Band 17 / 2024

GeoCROW:
Geosemantische Datenanalyse und Virtual Reality-gestützte Visualisierung für humanitäre Einsätze im digitalen Zeitalter

Johannes Göllner, Johannes Scholz (Hrsg.)

Schriftenreihe der
Landesverteidigungsakademie

UNSER HEER

Schriftenreihe der
Landesverteidigungsakademie

Johannes Göllner, Johannes Scholz (Hrsg.)

GeoCROW:

Geosemantische Datenanalyse und Virtual Reality-gestützte Visualisierung für humanitäre Einsätze im digitalen Zeitalter

17/2024
Wien, Dezember 2024

Impressum:

Medieninhaber, Herausgeber, Hersteller:
Republik Österreich / Bundesministerium für Landesverteidigung
Rossauer Lände 1
1090 Wien

Redaktion:
Landesverteidigungsakademie
Zentraldokumentation
Stiftgasse 2a
1070 Wien

Schriftenreihe der Landesverteidigungsakademie

Copyright:
© Republik Österreich / Bundesministerium für Landesverteidigung
Alle Rechte vorbehalten

Dezember 2024
ISBN 978-3-903359-02-2

Druck:
ReproZ W 24-5932
Stiftgasse 2a
1070 Wien

Inhaltsverzeichnis

Inhaltsverzeichnis 3
Vorwort 5
Abstract 7
Einführung 9
Teil A: Rahmenbedingungen für GeoCROW 11
1. Fallbeispiele 12
1.1 Niger 12
1.2 Türkei 13
2. Unterliegender Prozess bei humanitären Missionen 15
3. Rechtliche Rahmenbedingungen 18
4. Ethische Rahmenbedingungen 25
5. Technische Rahmenbedingungen 32
Teil B: Die GeoCROW-Systemarchitektur 33
1. Architektur 35
2. Komponenten 39
2.1 Workflow Engine 39
2.2 Datenbank und Datenbank-Kommunikationsservice 39
2.3 Komponenten zur Datengewinnung und Vorbereitung 40
2.4 Services zur Event Extraktion 42
2.5 Komponenten zur Aggregierung und Zusammenfassung 43
Teil C: Das GeoCROW-System 45
1 Datenanalyse 47
1.1 Analyse eines Newsitems 47
1.2 Die GeoCROW-Ontologie 48
2 Einordnung der Kernfunktionen des GeoCROW-Systems 52
2.1 Extraktion der räumlichen Komponenten 52
2.2 Extraktion von temporären Komponenten 54
2.3 Extraktion von Ereignistypen 54
2.4 Zusammenfassung zu Ereignissen und Events 56
3 Visualisierung der extrahierten Informationen 57
Konklusion 67
Literaturverzeichnis 69
Abbildungsverzeichnis 75
Autoren / Herausgeber und Autoren 77

Vorwort

In einer Welt, die durch zunehmende Unsicherheiten, Umwelt- und Naturkatastrophen und multiple und komplexe Krisensituationen geprägt ist, stehen humanitäre Organisationen und Einsatzkräfte vor immer größeren Herausforderungen. Die Fähigkeit, schnell und präzise auf diese Ereignisse zu reagieren, ist entscheidend, um Menschenleben zu retten und Schäden zu minimieren. In diesem Kontext bietet das GeoCROW-Projekt einen innovativen Ansatz, der moderne Technologien wie semantische Analyse, Crowdsourcing und Virtual Reality kombiniert, um die Entscheidungsfindung in Krisensituationen zu verbessern.

Dieses Vorhaben ist das Ergebnis einer engen Zusammenarbeit zwischen Forschungsinstituten, technologischen Partnern und humanitären Organisationen. Ziel ist es, ein System zu entwickeln, das große Mengen an Daten aus verschiedenen Quellen sammelt, analysiert und aufbereitet, um eine umfassende Lageerfassung zu ermöglichen. GeoCROW hebt sich durch die Integration von Virtual Reality besonders hervor, indem es eine immersive Visualisierung der gesammelten Daten bietet. Dies ermöglicht es den Einsatzkräften, sich ein realistisches Bild der Situation vor Ort zu machen, bevor sie tatsächlich in das Einsatzgebiet aufbrechen.

Das vorliegende Dokument beschreibt die Ergebnisse des GeoCROW-Projekts und gibt einen detaillierten Einblick in die technologischen, rechtlichen und ethischen Aspekte, die während der Entwicklungsphase berücksichtigt wurden. Es verdeutlicht, wie moderne Technologien dazu beitragen können, humanitäre Einsätze effizienter und sicherer zu gestalten.

Die Motivation für dieses Projekt ergibt sich aus der wachsenden Komplexität humanitärer Einsätze. Naturkatastrophen wie Erdbeben, Überschwemmungen und Wirbelstürme sind oft unvorhersehbar und erfordern eine schnelle Reaktion. Gleichzeitig führen politische Unruhen und Konflikte zu instabilen Situationen, die eine umfassende Analyse und gezielte Intervention erfordern. GeoCROW wurde entwickelt, um diesen Herausforderungen zu begegnen, indem es aktuelle Technologien nutzt, um schnelle, präzise und kontextualisierte Informationen bereitzustellen.

Der Einsatz von Daten aus sozialen Medien, Nachrichtenquellen und anderen digitalen Plattformen stellt eine wertvolle Ergänzung zu traditionellen Informationsquellen dar. Allerdings bringt die Nutzung dieser Daten auch

neue Herausforderungen mit sich, insbesondere im Hinblick auf Datenschutz und ethische Fragen. GeoCROW hat sich von Beginn an darauf konzentriert, diese Herausforderungen durch klare rechtliche Rahmenbedingungen und ethische Richtlinien zu adressieren.

Die Entwicklung eines so komplexen Systems wie GeoCROW erfordert nicht nur technisches Know-how, sondern auch eine enge Zusammenarbeit zwischen verschiedenen Disziplinen und Akteuren. Im Laufe des Projekts haben verschiedene Akteure gemeinsam daran gearbeitet, eine Lösung zu schaffen, die den realen Bedürfnissen vor Ort entspricht. Diese interdisziplinäre Zusammenarbeit hat es ermöglicht, innovative Technologien wie maschinelles Lernen, semantische Datenverarbeitung und Virtual Reality in einem System zu vereinen, das sowohl leistungsfähig als auch praxisnah ist.

Ein Projekt wie GeoCROW wäre ohne die Unterstützung und das Engagement zahlreicher Partner und Mitwirkender nicht möglich gewesen. Wir möchten an dieser Stelle unseren Dank an alle beteiligten Institutionen, Organisationen und Einzelpersonen aussprechen. Ihr Beitrag war entscheidend für den Erfolg des Projekts.

Das vorliegende Dokument bietet eine umfassende Darstellung der Ergebnisse des GeoCROW-Projekts und zeigt die Möglichkeiten auf, die moderne Technologien für die humanitäre Hilfe bieten. Gleichzeitig wird deutlich, dass GeoCROW nicht das Ende, sondern vielmehr ein Anfang ist. Wir hoffen, dass dieses Dokument nicht nur einen Einblick in die technische und wissenschaftliche Arbeit von GeoCROW gibt, sondern auch als Inspiration für zukünftige Projekte dient, die sich der Bewältigung globaler Herausforderungen widmen.

Johannes Scholz,
Paris-Lodron-Universität-Salzburg und Projektleiter GeoCROW
Johannes Göllner,
Landesverteidigungsakademie des BMLV und Bedarfsträgervertreter

Abstract

Das GeoCROW-Projekt (GeoSemantic and Crowdsourced enhanced Virtual Reality for Situational Awareness) kombiniert semantische Technologien, Crowdsourcing und Virtual Reality, um die Entscheidungsfindung und Einsatzplanung in humanitären Krisensituationen zu unterstützen. Ziel des Projekts ist die Entwicklung eines Systems, das große Mengen an unstrukturierten Daten aus verschiedenen Quellen wie Nachrichtenmedien, sozialen Netzwerken und geografischen Datenbanken automatisch sammelt, analysiert und visualisiert. Das Ergebnis ist ein umfassendes Lagebild, das Einsatzkräfte nutzen können, um sich effektiv auf Krisensituationen vorzubereiten und fundierte Entscheidungen zu treffen.

Der Kern des GeoCROW-Systems besteht aus einer modularen Architektur, die auf Microservices basiert. Zu den Hauptkomponenten gehören ein Geoparser zur Identifikation und Georeferenzierung von Ortsangaben, eine temporale Extraktionseinheit zur Identifizierung von Zeitangaben und ein Ereignistyp-Extraktionsservice, der Ereignisse wie Naturkatastrophen oder soziale Unruhen klassifiziert. Die extrahierten Informationen werden in einem Wissensgraphen gespeichert, der auf einer speziell entwickelten Ontologie basiert. Dieser Wissensgraph ermöglicht es, Beziehungen zwischen den Daten zu erkennen und komplexe Zusammenhänge transparent darzustellen.

Ein zentrales Feature des Projekts ist die Visualisierung der gesammelten Daten in einer Virtual-Reality-Umgebung. Diese immersive Darstellung bietet Einsatzkräften eine intuitive Möglichkeit, sich in das Einsatzgebiet „einzuleben“, bevor sie vor Ort eintreffen. Die VR-Umgebung ist in zwei Modi verfügbar: Im „Roomscale“-Modus können Nutzer eine Übersichtskarte erkunden, während der „Lifescale“-Modus eine detaillierte Erkundung einzelner Gebiete in Originalgröße ermöglicht.

Das System wird an zwei Fallstudien getestet: in Niger, wo es zur Analyse von sozialen und politischen Spannungen zum Einsatz kommt und in der Türkei, wo es zur Analyse der Lage nach dem schweren Erdbeben im Februar 2023 eingesetzt wird. Neben den technologischen Fortschritten berücksichtigt das Projekt auch rechtliche und ethische Rahmenbedingungen. Der Schutz personenbezogener Daten und die Einhaltung von Datenschutzgesetzen spielen eine zentrale Rolle. Durch die Anwendung von „Privacy by

Design“-Ansätzen wird sichergestellt, dass sensible Daten angemessen verarbeitet und Grundrechte gewahrt werden.
GeoCROW bietet eine vielversprechende Lösung, um den steigenden Anforderungen in der humanitären Hilfe gerecht zu werden. Es zeigt, wie moderne Technologien die Lageerfassung und Einsatzvorbereitung verbessern können, ohne dabei ethische und rechtliche Aspekte zu vernachlässigen.

Einführung

In einer zunehmend komplexen und vernetzten Welt stellt die effiziente Bewältigung humanitärer Krisen eine zentrale Herausforderung dar. Umwelt- und Naturkatastrophen, politische Unruhen und andere Krisensituationen erfordern eine schnelle, fundierte Entscheidungsfindung, um Menschenleben zu retten und die Infrastruktur wiederherzustellen. In diesem Kontext gewinnen die Verfügbarkeit und Analyse relevanter Daten eine entscheidende Bedeutung. Das GeoCROW-Projekt setzt genau hier an und entwickelt ein System, das semantische Technologien, Crowdsourcing und Virtual Reality (VR) kombiniert, um Einsatzkräfte in Krisengebieten zu unterstützen.

