten Struktur freigegeben, steht auch einer automatisierten Visualisierung direkt in den Open Data Portalen nichts mehr im Wege. Davon profitieren Bürgerinnen und Bürger, denn es entfällt die sonst aufwendige Aufbereitung der Rohdaten.

Begleitend zur Struktur der Daten muss ein Weg gefunden werden, Datensätze in einem Schnellverfahren automatisch auf ihre Tauglichkeit zu prüfen. Enthalten offene Daten in etwa Zeitreihen, oder ähnlich lineare Verläufe in ihren Datenpunkten, kann direkt zu den Rohdaten bereits ein Diagramm geliefert werden, das unverzüglich im Open Data Portal zur Verfügung steht. Durch Einführen interaktiver Elemente in diese Diagramme, wie etwa klickbare Buttons zur Navigation durch den Datensatz, können die Rohdaten in beliebiger Länge vorhanden sein, ohne die automatische Generierung zu behindern. Eine interaktive Visualisierung lädt den Betrachter zudem ein, sich intensiver mit den Daten zu beschäftigen und sie zu „erforschen“. Das Projekt befasst sich daher mit einer Vielzahl möglicher Visualisierungen und der automatisierten Erzeugung dieser. Zu diesem Zweck sollen in der frühen Evaluierungsphase des Projektes Ansätze und Methodiken der künstlichen Intelligenz auf ihre Tauglichkeit bei der Umsetzung der Softwarekomponenten geprüft werden. Es werden im Projektverlauf Softwarekomponenten entwickelt, um einen Datensatz anhand dieser Regelung einer oder mehrerer geeigneter Visualisierungen zuzuordnen. Eine weitere Softwarekomponente erstellt dann anhand der Auswahl eine interaktive Visualisierung. Ziel ist es, diese Komponenten so zu entwickeln, dass sie sich in bestehende Open Data Portale einbinden lassen und die Rohdaten automatisch visualisieren. Nutzer können dabei jederzeit separat auf die Rohdaten zugreifen.

Ziel des Projekts ist es, den Bereitstellungsprozess von offenen Daten erheblich zu erleichtern und für Bürgerinnen und Bürger anschaulich darzustellen. Die Open Data Portale der Modellregionen und das des RVR werden Softwarekomponenten integrieren, um ihre Daten in interaktiven Diagrammen anzubieten, die dann durch die Nutzer erforscht werden können. Dies führt zu einer signifikanten Steigerung der Vertrauensbasis von Bürgern und Behörden, da die dort erhobenen Daten in nachvollziehbarer Form zur Verfügung stehen.

4.2 Vorgehensweise

Das Projekt gliedert sich in mehrere Arbeitspakete:

4.2.1 Arbeitspaket 1: Evaluierung der Open Data Portale

Zum Erreichen des Ziels sind die bisher verfügbaren Datensätze auf den Open Data Portalen der im Antrag genannten Regionen auf ihre Beschaffenheit zu untersuchen. Diese Untersuchung wurde händisch durch die Projektmitarbeiter vorgenommen, da hier besonderes Augenmerk auf die Lesbarkeit der Datensätze gelegt wurde. Im Fokus stehen die Anforderungen an die maschinelle Verarbeitung der verfügbaren Datensätze. Dazu war es nötig, die Beschaffenheit der Datenqualität zu prüfen. Die Untersuchung der Datensätze ist von allen im Projekt beschäftigten Mitarbeitern erfolgt, da hier eine gemeinsame Diskussionsbasis zu schaffen war, die im späteren Projektverlauf wichtig war. Die Untersuchung konnte im eingeplanten Zeitraum abgeschlossen werden.

4.2.2 Arbeitspaket 2: Erstellen der Spezifikation

Für den Einsatz eines automatisch arbeitenden, visuellen Werkzeugs ist es notwendig, die möglichen Eingabedaten in einer bestimmten Güte vorzulegen. Im Rahmen dieses Arbeitspakets wurde eine Regelstruktur zur Veröffentlichung offener Daten formuliert. Anhand dessen sollen sich zukünftige Datensätze mit einer erhöhten Datengüte im Hinblick auf die maschinelle Verarbeitung erzeugen lassen. Die Strukturierung zur Veröffentlichung stellt einen wichtigen Schritt in Richtung Standardisierung dar. Die Projektplanung sieht vor, eine prototypische Kategorisierung verschiedener Datensätze je nach Eignung zur Visualisierung zu erstellen. Diese wurde fortschreitend im Projektverlauf angepasst, um neue Erkenntnisse während des Projektverlaufs, gerade während der Implementierung von Softwarekomponenten, anhand der Anforderungen einfließen zu lassen. Die Arbeiten an der Strukturierung sind vorwiegend durch die beiden wissenschaftlichen Mitarbeiter erfolgt. Die Anforderungen an die Datenstrukturen sind dadurch je nach Kategorie und möglicher Visualisierung festgehalten worden. Das Arbeitspaket ist daher parallel zu den folgenden Arbeitspaketen bis einschließlich AP4 fortgeführt und abgeschlossen worden.