GeoCROW steht für „GeoSemantic and Crowdsourced enhanced Virtual Reality for Situational Awareness" und verfolgt das Ziel, humanitäre Organisationen und militärische Einheiten bei der Lageerfassung und Einsatzplanung zu unterstützen. Durch die automatisierte Analyse großer Datenmengen aus diversen Quellen wie Nachrichtenportalen, sozialen Medien und geografischen Datenbanken können umfassende Lagebilder erstellt werden. Diese Daten werden durch VR-Umgebungen visualisiert, was den Nutzern eine interaktive und immersive Erfahrung bietet. Dies soll helfen, kritische Informationen schneller zu verstehen und fundierte Entscheidungen zu treffen.

Die Motivation für das GeoCROW-Projekt entspringt der wachsenden Zahl von Krisensituationen weltweit. Die Herausforderung liegt darin, in diesen dynamischen und oft unübersichtlichen Situationen relevante Informationen zu sammeln, zu analysieren und bereitzustellen. Moderne Technologien wie Machine Learning, Natural Language Processing (NLP) und Virtual Reality bieten neue Möglichkeiten, große Datenmengen effizient zu verarbeiten und nutzbar zu machen. GeoCROW kombiniert diese Technologien, um ein System zu schaffen, das Informationen nicht nur analysiert, sondern sie auch in einer benutzerfreundlichen und zugänglichen Weise präsentiert.

Das Hauptziel von GeoCROW ist es, ein innovatives System zu entwickeln, das die situative Wahrnehmung in humanitären Krisen verbessert. Konkret verfolgt das Projekt folgende Teilziele:

1. **Automatisierte Datensammlung und -analyse**: Entwicklung von Algorithmen zur Extraktion von relevanten Informationen aus unstrukturierten Datenquellen. Dazu gehören Nachrichtenartikel, Social-Media-Posts und andere digitale Inhalte, die wichtige Hinweise zu Ereignissen in Krisengebieten enthalten.

2. **Erstellung eines Wissensgraphen**: Aufbau einer Ontologie, die es ermöglicht, die extrahierten Informationen strukturiert zu speichern und Beziehungen zwischen verschiedenen Datenpunkten sichtbar zu machen.

3. **Immersive Datenvisualisierung**: Implementierung von VR-Technologien, um die gesammelten Daten in einer virtuellen Umgebung darzustellen. Dies ermöglicht den Nutzern eine intuitive und interaktive Erkundung der Lage vor Ort.

4. **Unterstützung der Einsatzplanung**: Bereitstellung präziser und aktueller Informationen, die die Entscheidungsfindung und operative Planung in Krisengebieten erleichtern.

GeoCROW leistet einen wichtigen Beitrag zur Verbesserung der humanitären Hilfe. Durch die Kombination modernster Technologien schafft es ein System, das nicht nur die situative Wahrnehmung in Krisengebieten verbessert, sondern auch die Sicherheit und Effizienz von Einsätzen erhöht. Langfristig könnte GeoCROW als Modell für ähnliche Systeme dienen und in internationalen Netzwerken der Katastrophenhilfe eine zentrale Rolle spielen.

Die vorliegende Arbeit ist in mehrere Abschnitte unterteilt. Zunächst werden die Rahmenbedingungen des Projekts beschrieben, darunter rechtliche, ethische und technische Aspekte. Im Anschluss wird die Systemarchitektur von GeoCROW detailliert erläutert. Der dritte Abschnitt widmet sich den zentralen Funktionalitäten des GeoCROW-Systems. Abschließend folgt eine Konklusion, die die wichtigsten Erkenntnisse zusammenfasst und einen Ausblick auf zukünftige Entwicklungen gibt.

Teil A: Rahmenbedingungen für GeoCROW

1. Fallbeispiele

Die folgenden beiden Kapitel gehen auf die beiden Fallbeispiele ein, die im Projekt GeoCROW behandelt werden. Dabei handelt es sich zum Einen um Niger in Nordafrika mit einem Fokus auf soziale Unruhen im Zuge des Militärputschs im Juli 2023, zum Anderen um die Erdbeben im türkisch-syrischen Grenzbereich im Februar 2023.

1.1 Niger

Der Binnenstaat Niger im Zentrum Nordafrikas umfasst eine Fläche von 1,267,000 km^2. Die Hauptstadt Niamey liegt im Westen des Landes. Insgesamt hat das Land eine Einwohnerzahl von ca. 17.13 Millionen Einwohnern. Außerdem zählt es zu den ärmsten Ländern der Welt. An das Land Niger grenzen die afrikanischen Staaten Libyen, Tschad, Nigeria, Benin, Burkina Faso, Mali und Algerien.
Bei Niger handelt es sich um einen multiethnischen Staat, der mit 53% hauptsächlich durch die Gruppierung der Haussa geprägt ist. Danach folgen Gruppierungen wie die Djerma/Songhai, Fulbe, Tuareg und Kanuri. Aus religiöser Perspektive prägen das Land sunnitische Muslime (98%). Das Christentum und andere Religionen teilen sich die anderen 2%. Als Amtssprache wird in Niger französisch verwendet, wobei sieben weitere Sprachen offiziell anerkannt sind (z.B.: Arabisch, Fulfude, Zorma).
Aus infrastruktureller Sicht ist Niger durch ein gering asphaltiertes Straßennetz geprägt. Dazu kommt, dass kein Eisenbahnnetz und keine nationale Fluglinie vorhanden sind. Zentrale Handelspartner des Nigers stellen Frankreich und China dar. Dabei zeichnet sich das Land vor allem durch die Agrarwirtschaft und den Uranbergbau aus.
Die sozialen Spannungen im Niger kommen vor allem von extremen materiellen Ungleichverteilungen und ethnischen Konflikten. Zusätzlich werden durch Dürreperioden Hungersnöte ausgelöst. Das Land wird auch von islamistischen Gruppierungen als Rückzugsort genutzt. Weitere gefahren für ein Einsatzkontigent im Niger können durch Dürren, Überschwemmungen, Sand-/Staubstürme, Hitze, Schwüle etc. entstehen.
Das im Projekt GeoCROW behandelte Einsatzszenario für das Land Niger fokussiert sich auf eine humanitäre Mission mit einer langen Vorbereitungszeit von bis zu 6 Monaten. Dabei sollen vor allem Informationen zur politischen und sozialen Gesamtsituation gesammelt werden, um Gefahren für

das Einsatzkontigent frühzeitig zu erkennen.
Radio und Fernsehen stellen die beliebtesten traditionellen Medientypen des Landes dar, weil im Niger mit 38% eine noch immer hohe Analphabetenrate vorliegt.[1] [2] Der World Press Freedom Index im Niger ist mit 60/100 im Jahr 2024 zwar verbesserungswürdig, aber besser als in anderen afrikanischen Staaten (z.B.: Sudan 36/100, Libyen 41/100).[3] Die relevantesten Printmedien im Niger sind die staatliche Tageszeitung Le Sahel, sowie die nationale Nachrichtenagentur von Niger, als auch L'Enqueteur. Des Weiteren gibt es ein breites Angebot an digitalen Medien und Nachrichtenseiten, wie etwa Actu-Niger, NigerInter, und Tamtaminfo.[4] Die Hauptberichtserstattungssprache ist Französisch. Berichterstattungen in anderen Nationalsprachen (z.B.: Hausa) sind nur gering ausgeprägt. Auch internationale Medien, wie Africanew, BBC Afrique und Aljazeera spielen eine Rolle neben den lokalen Medien. Nur etwa 2% der Bevölkerung nutzt aktiv Social Media Plattformen, wobei Facebook mit Abstand am häufigsten genutzt wird.[5] [6]

1.2 Türkei

Die Republik Türkei liegt im Südosten Europas und ihre Hauptstadt ist Ankara. Dort leben auf einer Fläche von etwa 783,562 km^2 ca. 83.61 Millionen Einwohner. Mit 15.46 Millionen Einwohnern ist Istanbul die größte Stadt der Türkei. Mit 70-75% bilden die Türken die größte Bevölkerungsgruppe des Landes gefolgt von ca. 19% Kurden, die meist im Südosten des Landes leben. Andere Minderheiten sind beispielsweise die Zaza, Tscherkessen, Araber oder Griechen. Die Amtssprache in der Türkei ist türkisch und wird von etwa 85% der Einwohner gesprochen.
In der Türkei sind individuelle Gewissens- und Glaubensfreiheit garantiert, wobei die Religion strikt dem säkularen Staat untergeordnet ist. Mit 99% ist

[1] BBC, "Niger Media Guide," 2023, https://www.bbc.com/news/world-africa-13943665.
[2] Statistics, UNSECO Institute for; UNESCO Institute for Statistics, "Literacy Rate, Adult Total," 2024, https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=NE.
[3] Reporter ohne Grenzen, "World Press Freedom Index," 2024, https://rsf.org/en/index.
[4] Media Landscapes, "Niger," 2024, https://medialandscapes.org/country/niger.
[5] Simon Kemp, "Digital 2023: Niger," 2023, https://datareportal.com/reports/digital-2023-niger.
[6] Statcounter, "Social Media Stats Niger," 2024, https://gs.statcounter.com/social-media-stats/all/niger/#monthly-201901-202305.

ein Großteil der Einwohner der Türkei muslimisch, wobei die sunnitischen Muslime mit 70-85% den Großteil bilden. Anerkannte religiöse Minderheiten sind beispielsweise das Christentum und das Judentum. Aus geographischer Perspektive ist die Türkei im Norden und Süden von Gebirgsketten umrahmt. Diese Gebiete zeichnen sich auch durch eine erhöhte Erdbebengefahr aus. So kam es erst im Februar 2023 im türkisch-syrischen Grenzgebiet zu verheerenden Erdbeben bei denen mehr als 50,000 Menschen gestorben sind und schwere Schäden entstanden sind.[7] Bei dieser Katastrophe war auch die österreichische Katastrophenhilfeeinheit AFDRU (Austrian Forces Disaster Relief Unit) des Österreichischen Bundesheeres (Kompetenzzentrum: ABC-Abwehr) vor Ort im Einsatz. Das Einsatzszenario für das Projekt GeoCROW unterscheidet sich im Fall der Türkei von dem im Niger in zwei wesentlichen Punkten. Zum Einen gibt es keinerlei Vorlaufzeit und es muss nach dem Auftreten des Erdbebens schnell agiert werden. Zum Anderen handelt es sich hier nicht um soziale Unruhen, sondern um eine Naturkatastrophe. Damit die Einsatzteams sich vor ihrer Abreise effektiv auf das Krisengebiet vorbereiten können, müssen vor allem Informationen zu beschädigter Infrastruktur und Versorgungsknappheiten gewonnen werden.

Die Medienlandschaft in der Türkei zeichnet sich durch einen Trend hin zum Internet aus. Dennoch ist das Fernsehen das führende Nachrichtenmedium.[8] Aufgrund der geringen Pressefreiheit (World Press Freedom Index 34/100) wird die Medienlandschaft vorwiegend von regierungsnahen Outlets dominiert. Kritische Stimmen beschränken sich auf einige wenige Publikationen mit geringer Auflage, oder wählen digitale Medien als Outlet.[9] Zu den relevantesten News Plattformen zählen die Nachrichtenagenturen Doğan Haber Ajansı (DHA), ANKA Haber Ajansı,und die staatliche Anadolu Ajansı(AA), die Zeitungen Hurriyet und Sabah. Neben Türkisch werden viele der Medien auch in englischer Sprache angeboten. Dazu kommen internationale Medien, wie etwa euronews, BBC, und Aljazeera. Auch die

[7] Tagesschau, "Zahl Der Toten Steigt Auf Mehr Als 50.000," 2023, https://www.tagesschau.de/ausland/asien/erdbeben-tote-tuerkei-syrien-101.html.

[8] BBC, "Turkey Media Guide," 2023, https://www.bbc.com/news/world-europe-17992011.

[9] Reporter ohne Grenzen, "World Press Freedom Index."

Nutzung von sozialen Medien ist mit 62 Millionen Nutzern in der Türkei weit verbreitet. Das heißt etwa 73% der Gesamtbevölkerung nutzen Facebook und Co. Lange Zeit dominierte Facebook den Markt für Social-Media-Plattformen in der Türkei. Allerdings hat sich dies in den letzten Jahren gewandelt, wodurch Facebook, Youtube und Instagram nun gleichermaßen beliebt sind. Auch Messaging-Dienste wie Whatsapp, FB Messenger und Telegram verzeichnen hohe Benutzerzahlen.[10] [11] [12]

2. Unterliegender Prozess bei humanitären Missionen

Abbildung 1 stellt den Prozessablauf bei humanitären Missionen dar. Aus der Abbildung kann entnommen werden, dass der Prozess auf dem Völkerrecht basiert und entlang der Vorgaben der INSARAG-Guidelines der Vereinten Nationen und dem Management Verfahren des österreichischen Heers verläuft.
Der Prozess beginnt mit einer Hilfeanfrage des betroffenen Staates an Institutionen wie die UN oder die EU. Diese leiten die Anfrage an potenzielle Unterstützerstaaten weiter, die daraufhin entscheiden, ob sie ihre Hilfe anbieten oder ablehnen. Im Falle einer Zustimmung in Österreich wird die Anfrage an das Bundesministerium für Landesverteidigung (BMLV) und weitere Hilfsorganisationen übermittelt. Das BMLV aktiviert daraufhin die Einsatzteams, wie AFDRU sowie das Institut für Militärisches Geowesen (IMG), welches für die Bereitstellung von Einsatzmaterialien verantwortlich ist. Ferner kann ad hoc auch ein ereignisspezifisches Situation Awareness Center/Portal durch die Abteilung Zentraldokumentation der Landesverteidigungsakademie des BMLV temporär eingerichtet bzw. aktiviert werden. Zu den Vorbereitungen gehören die Zusammenstellung von Informationen über kulturelle, soziale und geografische Bedingungen im Einsatzgebiet. Diese Informationen werden als Berichte, Karten und Briefings aufbereitet und den Teams bereitgestellt. Hierbei kommt das GeoCROW-System zum Einsatz, das automatisch Daten zum Einsatzgebiet sammelt und den Teams

[10] OOSGA, "Social Media in Turkey - 2023 Stats & PLatform Trends," 2023, https://oosga.com/social-media/tur/.
[11] Statcounter, "Social Media Stats Turkey," 2024, https://gs.statcounter.com/social-media-stats/all/turkey.
[12] Datareportal, "Digital 2023: Turkey," 2023, https://datareportal.com/reports/digital-2023-turkey.

in VR zur Vorbereitung zur Verfügung stellt.
Das GeoCROW-System richtet sich sowohl an das Erkundungsteam als auch an das Hauptteam. Das Erkundungsteam reist innerhalb von 3-4 Stunden ins Einsatzgebiet und meldet sich beim On-Site Operations Coordination Centre (OSOCC) der UN-OCHA (Genf), das die Einsatzkräfte auf die betroffenen Gebiete verteilt. Vor Ort erfasst das Erkundungsteam Informationen zu benötigten Hilfsgütern, Einsatzkräften, zum Schadensausmaß und weiteren relevanten Daten.
Das Hauptteam folgt innerhalb von 24 Stunden und meldet sich ebenfalls beim OSOCC. Nach der Zuweisung eines Einsatzgebiets durch das OSOCC beginnt der Einsatz des Hauptteams, der in der Regel nach 100 Stunden abgeschlossen ist.

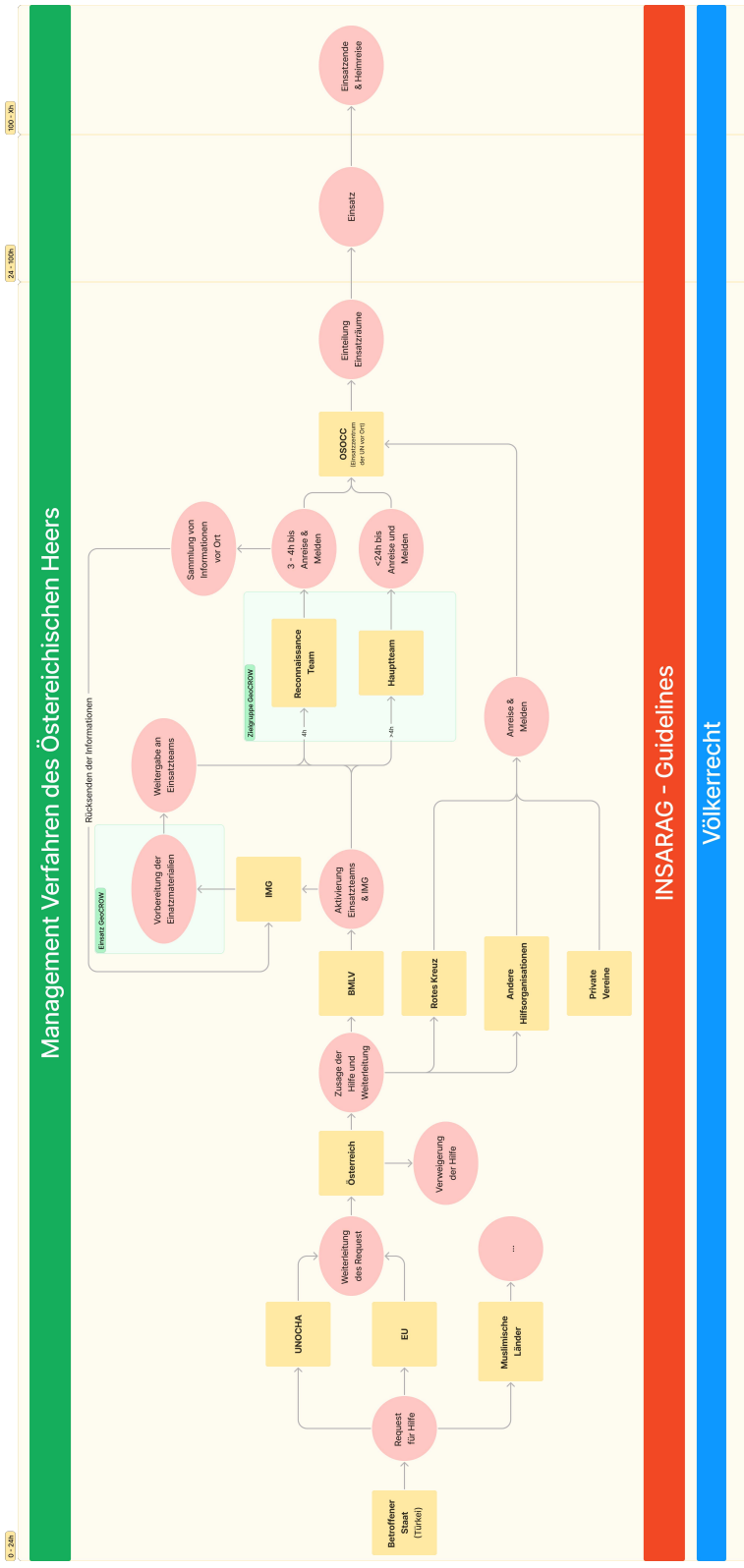

Abbildung 1: Prozessablauf bei humanitären Missionen.

3. Rechtliche Rahmenbedingungen

Ausgehend von unter den Projektpartnern erarbeiteten Parametern wurden bereits früh im Projekt auch die rechtlichen Rahmenbedingungen erarbeitet. In Hinblick auf die rechtlichen Rahmenbedingungen ist besonders zu beachten, dass diese immer nur eine Art *Momentaufnahme* der jeweils aktuellen, bekannten und als wesentlich erachteten Aspekte sein kann, da sich unter Umständen bereits kleine Änderungen stark auf die einschlägige Fakten- bzw. Rechtslage auswirken können. Es handelt sich bei der zu beurteilenden Situation somit gewissermaßen um einen beweglichen Rahmen, wobei sich die relevanten Parameter im Laufe des Projekts mitunter geändert haben und auch weiterhin ändern können (so etwa auch die einschlägige Rechtslage). Im Zuge der rechtlichen Analysen im Rahmen des Projekts konnte demnach lediglich eine möglichst konkrete Beurteilung des aktuellen Stands bzw. der groben rechtlichen Umgebung vorgenommen werden.
Zu beachten ist dabei weiters, dass aus rechtlicher Sicht die Vorgänge im Forschungsprojekt (z.B. Entwicklung und Testbetrieb von Systemen) von einem etwaigen Echtbetrieb (dem tatsächlichen Einsatz) entsprechender Systeme und den damit zusammenhängenden Vorgängen unterschieden werden müssen, weil sich dabei oft unterschiedliche Implikationen ergeben.

Im Rahmen der ersten rechtlichen Analyse wurde der Fokus insbesondere auf wesentliche Rahmenbedingungen in Hinblick auf staatliches Handeln unter dem Gesichtspunkt der Verarbeitung personenbezogener Daten gelegt.
Insbesondere folgende Aspekte wurden als wesentliche Ausgangssituation in dieser ersten rechtlichen Analyse herangezogen: Die zu entwickelnden Systeme sollen von entsprechenden Einheiten des Österreichischen Bundesheers im Rahmen von internationalen (Natur-)Katastrophenhilfeeinsätzen, insbesondere zur Vorbereitung, vor allem entsprechender Aufklärungseinheiten (Vorabtruppen), auf den Einsatz, eingesetzt werden. Dies betrifft insbesondere die Phase der Vorbereitung ab der (politischen) Entscheidung, dass sich die Republik Österreich an einem entsprechenden Auslandseinsatz beteiligt. Ausbildungs- und Übungsmaßnahmen sowie die Einsatzplanung

können dabei mitunter andere rechtliche Implikationen haben. Insbesondere aber der Einsatz unmittelbar zur militärischen Landesverteidigung wurde **nicht** in die zu beurteilende Ausgangssituation einbezogen. Die erste rechtliche Analyse hat sich damit auch vor allem auf einen etwaigen Echtbetrieb entsprechender Systeme bezogen.

In der Folge hat sich zunächst die Frage nach der Einordnung entsprechender (staatlicher) Handlungen – unter dem besonderen Blickwinkel der Verarbeitung personenbezogener Daten – gestellt.