4.2.3 Arbeitspaket 3: Evaluierung Softwarekomponenten

Die benötigten Softwarekomponenten lassen sich in zwei Modulen abbilden, eines zur Analyse der Datenstruktur und Empfehlung einer Visualisierung basierend auf Ansätzen der künstlichen Intelligenz. Ein weiteres Modul zum Darstellen der geeigneten Visualisierung selbst. Dazu ist es nötig, Recherchen zum Einsatz von Software-Frameworks, geeigneter Programmiersprachen und dem geplanten Einsatzumfeld durchzuführen. Diese Arbeiten sind allen im Projekt beschäftigten Mitarbeitern übertragen worden. Die Recherchen beliefen sich auf verfügbare Softwarekomponenten im Internet, sowie Methodiken in verfügbaren Publikationen und Fachbüchern. So ließen sich Anforderungen an die Entwicklung formulieren, sowie geeignete Maßnahmen zur Umsetzung definieren.

4.2.4 Arbeitspaket 4; Umsetzung der evaluierten Softwarekomponenten.

Die Umsetzung der benötigten Softwarekomponenten setzt sowohl die Erweiterung bestehender Software voraus, sowie die Neuentwicklung von Software anhand der ermittelten Anforderungen. Die ursprüngliche Planung sah einen Start des Arbeitspakets erst nach vollständigem Abschluss der Arbeitspakete 1 – 3 vor. Da jedoch vor allem AP2 eine ständige Anpassung im Projektverlauf erfährt, ist die Entwicklung vorgezogen worden, da hier eine gemeinsame Programmierbasis für alle Projektmitarbeiter geschaffen wurde. Als Besonderheit hat sich ergeben, dass eine der studentischen Hilfskräfte eine Erweiterung der bestehenden Softwarekomponente

zum Visualisieren von Datenpunkten um eine Geodarstellung im Rahmen einer Abschlussarbeit erweitert.

Aufgrund der Covid-19 Pandemie haben sich die Arbeiten an den Softwarekomponenten aufgrund der Neuorganisation in Telearbeit verzögert, konnten aber gegen Ende des Projektes in einer Programmierschnittstelle zusammengeführt werden, die sowohl Online als auch in Form von Quellcode, durch das Institut für Internet-Sicherheit bereitgestellt wurde.

Die Komponenten wurden als Schnittstelle im Internet implementiert und zur Demonstration freigegeben. Diese Schnittstelle orientiert sich an der erstellten Spezifikation zur Veröffentlichung offener Daten. Die erstellten Komponenten sind abgeschlossen und im Rahmen einer Präsentation den beteiligten Partnern aus Gelsenkirchen und dem KDN per Videokonferenz demonstriert worden.

4.2.5 Arbeitspaket 5: Integration der Softwarekomponenten und abschließende Test

Die Möglichkeiten zur Integration in die bestehenden Portale ist durch die während des Projekts bestehende Covid-19 Pandemie nur theoretisch erfolgt, da die Koordination mit den entsprechenden Stellen nur per Telefonkonferenz erfolgen konnte. Die Integration der Softwarekomponenten wurde daher in einer eigenen Webapplikation durchgeführt, die eine Integration der erstellten Softwarekomponenten demonstriert.

5 Ergebnisse

5.1 Projektergebnisse

Eine Beispielimplementierung zur Einbettung in Webdokumente liegt in Form eines Demonstrators vor. Dieser basiert auf dem geläufigen Web-Framework React. Sie dient auch als visuelle Schnittstelle zur Administration. Die Dokumentationen des Konzepts wurde in Form einer Schnittstellenspezifikation zur Verfügung gestellt, die Lösungen zur Optimierung von Open Data zeigt und die Funktionsfähigkeit der Schnittstelle abbildet.