Im gegenständlichen Forschungsprojekt wird „Social-Media-Monitoring" betrieben. Dabei wurde untersucht wie das Social-Media-Monitoring einzuordnen ist und festgestellt, dass grundsätzlich zwischen Monitoring auf Grundlage aggregierter Massendaten und Monitoring ad personam zu unterscheiden ist: Social-Media-Monitoring – das wie im gegenständlichen Projekt auf der Grundlage aggregierter Massendaten erfolgt, konzentriert sein Erkenntnisinteresse auf die Analyse von Stimmungen, Trends und Entwicklungen.
Die Tätigkeit der Internet-Recherche oder Monitorings durch Verwaltungsbehörden ist verwaltungsrechtlich nicht leicht einzuordnen, dabei könnte es sich um schlicht-hoheitliches Verwaltungshandeln handeln, worunter jene Verwaltungsorganhandlungen verstanden werden, die zwar kein Hoheitsakt im engeren Sinne sind, aber eine „funktionelle Verbindung zur staatlichen Hoheitsverwaltung aufweisen". Schlichte Hoheitsverwaltung erfolgt „im Zusammenhang mit" der Hoheitsverwaltung. Sie ist nicht normativer, sondern bloß faktischer Natur, das heißt Handlungen, die nicht als Hoheitsakte, sondern regelmäßig als „Vorstufe" oder „im Kontext" hoheitlichen Verwaltungshandelns zu definieren und damit im „niederschwelligen" Bereich anzusiedeln sind.
Die Zuordnung des Verwaltungshandelns zur Hoheitsverwaltung wirft Fragestellungen betreffend die für diesen Bereich geltenden Anforderungen hinsichtlich des Legalitätsprinzips, der Kompetenzverteilung, der Verfahrensprinzipien sowie der Grundrechtsbindung, aber auch hinsichtlich des Rechtsschutzes und der Amtshaftung auf.
Inwieweit „schlicht-hoheitliches Verwaltungshandeln" dem Geltungsbereich des Art 18 Abs 1 B-VG unterliegt, ist umstritten. In der Lehre wird

dies teilweise bejaht, teilweise verneint. Berka geht hinsichtlich des schlichthoheitlichen Verwaltungshandelns von abgeschwächten Bestimmtheitserfordernissen und einer verringerten Intensität aus. Jedoch bestünden auch im Rahmen der schlichten Hoheitsverwaltung dann erhöhte Bestimmtheitserfordernisse im Sinne eines differenzierten Legalitätsprinzips, wenn es zu Grundrechtseingriffen kommt bzw. kommen kann.
Hier ist darauf hinzuweisen, dass der Einsatz von Social Media-Monitoring ebenfalls in den Bereich der grundrechtssensiblen Tätigkeiten vorstoßen kann. Dies kann insbesondere den Eingriff in das Grundrecht auf Datenschutz betreffen, da regelmäßig eine Verarbeitung von personenbezogenen Daten erfolgt. Wenn sich die Monitoring-Auswertungen auf anonymisierte Inhaltsdaten beschränken, die keinen Rückschluss auf bestimmbare Personen (wie dies häufig bei allgemeinen Stimmungserkundungen und Trendanalyse der Fall ist) zulassen, ist der Anwendungsbereich des Datenschutzrechts (nur) für die Filterung personenbezogener Daten aus dem Datenbestand eröffnet, da das Ziel von Beobachtungen dabei nicht die ad personam erfolgende Inhaltsauswertung, sondern die Aggregierung von Metadaten für überblicksartige Einschätzungen ist.

In weiterer Folge wurden auch die (militär-)verfassungsrechtlichen Grundlagen der internationalen Hilfe, insbesondere durch das Österreichische Bundesheer, und schließlich auch internationale und unionsrechtliche Rahmenbedingungen in diesem Zusammenhang analysiert.
Dabei hat sich gezeigt, dass eine komplexe bzw. vielschichtige rechtliche Situation vorliegt. Dies beginnt mit den verfassungsmäßigen Aufgaben des Bundesheers, denn Auslandseinsätze stellen prinzipiell eine eigene Aufgabe abseits der militärischen Landesverteidigung dar. Allerdings werden entsprechende Planungs- und Vorbereitungsmaßnahmen, ebenso wie sämtliche Ausbildungs- und Übungsmaßnahmen, der *allgemeinen Einsatzvorbereitung* und somit der Aufgabe der militärischen Landesverteidigung zugerechnet. Weiters sind auch die unionsrechtlichen Rahmenbedingungen relevant und gleichzeitig komplex. Hierbei bestehen nämlich verschiedene, potenziell relevante Kompetenzen und Regelungsgebiete. Im Übrigen gestaltet sich auch die „*Landkarte*" der internationalen Organisationen und Mechanismen im relevanten Bereich der internationalen Hilfe durchaus vielschichtig. Dies wurde ebenso dem Grund nach analysiert.

In weiterer Folge ergeben sich – insbesondere auch aus den genannten verfassungs- und unionsrechtlichen Gegebenheiten – differenzierte datenschutzrechtliche Rahmenbedingungen. Zunächst bestehen hierbei auf unionsrechtlicher Ebene neben der DSGVO noch die EU-Verordnung 2018/1725 für personenbezogene Datenverarbeitungen durch die Organe Einrichtungen und sonstigen Stellen der EU sowie die EU-Datenschutz-Richtlinie Polizei und Justiz (RL 2016/680; „DSRL-PJ"), welche potenziell einschlägig sein könnten.
Zusätzlich enthält auf nationaler Ebene vor allem das österreichische Datenschutzgesetz („DSG") relevante Regelungen. Dieses beinhaltet prinzipiell eigene Regelungen über seinen Anwendungsbereich (siehe insbesondere § 4 DSG). Dabei setzt das DSG aber vor allem auch die europäische Rechtslage auf umfassende Weise um, indem auch in Bereichen, in denen die DSRL-PJ bzw. die DSGVO eigentlich nicht anwendbar wären, entsprechende Regelungen dennoch für anwendbar erklärt werden. Denn die nationale Sicherheit und damit – nach dem Verständnis des österreichischen Gesetzgebers – auch etwa die Landesverteidigung ist eigentlich vom Anwendungsbereich des Unionsrechts und in der Folge auch von jenem der DSGVO und der DSRL-PJ ausgenommen. Demgegenüber setzt das DSG in seinem 3. Hauptstück zwar prinzipiell Regelungen der DSRL-PJ um, erklärt jedoch, dass dieses Hauptstück auch bei der Verarbeitung personenbezogener Daten durch zuständige Behörden zum Zweck der **nationalen Sicherheit**, **Nachrichtendienste** und **militärischen Eigensicherung** anwendbar sein sollen. Weiters regelt § 4 Absatz 1 des DSG im Wesentlichen, dass dessen (allgemeine) Bestimmungen **sowie** jene der DSGVO dann auf entsprechende Datenverarbeitungen anwendbar sein sollen, wenn nicht das dritte Hauptstück des DSG speziellere Regelungen vorsieht. Das hat demnach wiederum auch eine Erstreckung des Anwendungsbereichs der DSGVO auf Bereiche zur Folge, in denen diese eigentlich nicht zur Anwendung kommen würde. Aus diesen Vorschriften ergibt sich demnach ein gewisses Geflecht an potenziell datenschutzrechtlich relevanten Regimen für die einschlägigen Bereiche der internationalen Hilfe.
Unter dem Strich verhält es sich dabei demnach wohl so, dass in gewissen Bereichen die DSGVO direkt anwendbar und demnach vor allem diese maßgeblich wäre (prinzipiell aber national auch das DSG), für EU-Organe und -

Institutionen sowie deren sonstige Stellen wiederum die Verordnung 2018/1725. Hingegen sind diese Verordnungen in Bereichen, die der nationalen Sicherheit zuzuordnen sind, prinzipiell nicht direkt anwendbar. Insbesondere dort käme allerdings wiederum das DSG zum Tragen, das im Allgemeinen seine eigenen Bestimmungen neben jenen der DSGVO für anwendbar erklärt, sofern nicht die Bestimmungen in dessen drittem Hauptstück (das insbesondere die DSRL-PJ umsetzt) als die spezielleren vorgehen. Fraglich ist vor allem auch der Bereich der Gemeinsamen Außen- und Sicherheitspolitik der EU, der im Gegensatz zur nationalen Sicherheit prinzipiell dem Anwendungsbereich des Unionsrechts unterliegen würde. Das DSG würde nach dessen Wortlaut nämlich auch in diesem Bereich die DSGVO – entgegen ihren Regelungen über den Anwendungsbereich – national für anwendbar erklären.

In der Folge bestehen auch gewisse potenzielle Rechtsgrundlagen für die Verarbeitung personenbezogener Daten durch das Bundesheer im Rahmen eines etwaigen Einsatzes des GeoCROW-Systems. Solche Rechtsgrundlagen sind aus datenschutzrechtlicher Sicht eine wesentliche Voraussetzung für die Verarbeitung personenbezogener Daten, hängen jedoch im Detail auch vom anwendbaren Datenschutzregime ab. Insbesondere ist hier zum einen § 10 DSG zu nennen, welcher die personenbezogene Datenverarbeitung durch Verantwortliche des öffentlichen Bereichs und Hilfsorganisation im Katastrophenfall regelt, und zum anderen § 6a Auslandseinsatzgesetz 2001. Letztere Bestimmung regelt prinzipiell Aufgaben und Befugnisse im Auslandseinsatz und bezieht sich dabei unter anderem auf Personen, die im Zuständigkeitsbereich des Bundesministers für Landesverteidigung zu einem Auslandseinsatz zur solidarischen Teilnahme an Maßnahmen der humanitären Hilfe und der Katastrophenhilfe oder Maßnahmen der Such- und Rettungsdienste entsendet werden (§ 1 Ziffer 1 Buchstaben b und c des BVG über Kooperation und Solidarität bei der Entsendung von Einheiten und Einzelpersonen in das Ausland; „KSE-BVG“). Eine der dabei möglichen Befugnisse betrifft auch die Verarbeitung für die Wahrnehmung entsprechender Aufgaben notwendiger personenbezogener Daten (§ 6a Absatz 2 Ziffer 1 Auslandseinsatzgesetz 2001). In diesem Zusammenhang haben die gemäß KSE-BVG für die Entsendung zum jeweiligen Auslandseinsatz zu-

ständigen Organe durch eine Verordnung prinzipiell festzulegen, welche Befugnisse mit welchen Mitteln wahrzunehmen wären (siehe im Detail § 6a Absatz 3 Auslandseinsatzgesetz 2001). Die entsprechenden Gesetzesmaterialien halten in Hinblick auf einschlägige Datenverarbeitungen fest, dass dabei durch eine entsprechende Verordnung zu präzisieren sein werden, etwa in Hinblick auf die Festlegung gewisser Datenkategorien, den Betroffenenkreis oder mögliche Empfänger im Wege einer Übermittlung.

Abgesehen davon ist sowohl in Bezug auf § 10 DSG als auch § 6a Auslandseinsatzgesetz 2001 bis zu einem gewissen Grad fraglich, inwieweit diese in für den tatsächlichen Einsatz des GeoCROW-Systems relevanten Situationen nutzbar gemacht werden könnten. In Hinblick auf § 6a Auslandseinsatzgesetz 2001 betrifft dies etwa die Aspekte, welche Personen zur Datenverarbeitung ermächtigt werden könnten bzw. zu welchem Zeitpunkt. In Hinblick auf § 10 DSG könnte demgegenüber unter anderem prinzipiell überlegt werden, wie die Bestimmung im Detail auf Sachverhalte mit Auslandsbezug anwendbar ist. Zudem könnten die Aufbewahrung und Verwendung von personenbezogenen Daten für künftige Übungs- oder Ausbildungsmaßnahmen wegen der strikten Speicherbegrenzung in Absatz 6 wohl nicht auf diese Bestimmung gestützt werden.
Schließlich ist Hinblick auf den speziellen Sachverhalt, der potenziell bei einem entsprechenden Einsatz des Systems vorliegen würde, auch fraglich, ob diese Bestimmungen unter dem Blickwinkel der hinreichenden Genauigkeit von Regelungen, die in das Datenschutzgrundrecht und andere Grundrechte eingreifen, ausreichen würden: Aus dem Legalitätsprinzip (Art 18 Abs 1 B-VG) ergibt sich nämlich die Verpflichtung, dass Verwaltungshandeln nicht nur inhaltlich, sondern auch in formeller Hinsicht ausreichend determiniert werden muss. Es gibt jedoch keine ausdrückliche Formvorschrift, wie diese hoheitliche Ermächtigung durch den Gesetzgeber ausgestaltet ist. Jedoch müssen die wesentlichen Voraussetzungen und Inhalte des behördlichen Handelns umschrieben sein, und zwar so, dass die Übereinstimmung des behördlichen Handelns überprüft werden kann.
Hinsichtlich der erfolgenden Verarbeitungstätigkeiten betreffend die Verarbeitung von personenbezogenen Daten ist zwischen Verarbeitungstätigkei-

ten im Rahmen des Forschungsprojekts und im geplanten späteren Echtbetrieb zu unterscheiden. Werden personenbezogene Daten für wissenschaftliche Forschungszwecke verarbeitet, gelten dafür grundsätzlich die gleichen Standards zum Schutz betroffener Personen wie bei sonstigen Datenverarbeitungszwecken. Auch im Forschungsbereich erfordert somit jede Verarbeitung von personenbezogenen Daten einen Erlaubnistatbestand. Die DSGVO hat den einzelnen Mitgliedstaaten die Möglichkeit eröffnet, durch Öffnungsklauseln entsprechende gesetzliche Verarbeitungsgrundlagen zu schaffen. Für die Schaffung von national-gesetzlichen Verarbeitungsgrundlagen, sind dabei insbesondere diejenigen in den Art 6 und 9 DSGVO relevant. Für den Bereich der wissenschaftlichen Forschung bestehen einerseits im Datenschutzgesetz (DSG) Sonderregelungen. Um Datenschutz und das Interesse an Wissenschaft und Forschung bestmöglich zu vereinbaren, erfolgte zudem die Novellierung des Forschungsorganisationsgesetzes (FOG) durch das WFDSAG 2018 welches ebenfalls Sonderregelungen betreffend Datenverarbeitungen für Forschungszwecke beinhaltet. Die Privilegierungen des FOG werden in der Literatur, vom Datenschutzrat und der DSB zwar kritisch gesehen. Bislang fehlt jedoch weitgehend Rechtsprechung der DSB zum FOG und kann dieses daher – bis zu einer anderslautenden Spruchpraxis – als Rechtsgrundlage für Verarbeitungstätigkeiten in GeoCROW - herangezogen werden.
Hinsichtlich eines allfälligen künftigen Echtbetriebs wurde festgestellt, dass auch bei „schlicht-hoheitlichem Verwaltungshandeln“ erhöhte Bestimmtheitserfordernisse im Sinne eines differenzierten Legalitätsprinzips bestehen, wenn es zu Grundrechtseingriffen – wie etwa der Verarbeitung von personenbezogenen Daten - kommt bzw. kommen kann. In diesem Bereich wäre es daher in eventu wohl empfehlenswert, mit einer exakt passenden gesetzlichen Bestimmung nachzuschärfen.

Neben der in Auszügen soeben beschriebenen, grundlegenden rechtlichen Analyse und einer allgemeinen datenschutzrechtlichen Begleitung der laufenden Fortschritte im Projekt (etwa im Zusammenhang mit der weitestgehenden Anonymisierung potenzieller personenbezogener Daten oder des Hostings) wurde das Projekt im Rahmen der rechtlichen Evaluierung des GeoCROW-Systems bereits von einem frühen Stadium an auch durch eine

Datenschutz-Folgenabschätzung begleitet. Hierbei wurde der Fokus allerdings nicht auf einen potenziellen Echtbetrieb des Systems gelegt, sondern insbesondere der letztaktuelle Stand des Forschungsprojekts und relevanter Vorgänge beurteilt, um vor allem die (bis zu einem gewissen Grad) bereits feststehenden Aspekte des Sachverhalts analysieren zu können, weil veränderliche, zukünftige oder unsichere Aspekte dabei nur schwer mit der notwendigen Klarheit beurteilt werden können. In Hinblick auf eine tatsächliche Inbetriebnahme wäre demnach gegebenenfalls eine spezifischere Datenschutz-Folgenabschätzung durchzuführen.

Im Allgemeinen wird im Zuge dessen vereinfacht formuliert der datenschutzrechtlichen Rahmen des Untersuchungsgegenstands im Detail beurteilt, entsprechende Grenzen abgesteckt bzw. auch Verbesserungsmöglichkeiten und -erfordernisse aufgezeigt, insbesondere nach den Vorgaben von Artikel 35 DSGVO. Dabei steht die Analyse der Notwendigkeit und Verhältnismäßigkeit der jeweiligen Verarbeitungen von personenbezogenen Daten im Fokus und dabei auch etwa die Beurteilung von Möglichkeiten zur Anonymisierung von Daten. Der gesamte Prozess wird hierbei in einem entsprechenden Bericht beschrieben, bereits beginnend mit Überlegungen dazu, inwiefern überhaupt eine DSFA durchzuführen ist bzw. durchgeführt wird (Schwellenwertanalyse). Im Zentrum der Datenschutz-Folgenabschätzung steht die Beurteilung von einschlägigen Risiken und die Eruierung entsprechender Mitigierungsmaßnahmen. In Hinblick auf das GeoCROW-Projekt ist bezüglich der entsprechenden Dokumentation insbesondere auf den finalen Evaluationsbericht (Deliverable 5.3) zu verweisen.

4. Ethische Rahmenbedingungen

Aus ethischer Perspektive wirft das Projektvorhaben mehrere Fragestellungen auf, die aus Sicht unterschiedlicher moralphilosophischer Paradigmata im Rahmen verschiedener Abstraktionsebenen untersucht werden können. Um eine möglichst umfassende ethische Evaluierung durchführen zu können, erweist sich daher als sinnvoll, die Analyse einerseits nach den drei herrschenden moralphilosophischen Paradigmata, andererseits nach fünf einschlägigen Abstraktionsebenen zu unterteilen.

Die drei herrschenden moralphilosophische Paradigmata sind (1) die Tugendethik, (2) der Konsequentialismus, (3) die deontologische Ethik.[13]
Tugend- oder aretaische Ethiken basieren auf der Vorstellung, dass das ethische Handeln durch die Entwicklung und das Ausleben bestimmter Charaktereigenschaften, d.h. der Tugenden gekennzeichnet wird.[14] Meistens wird hierzu eine Form von moralischer Weisheit oder *Phronesis* (*φρονήσει*) postuliert, die die praktische Ausübung tugendhaften Handelns leiten soll.[15] Tugendhaftes Handeln sei dabei der Weg zur Erreichung eines glücklichen bzw. erfüllten Lebens, der sog. *Eudaimonia* (*εὐδαιμονίας*). Damit eng verwandt sind Werteethiken, die diese Vorstellung eines guten bzw. erfüllten bzw. lebenswerten Lebens (der *Eudaimonia*) wiederum mit der Verwirklichung bestimmter Werte in Verbindung setzen.[16] Tugend- und werteethische Aspekte spielen bei modernen Ansätzen zur Technologie- und Computer- bzw. zur digitalen Ethik eine zentrale Rolle.[17] Im Kontext von GeoCROW dürften sich tugendethische Aspekte u.a. bei der Betrachtung von Herangehensweisen bzw. sog. *best practices* insbesondere bei der Entwicklungsphase von neuen Produkten und Technologien zeigen. Darin sind auch Vorstellungen wie *Ethics by Design* einzuordnen. Verwandte werteethische Erwägungen treten wiederum in den Vordergrund, wenn es um die Bestimmung bzw. um die Verfolgung erstrebenswerter Ziele oder Werte geht, die durch neue Technologien oder Produkte verfolgt, verwirklicht oder zumindest berücksichtigt werden müssen. Beispiele dafür im Kontext von GeoCROW sind Datenschutz und *Privacy*, Sicherheit, Effizienz, Umweltschutz, Demokratie, Rechtstaatlichkeit usw.
Konsequentialistische Ethiken legen den Fokus auf die Folgen bzw. die

---

[13] Rosalind Hursthouse and Glen Pettigrove, "Virtue Ethics," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta and Uri Nodelman, Fall 2023 (Metaphysics Research Lab, Stanford University, 2023), https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/.

[15] Hursthouse and Pettigrove, "Virtue Ethics."

[16] M. Scheler and C. Bermes, *Der Formalismus in Der Ethik Und Die Materiale Wertethik: Neuer Versuch Der Grundlegung Eines Ethischen Personalismus*, Philosophische Bibliothek (Felix Meiner Verlag, 2014), https://books.google.at/books?id=OwuKDwAAQBAJ.

[17] K. Davis, *Ethics of Big Data: Balancing Risk and Innovation* (O'Reilly Media, 2012), https://books.google.at/books?id=SG5F3vxBRu4C.

Konsequenzen der jeweils zu evaluierenden Handlungen und Verhaltensweisen. Die meisten konsequentialistischen Ethiken basieren auf dem Prinzip des Nutzens (*Principle of Utility*), wonach der moralische Wert einer Handlung ihren Folgen zur Vermehrung (oder Verminderung) des Glücks (genauer des Nutzens) einer Gruppe entspricht.[18] Das Prinzip des Nutzens kann mit dem Lemma „Das größte Glück der größten Zahl" (*„The greatest good for the greatest number"*)[19] zusammengefasst werden, wodurch die Maximierung des Gesamtnutzens als endgültiges ethisches Ziel postuliert wird. Im Zusammenhang mit GeoCROW werden utilitaristische Aspekte insbesondere bei der Betrachtung von effektiv mess- und vergleichbaren quantitativen Folgen mit moralisch relevantem Inhalt eine wichtige Rolle spielen.
Während konsequentialistische Theorien das Gute (d.h. den Nutzen) als primitivbegriff annehmen und daraus die jeweilige moralische Pflicht – d.h. die zur Maximierung des Nutzens führende Handlung – herleiten, gehen deontologische Ethiken genau umgekehrt vor und leiten dementsprechend das Gute aus einer für primitiv angenommene Pflicht her, etwa nach dem Lemma: *Gut ist das, was die Pflicht erfüllt*.[20] Im Kontext von GeoCROW dürfen Aspekte der deontologischen Ethik insbesondere dann in den Vordergrund treten, wenn es um die Abwägung von Grund- und Menschenrechten moralischer Natur (bspw. im Zusammenhang mit Persönlichkeitsrechten und mit Privatsphäre) geht. Da diese Normen auch ins geltende Rechtssystem übernommen wurden, werden sie i.d.R. eher im Rahmen der Analyse bzw. der Evaluierung der rechtlichen Anforderungen an GeoCROW zu betrachten sein.

---

[18] J. Bentham, *An Introduction to the Principles of Morals and Legislation* (Dover Publications, 2012), https://books.google.at/books?id=GcnssaU-1wgC.

[19] Julia Driver, "The History of Utilitarianism," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta and Uri Nodelman, Winter 2022 (Metaphysics Research Lab, Stanford University, 2022), https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history/.

[20] Larry Alexander and Michael Moore, "Deontological Ethics," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2021 (Metaphysics Research Lab, Stanford University, 2021), https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/.