Eine bereits existierende Visualisierungskomponente (pive) wurde angepasst, erweitert und anschließend in eine auf Django REST Framework basierende Servekomponente eingebettet.

Dieses Dokument zur Spezifikation der eingehenden Datenstrukturen zur automatisierten Visualisierung, die Anleitung zum eigenen Betrieb und zur Nutzung der Schnittstelle, ist im Rahmen des Projekts „Interaktive Visualisierung von Open Data – IVOD“, erstellt worden. Die Schnittstelle wickelt den Prozess der automatisierten Erzeugung interaktiver Visualisierungen, sowie dem Vorschlagsystem zur Ermittlung geeigneter Visualisierungsmöglichkeiten bei bestimmten Eingabedaten ab. Die Art der Eingabedaten wird hier im Detail erläutert. Das Dokument richtet sich an Entwickler und Systemintegratoren, die einen fortlaufenden Betrieb der Schnittstelle sowie deren Weiterentwicklung unterstützen möchten.

Die Schnittstelle selbst kapselt die Funktionen der im Projekt erweiterten Open Source Software „pive“ (<https://github.com/daboth/pive>) und übernimmt die Bereitstellung der dort bestehenden Funktionen. Sie wurde zudem nach modernem Standard um IT-Sicherheitsfunktionalitäten ergänzt. Die Schnittstelle selbst wird unter der offenen BSD 2-Clause "Simplified" License veröffentlicht.

Der Quellcode der Schnittstelle wird nach Projektabschluss bereitgestellt als Open Source auf dem Portal Github unter der URL: <https://github.com/internet-sicherheit/ivod-platform> und kann im Anschluss an das Projekt von jedermann weiterentwickelt werden.

Dokumente und Quellcode sind in einer passwortgeschützten Dateifreigabe dem KDN und der Dienststelle in Gelsenkirchen direkt nach Ende des Projekts übermittelt worden.

5.2 Lehre

Im Projektverlauf konnte begleitend zur Lehre an der Westfälischen Hochschule eine Abschlussarbeit mit dem Grad Bachelor of Science entstehen. Des Weiteren sind zwei Seminararbeiten zum Projektthema durch Studierende abgeschlossen worden. Es wurde außerdem die Durchführung eines Softwareprojektes zum Thema Open Data begleitet. Die Ergebnisse sind in das Projekt eingeflossen.

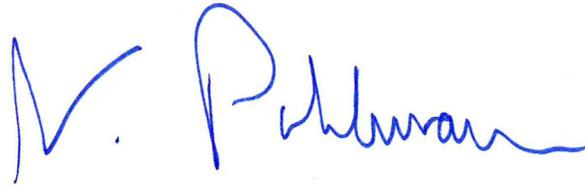
5.3 Verwertung und Anschlussfähigkeit

Durch das Projekt wurde eine zeitgemäße Veröffentlichung von offenen Daten angestrebt. Viele Datensätze offener Daten enthalten zusätzliche, nur von Menschen lesbare Informationen, sind unstrukturiert oder liegen in Dateiformaten vor, die eine Maschinenlesbarkeit stark erschweren. Offene Daten sind in vielen Formen verfügbar, denn Behörden sind lediglich dazu verpflichtet offene Daten bereitzustellen, jedoch ohne festgelegte Form.

Deshalb wurde im Rahmen des Projektes eine Regelung zur Struktur der Veröffentlichung von Rohdaten ermittelt, wodurch sich diese in reinen, definierten Datenstruktur veröffentlichen lassen. Die Erläuterungen zu den Daten werden in der Schnittstellenspezifikation festgehalten und veröffentlicht. Wenn die Daten in dieser vorher definierten Struktur zur Verfügung gestellt werden, kann auch eine automatisierte Visualisierung des Datensatzes direkt in den Open Data Portalen vorgenommen werden.

Die Ergebnisse können durch den Einsatz als Open Source von freien Entwickler:innen beliebig angepasst und weiterentwickelt werden. Hierdurch können diverse Folgeprojekte angestoßen werden, da Innovationen unabhängig entstehen und durch jedermann weiterentwickelt werden. Die Darstellung als Open Source Projekt trägt zudem maßgeblich zur Transparenz zwischen Behörden und Bürgern bei.

6 Unterschrift Zuwendungsempfänger – Institut für Internet-Sicherheit – if(is)



Gelsenkirchen, 15.08.22

Ort, Datum

Norbert Pohlmann