Die ethische Evaluierung wird zudem in Anlehnung an *Luciano Floridi*[21] auf fünf Abstraktionsebenen entwickelt:

1. **Soziopolitische Ebene**: Die soziopolitische Abstraktionsebene besteht aus Beobachtbaren, die allgemeine soziokulturelle und politische Aspekte betreffen, etwa hinsichtlich der Staats- und Regierungsform (bspw. Rechtsstaatlichkeit, Autokratie/Demokratie, Zivilgesellschaft/Militär) und der allgemeinen Gesellschaftsstruktur (Bräuche und Sitten, Religion, Sprache, Geschichte und Erinnerungskultur).
2. **Menschenebene**: Die Beobachtbaren in der Menschenebene betreffen primär die Menschen, die bei der Entwicklung und Anwendung der jeweiligen Technologien und Produkte aktiv oder passiv beteiligt sind.
3. **Computerebene**: Die Beobachtbaren dieser Ebene betreffen primär die Eigenschaften der jeweils entwickelten Technologien und Produkte, insbesondere in Anbetracht ihrer jeweiligen gesellschaftlichen Funktion.
4. **Informationsebene**: Die Informationsebene besteht aus Beobachtbaren betreffend die verschiedenen Aspekte der Information, die im Kontext der Entwicklung neuer Technologien und Produkte erhoben, gespeichert, geändert, vermittelt und gelöscht wird.
5. **Datenebene**: Strenggenommen verarbeiten Rechner keine Informationen, sondern Daten. Information kann als die semantische Deutung von Daten definiert werden; Daten wiederum als rein syntaktische Reihen von Zeichen. Durch Betrachtungen im Rahmen der Datenebene wird der Umstand hervorgehoben, dass Daten bzw. ihre Verarbeitung insbesondere in großen Mengen auch unabhängig von

[21] Luciano Floridi, "The Method of Levels of Abstraction," *Minds and Machines* 18, no. 3 (September 1, 2008): 303–29, https://doi.org/10.1007/s11023-008-9113-7. Floridi selbst behauptet die grundsätzliche Anwendbarkeit der Methode der Abstraktionsebenen auf ethische Fragestellungen etwa in Luciano Floridi and Mariarosaria Taddeo, "What Is Data Ethics?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, no. 2083 (December 28, 2016): 20160360, https://doi.org/10.1098/rsta.2016.0360.

ihrem unmittelbaren Informationsgehalt weitrechende Konsequenzen haben können.

Darüber hinaus wird das Projektvorhaben auch hinsichtlich der Beachtung von Grund- und Menschenrechten evaluiert. In Art 8 der Konvention zum Schutze der Menschenrechte und Grundfreiheiten (EMRK),[22] die in Österreich im Verfassungsrang steht, wird die Achtung des Privat- und Familienlebens festgelegt. Demnach hat jedermann Anspruch auf Achtung seines Privat- und Familienlebens, seiner Wohnung und seines Briefverkehrs. Der Eingriff einer öffentlichen Behörde in die Ausübung dieses Rechts ist nur statthaft, insoweit dieser Eingriff gesetzlich vorgesehen ist und eine Maßnahme darstellt, die in einer demokratischen Gesellschaft für die nationale Sicherheit, die öffentliche Ruhe und Ordnung, das wirtschaftliche Wohl des Landes, die Verteidigung der Ordnung und zur Verhinderung von strafbaren Handlungen, zum Schutz der Gesundheit und der Moral oder zum Schutz der Rechte und Freiheiten anderer notwendig ist. Die verschiedenen Formen der staatlichen Datensammlungen sind daher in Hinblick auf Art 8 EMRK rechtfertigungsbedürftig. Die Beobachtung gesellschaftlicher Entwicklungen und die Ermittlung spezifischer Gefahren, erfordert oft Grundrechtseingriffe, die sich bereits im Rahmen des Social Media-Monitorings durch die Verarbeitung von öffentlich zugänglichen personenbezogenen Daten manifestieren.[23] Selbst wenn von Unternehmen in der Social Media-Branche automatisierte Strategien verwendet werden, um Persönlichkeitsprofile zu erstellen und zukünftiges Verhalten vorherzusagen, stellt dies einen Eingriff in die Privatsphäre dar, der hier nicht nur durch staatliche Regulierung begrenzt wird, sondern auch durch die Anwendung von Computerprogrammen. Ein solcher Eingriff in die Grundrechte ist nur in Fällen der nationalen Sicherheit, öffentlichen Ruhe und Freiheit gerechtfertigt. Die Unternehmen beschränken das private digitale Leben ihrer Nutzer:innen, indem sie vermeintlich ideale Bedingungen schaffen, obwohl dies zu einem Missbrauch der Meinungsbildung führen kann.[24] Das von GeoCROW verfolgte Vorhaben soll

[22] "Konvention Zum Schutze Der Menschenrechte Und Grundfreiheiten," Pub. L. No. BGBl 1958/210 (n.d.).

[23] Scheichenbauer and Seidl, "Die Verwaltungsrechtliche Einordnung von Internet-Recherchen Durch Verwaltungs-Behörden," in *Jahrbuch Datenschutzrecht 2022*, 56, n.d.

[24] Victoria Bacher, "Regulierungsbedarf Beim Einsatz von Algorithmen in Sozialen Medien Zum Schutz Der Meinungsfreiheit," *ELSA Austria Law Review* 7, no. 1 (2022): 64–69, https://doi.org/10.33196/ealr202201006401.

mithin stets auf die Verhältnismäßigkeit zwischen den in seinem Rahmen zu entwickelnden neuen Technologien und Methoden einerseits und der Wahrung bzw. dem Schutz des Privat- und Familienlebens anderseits achten.
In Art 10 der EMRK wird die Freiheit der Meinungsäußerung festgelegt. Demnach hat jedermann Anspruch auf freie Meinungsäußerung. Dieses Recht schließt die Freiheit der Meinung und die Freiheit zum Empfang und zur Mitteilung von Nachrichten oder Ideen ohne Eingriffe öffentlicher Behörden und ohne Rücksicht auf Landesgrenzen ein. Da die Ausübung dieser Freiheiten Pflichten und Verantwortung mit sich bringt, kann sie bestimmten, vom Gesetz vorgesehenen Formvorschriften, Bedingungen, Einschränkungen oder Strafdrohungen unterworfen werden, wie sie in einer demokratischen Gesellschaft im Interesse der nationalen Sicherheit, der territorialen Unversehrtheit oder der öffentlichen Sicherheit, der Aufrechterhaltung der Ordnung und der Verbrechensverhütung, des Schutzes der Gesundheit und der Moral, des Schutzes des guten Rufes oder der Rechte anderer unentbehrlich sind, um die Verbreitung von vertraulichen Nachrichten zu verhindern oder das Ansehen und die Unparteilichkeit der Rechtsprechung zu gewährleisten.[25]
Der sachliche Anwendungsbereich dieses Grundrechts umfasst die freie Meinungsäußerung, somit eine Verhaltensweise, die schutzbedürftig ist und deren Eingriff einer Rechtfertigung bedarf.[26] Geschützt wird mit Art 10 EMRK die Meinungsfreiheit ohne Einschränkung auf bestimmte Kommunikationsmittel.[27] Damit sind unter anderem Twitter-Meldungen durch Art 10 EMRK prima facie geschützt. Darauf soll daher im Rahmen von GeoCROW besonders geachtet werden, da das Projektvorhaben die Auswertung von Social-Media-Daten, etwa Twitter-Nachrichten vorsieht.
Gerade im Zusammenhang mit Twitter ist auf die Problematik von Falschmeldungen und Fake News einzugehen. Der Schutzbereich von Art 10 EMRK greift nicht nur unabhängig vom gewählten Kommunikationsmittel, sondern auch unabhängig von dem Inhalt der Kommunikation.[28] Geschützt

[25] Spindler and Schuster, *Recht Der Elektronischen Medien*, 34–65, n.d.
[26] *Pöschl* in *Koziol* 40.
[27] *Pöschl* in *Koziol* 40.
[28] *Pöschl* in *Koziol* 41.

sind nach der Judikatur daher auch Falschmeldungen und Werturteile.[29] Lässt sich diese Judikaturlinie auch uneingeschränkt auf das Internet und Online-Meldungen übertragen? Nach *Pöschl* ist dies zu bejahen und es sind sowohl falsche Tatsachenbehauptungen als auch Werturteile prima facie vom Schutzbereich umfasst und erst auf der Ebene der Schranke zu prüfen, ob und inwieweit verhältnismäßig und angemessen eingegriffen werden darf.[30]

Ursprünglich als Schutzrechte gegenüber dem Staat konzipiert, werden in den Anwendungsbereich der Grundrechte zunehmend auch Dritte gezogen. Und dies aus gutem Grund – geht doch gerade bei der Meinungsfreiheit die Gefahr selten vom Staat direkt aus, sondern von Unternehmen wie Meta Platforms[31], Inc., Alphabet Inc. und Co. Dabei ist nicht zu übersehen, dass die von den Unternehmen gesammelten Daten bisweilen auch an staatliche Behörden weitergegeben werden.[32] Die damit verbunden Intransparenz für die Nutzer*innen kann zu Eingriffen in die Rede- und Meinungsfreiheit führen.[33] Generell wird also im Rahmen von GeoCROW auch auf die Berücksichtigung bzw. Erfüllung von Menschen- und Grundrechte zu achten sein.

Die hier skizzierte, allgemeine Analyse der ethischen Anforderungen an GeoCROW zeigt, dass das Projekt vielfältige, teilweise sehr komplexe ethische Fragestellungen aufwirft. Während der Durchführung des Vorhabens sind diese Fragestellungen und Anforderungen des Näheren zu konkretisieren und ggf. je nach Bedarf zu ergänzen. Hierfür erweist sich als ratsam, die oben festgesetzten Maßstäbe, d.h. die Grundsätze und Grundkriterien der drei etablierten Ethiktheorien (deontologischer Ethik, Tugendethik, Konsequentialismus) im Auge zu behalten. Ebenfalls ersprießlich erscheint die Raffinierung der jeweiligen Analysen durch die Methode der Abstraktionsebenen. Fragen betreffend die Erfüllung von Menschen- und Grundrechten sind dualer (d.h. ethischer und rechtlicher) Natur und sollen daher in engem Austausch mit der juristischen Evaluation des Projektvorhabens ausgearbeitet werden. Da die ethische Evaluierung größtenteils *in abstracto* erfolgt, hängt

---

[29] Vgl EGMR 26. 2. 2002, *Dichand ua/Österreich*, EGMR 2. 11. 2006, *Standard Verlags GmbH und Krawagna-Pfeifer/Österreich*, EGMR 7. 12. 1976, *Handyside/Vereinigtes Königreich*, EGMR 10. 7. 2003, *Murphy/Irland*.

[30] *Pöschl* in *Koziol* 43.

[31] *Brosch*, Alles neu macht das ... Facebook Oversight Board?, MMR 2021, 26.

[32] Magdalena Pöschl, "Neuvermessung der meinungsfreiheit," in *Tatsachenmitteilungen und werturteile: Freiheit und verantwortung*, ed. Helmut Koziol (Jan Sramek Verlag, 2018), 31–59.

[33] Pöschl.

sie nicht vom konkreten Entwicklungsstand der technischen Lösungen des Projekts ab.

5. Technische Rahmenbedingungen

Tabelle 1 zeigt die Hardware-Spezifikationen des vom Institut für Militärisches Geowese vorgegebenen Rechners. Den Prozessor (CPU = Central Processing Unit) bildet ein Intel Xeon Chip. Des Weiteren besitzt der Rechner eine Radeon Graphikkarte (GPU = Graphics Processing Unit). Hinsichtlich des Arbeitsspeichers (SDRAM = Synchronous Dynamic Random Access Memory) ist das System mit 128 GB ausgestattet. Zur Speicherung der Daten steht eine 500 GB SATA Hard Drive sowie zwei 1 TB SATA-Festplattenlaufwerke zur Verfügung. Als Betriebssystem läuft Windows 11 auf dem Server.
Unabhängig davon ist das Gesamtsystem so aufgebaut, dass es mit den zur Verfügung stehenden Hardware-Ressourcen skaliert. Das heißt stehen einem weniger Ressourcen zur Verfügung leidet die Geschwindigkeit, stehen einem mehr Ressourcen zur Verfügung steigt die Geschwindigkeit.

Tabelle 1: Hardware Spezifikationen des vom Institut für Militärisches Geowesen als Zielmaschine vorgegebenen Rechner DELL 7920 Rack.

CPU	Intel Xeon Gold 6140M 2.3GHz (3.7GHz Turbo 18C, 10.4GT/s 3UPI, 25MB Cache, HT (140W) DDR4-2666)
SDRAM	128GB (2x64GB) DDR4 2666MHz RDIMM ECC
GPU	Radeon Pro WX 7100, 8GB, 4DP, (7920 Rack)
Netzwerk	Intel I350 (4x1Gbit) vier Ports Netzwerk Tochterkarte
HDD	1x 2.5" 500GB 7200rpm SATA Hard Drive 2x 2,5-Zoll-SATA-Festplattenlaufwerk, 1 TB, 7.200 1/min
OS	Win11

Teil B: Die GeoCROW-Systemarchitektur

1. Architektur

Die in Abbildung 2 dargestellte Systemarchitektur resultiert aus einem iterativen Entwicklungsprozess während der gesamten Dauer des GeoCROW-Projekts. Die hier dargestellte finale Version des Systems besteht nur mehr aus den Kern-Komponenten die essentiell sind, um die Aufgaben des Systems zu erfüllen. Das GeoCROW-System basiert auf einer Microservices Architektur, das heißt, jeder Service ist für sich abgekapselt und erfüllt eine spezifische Funktion des Gesamtsystems. Im Falle von GeoCROW wird Docker genutzt, um die Microservices in Container laufen zu lassen und durch ein internes Docker-Netzwerk zu verbinden. Innerhalb dieses Netzwerks sind die Microservices über gewisse Ports erreichbar und ihre Funktionen können über ihre API-Endpoints ausgeführt werden. Diese Struktur sorgt für zusätzliche Sicherheit, da keine Komponente des Systems für externe zugänglich ist.
In Kombination mit der zentralen Datenbank GraphDB[34] und der Orchestrierungssoftware Airflow[35] sorgt diese Architektur für eine erhöhte Ausfallsicherheit und Fehlertoleranz des Systems. Die GraphDB dient dabei zur temporären Speicherung aller gewonnenen Informationen. Airflow hingegen definiert den Workflow und managet diesen, damit die einzelnen Services zum richtigen Zeitpunkt aufgerufen werden und mit den benötigten Daten versorgt werden.
Zur Gewährleistung der Modularität und der Minimierung von Abhängigkeiten erfolgt die Kommunikation mit der Datenbank ausschließlich über den CRUD-Service. Dieser besitzt API-Endpoints zur Speicherung (Create), Abfrage (Retrieve), Aktualisierung (Update) und Löschung (Delete) der verschiedenen im System verarbeiteten Informationen. Der Aufbau der Infrastruktur durch Microservices erleichtert außerdem die Wartung und Entwicklung des Systems, sowie die Handhabung komplexer Use Cases. Das Verpacken der Microservices als Docker-Container ermöglicht außerdem eine Betriebssystemunabhängigkeit.
Für die Interfaces der verschiedenen Services werden OpenAPI[36] und

[34] Ontotext, "GraphDB," 2023, https://graphdb.ontotext.com/.
[35] The Apache Software Foundation, "Apache Airflow," 2023, https://airflow.apache.org/.
[36] The Linux Foundation, "OpenAPI," 2022, https://www.openapis.org/.

FastAPI[37] verwendet. Die Dokumentation der Interfaces erfolgt mit Swagger[38]. Eine API (Application Programming Interface) ist eine Programmierschnittstelle, die es unabhängigen Anwendungen ermöglicht miteinander zu kommunizieren und Daten auszutauschen.[39] OpenAPI ist eine standardisierte Sprache für HTTP APIs. Diese gibt Struktur und Syntax der API unabhängig von der genutzten Programmiersprache vor. FastAPI ist ein modernes high-performance Web Framework, das auf den Spezifikationen von OpenAPI aufbaut, um APIs mit Python zu erstellen. Swagger nutzt ebenfalls OpenAPI und bietet eine Infrastruktur, um APIs zu designen, zu entwickeln, zu dokumentieren und zu testen.
Um funktional zu sein, benötigt das System verschiedene Inputs. Zu diesen Inputs zählen Informationen, die der Nutzer stellen muss, Websources, eine Incident Taxonomy, GeoNames[40] und die OpenStreetMap[41]. Zu den Nutzerinformationen zählen beispielsweise die geographische Ausdehnung des Untersuchungsgebiets sowie der Zeitraum, der von Interesse ist. Außerdem benötigt das System eine Liste vordefinierter Websources, die zur Gewinnung von Daten gescraped werden. Die Ereignistyptaxonomie stellt eine Sammlung von Begriffen dar, die die Ereignisse festlegt, die von Interesse sind. GeoNames ist eine Gazetteer (Ortsverzeichnis), das heißt eine Ansammlung von Ortsnamen mit zugehörigen Koordinaten. Dieses wird benötigt, um beim Geoparsing Ortsnamen in den unstrukturierten Texten zu finden und sie korrekt zu georeferenzieren. Die OpenStreetMap dient dazu, die Umgebung im Einsatzgebiet in der Virtual Reality abzubilden. Der Ablauf sieht vor, dass Airflow die Services entsprechend eines vordefinierten Workflows aufruft, diese die benötigten Daten über den CRUD-Service aus der Datenbank beziehen, die Daten verarbeiten und ihre Ergebnisse über den CRUD-Service wieder in die Datenbank schreiben.
Abschließend werden die relevanten Ergebnisse als GeoJSON exportiert und in die VR-Umgebung eingebunden. Der Teil der Virtual Reality ist in gewisser Weise vom restlichen System abgekoppelt, da dieser vollkommen

[37] FastAPI, "FastAPI," 2024, https://fastapi.tiangolo.com/.
[38] SmartBear Software, "Swagger," 2023, https://swagger.io/.
[39] Red Hat, "Was Ist Eine API?," 2023, https://www.redhat.com/de/topics/api/what-are-application-programming-interfaces.
[40] GeoNames, "GeoNames," 2023, https://www.geonames.org/.
[41] OpenStreetMap, "About OpenStreetMap," 2024, https://www.openstreetmap.org/about.

offline funktionieren muss. Somit werden die exportierten Ergebnisse gemeinsam mit der OpenStreetMap in die VR-Umgebung integriert. In dieser Umgebung kann der Nutzer schließlich interaktiv und immersiv die Informationen durchsuchen.

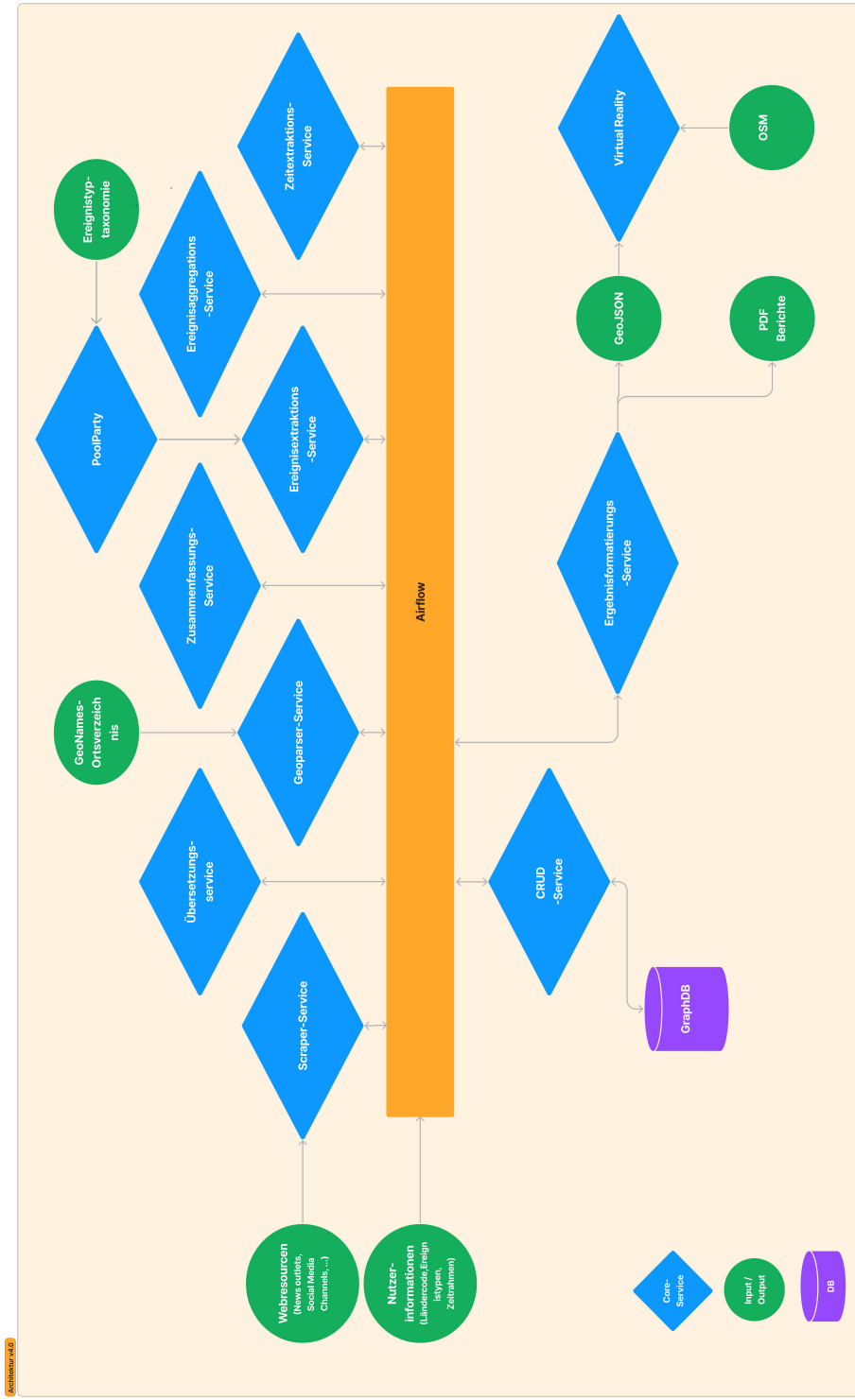

Abbildung 2: GeoCROW-Systemarchitektur mit allen Services, Inputs und Outputs, sowie der Orchestrierungskomponente Airflow und der zentralen Datenbank GraphDB.

2. Komponenten

Im Folgenden werden die Komponenten der Architektur einschließlich ihres Zweckes und Aufbaus kurz beschrieben. Jene Komponenten, die den Kern der Ereignisextraktion bilden, werden weiters ausführlicher in Abschnitt C hinsichtlich ihrer Funktionsweise beschrieben.

2.1 Workflow Engine

Der Ablauf der gesamten Datenpipeline und somit des Gesamtsystems wird von der zentralen Orchestrierungssoftware Apache Airflow[42] gemanagt. Dabei handelt es sich um eine Plattform, die es ermöglicht Workflows zu kreieren, planen und überwachen Bei Airflow stehen die vier Prinzipien Skalierbarkeit, Dynamik, Erweiterbarkeit und Eleganz im Vordergrund. Aufgrund seiner Architektur ist es leicht skalierbar. Des Weiteren können Pipelines dynamisch generiert werden. Es können leicht eigene Operatoren definiert werden und so die Funktionalitäten von Airflow erweitert werden. Außerdem ist es elegant, da die Definition der Pipelines mit geringem Aufwand verbunden ist. Apache Airflow wird folglich im Gesamtsystem genutzt, um einen Workflow zu programmieren, der die einzelnen Microservices aufruft und neue Prozess in Gang bringt, nachdem die vorherigen abgeschlossen wurden.
In Airflow werden, die einzelnen Verarbeitungsschritte des GeoCROW-Systems als Tasks definiert. Jeder Task stellt somit einen API-Aufruf zu einem der anderen Services dar. Airflow führt die einzelnen Tasks in einer vordefinierten Reihenfolge aus und sorgt so für eine orchestrierte Ausführung des Gesamtsystems. Außerdem ermöglicht es Airflow dem Nutzer gewisse Input-Parameter festzulegen, die dann auch an die einzelnen Tasks weitergegeben werden können. Diese Input-Parameter sind beispielsweise der Zeitraum, sowie die Region, die für den Nutzer von Interesse sind.

2.2 Datenbank und Datenbank-Kommunikationsservice

[42] The Apache Software Foundation, "Apache Airflow."

Der CRUD-Service, sowie die GraphDB sorgen zusammen mit der Workflow Engine Airflow für eine erhöhte Ausfallsicherheit und Fehlertoleranz des Systems. Allerdings erfüllt vor allem die GraphDB eine Vielzahl an Aufgaben, so ermöglicht sie neben der Speicherung der Daten auch das Erforschen von Beziehungen zwischen den Daten aufgrund der Graphenstruktur. Dabei stellt die GraphDB einen RDF-Triple-Store dar, d.h. Daten werden in der Form von Triples gespeichert. Ein Triple besteht immer aus einem Subjekt (S), Prädikat (P) und Objekt (O). Zur Abfrage der Daten unterstützt GraphDB die Abfragesprache SPARQL und ihrer geographische Erweiterung GeoSPARQL.[43]
Die Datentriples, die Entitäten und ihre Verbindungen beschreiben werden durch den CRUD-Service zur Speicherung an die Datenbank übermittelt und bei Bedarf auch wieder abgefragt.
Somit fungiert der CRUD-Service als Vermittler zwischen den verschiedenen Datenverarbeitungsservices und der GraphDB. Dadurch sorgt der CRUD-Service für einen effizienten und nahtlosen Datentransfer. Durch diesen Service wird die komplexe Kommunikation mit der Datenbankinfrastruktur abstrahiert, was nicht nur die Implementierung der verarbeitenden Services vereinfacht, sondern auch die Wartung der gesamten Systemarchitektur aufgrund der reduzierten Abhängigkeiten einfacher gestaltet.
Dabei bietet der CRUD-Service Funktionen zur Speicherung (Create), Abfrage (Retrieve), Aktualisierung (Update) und Löschung (Delete) verschiedener Datenkonzepte an. Beispiele für solche Konzepte sind Newsitems oder Events. Die Definitionen der einzelnen Konzepte werden dabei neu durch die GeoCROW-Ontologie erstellt oder aus bestehenden Ontologien (z.B.: NIF) entnommen.

2.3 Komponenten zur Datengewinnung und Vorbereitung

Einen zentralen Bestandteil der Datenprozessierungspipeline, von dem alle anderen GeoCROW-Services abhängig sind, bilden die Komponenten zur Datengewinnung und Vorbereitung. Dabei handelt es sich vor allem um den Scraper-Service und den Übersetzungs-Service. Die Aufgabe des Scraper-

[43] Ontotext, "About GraphDB," 2023, https://graphdb.ontotext.com/documentation/10.0/about-graphdb.html.

Service liegt in der Sammlung von relevanten Daten von unterschiedlichen News Outlets, wobei sowohl klassische (digitale) Median, als auch Social-Media-Kanäle berücksichtigt werden.
Der GeoCROW-Scraper nutzt das Python-Paket Scrapy[44], ein open-source scraping Paket, dass es ermöglicht auf effiziente Art und Weise Daten von Websiten zu extrahieren. Scrapy nutzt das Konzept von sogenannten „Spiders", welche es ermöglichen unkompliziert und flexibel Parser für individuelle Website-Strukturen zu erstellen und so die gewünschten Inhalte von einer URL zu extrahieren. Jede Spider gibt das Ergebnis des Parsing-Prozesses in einer vordefinierten Struktur zurück.
Als Input benötigt der Scraper die Spezifikation eines Interessensgebiets mithilfe von Ländercodes, sowie den Zeitraum, der von Interesse ist. Als Ländercodes mit definierten Quellen und Spiders stehen Im Projekt TR (Türkei) und NR (Niger) zur Verfügung. Der Interessenszeitraum wird in Tagen angegeben und vom Tag des "Scrapens" zurückgerechnet. Die gesammelten Daten werden hinsichtlich der Quelle, des Erscheinungsdatums, des Titels sowie des Inhalts strukturiert.
Da der Scraper-Service die verschiedenen Quellen in ihrer Originalsprache extrahiert (z.B.: türkisch, französisch, etc.) ist es notwendig diese in eine einheitliche Sprache zu übersetzen, die sowohl der Nutzer als auch alle anderen GeoCROW-Services verstehen können. Diese Aufgabe übernimmt im GeoCROW-System der Übersetzungs-Service, der die ins System kommenden Informationen (Newsitems) in die englische Sprache übersetzt. Die Übersetzung ersetzt die originale Sprache nicht, sondern ergänzt diese. Das Ziel dieser Übersetzung ist die Reduktion der Sprachbarriere bei der Auswertung und Interaktion mit fremdsprachlichen Informationen. Basierend auf der Übersetzung erfolgen alle weiteren, sprachbezogenen Operationen, wie etwa das Geoparsing.
Im ersten Schritt erkennt der Service um welche Sprache es sich beim Originaltext handelt. Hier kommt die Python-Bibliothek Lingua[45] zum Einsatz, welche mittels Sprachmodellen die Originalsprache feststellt. Dies ist notwendig, da es – abhängig von der Originalsprache und dem vorliegenden Material und dessen Umfang – zu Mehrdeutigkeiten kommen kann. Für die eigentliche Übersetzung wird das Transformer-based Pretrained Language

[44] Scrapy, "Scrapy," 2024, https://scrapy.org/.
[45] Lingua, "Lingua-Language-Detector 2.0.2," 2023, https://pypi.org/project/lingua-language-detector/.

Model (T-PTLM) MBart-50 genutzt. Dieses Sprachmodell beherrscht über 50 verschiedene Sprachen und ist somit sehr flexibel hinsichtlich der zu übersetzenden, textuellen Informationen.

2.4 Services zur Event Extraktion

Die gescrapten und übersetzten Datenobjekte werden danach von verschiedenen GeoCROW-Services weiterverarbeitet um wichtige Informationen zu gewinnnen. Dazu zählen die Extraktion von Orten, Zeiten, Ereignistypen und deren Aggregation zu Ereignissen.
Die Extraktion der Orte übernimmt dabei der Geoparser-Service, der zunächst die Ortsbeschreibungen in den Newsitems markiert und anschließend mit der korrekten Entität in dem Ortsverzeichnis GeoNames in Verbindung bringt. Ein Ortsverzeichnis ist eine Ansammlung von Ortsnamen in einer oder mehreren Sprachen. Im Falle von GeoCROW wird der Geoparser Mordecai3, der mit dem Ortsverzeichnis GeoNames arbeitet, genutzt. GeoNames ist eine geographische Datenbank, die unter einer Creative Commons Attribution License frei nutzbar ist. Das Ortsverzeichnis enthält über 25 Millionen Ortsnamen. Die alle mit einer geographischen Koordinate versehen sind. Die Geoparser-Software Mordecai3 nutzt in einem ersten Schritt (1) Named-Entity-Recognition Modelle von spaCy um Ortsnamen in Texten zu finden (Geotaggen) und disambiguiert diese in einem zweiten Schritt, (2) indem der Ortsnamen im GeoNames Ortsverzeichnis gesucht wird und die potentiellen Ergebnisse durch ein neuronales Netzwerk gereiht werden (Geocoding/Place Name Disambiguation). Das Ergebnis mit der besten Reihung wird für die Georeferenzierung des getagten Ortsnamens genutzt.
Die Extraktion von Zeiten erfolgt durch den Zeitextraktions-Service, der sowohl absolute als auch relative Zeitangaben extrahiert. Dieser Service basiert auf HeidelTime[46], einem mehrsprachigen, Zeittagger, der Zeitausdrücke aus Dokumenten extrahiert und diese normalisiert. Dabei nimmt er den übersetzten Newsitemtext als Input, sowie ein Referenzdatum, und extrahiert alle

[46] Jannik Strötgen and Michael Gertz, "A Baseline Temporal Tagger for All Languages," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal: Association for Computational Linguistics, 2015), 541–47, https://doi.org/10.18653/v1/D15-1063.

Zeitangaben aus dem Text. Relative Zeitangaben, wie zum Beispiel „Gestern", werden mittels des Referenzdatums disambiguiert. Den Output des Services bilden alle gefundenen und disambiguierten Zeitannotationen.
Der Ereignistypextraktions-Service nutzt sowohl den PoolParty-Service der Semantic Web Company als auch die im Laufe des Projekts erstellte Ereignistaxonomie, um Ereignistypen aus den Newsitems herauszufiltern und als Annotationen zu retournieren.
Die extrahierten Orte, Zeiten und Ereignistypen je Newsitem dieser Komponenten werden vom Ereignisaggregations-Service genutzt, um diese zu Ereignissen zu aggregieren. Er hat dabei zwei Aufgaben: (1) die extrahierten Annotationen zu Events zusammenzufassen und (2) Events mit gleichem Ort, gleicher Zeit und gleichem Ereignistyp zu Sub-Events zusammenzufassen. Zur Erfüllung der ersten Aufgabe wird eine einfache Logik genutzt bei der jene Orte, Zeiten und Ereignistypen zu Events gruppiert werden, die im jeweiligen Newsitemtext am nächsten beieinander liegen. Auch die zweite Aufgabe wird durch eine einfache Logik gelöst, indem die Orte, Zeiten und Ereignistypen der einzelnen Events verglichen werden und jene zu Sub-Events gruppiert werden, die alle drei gleich haben. Basierend auf der Anzahl der gruppierten Sub-Events kann auch die Zuverlässigkeit der Information evaluiert werden, da die Wahrscheinlichkeit das eine Information wahr ist steigt je mehr Quellen darüber berichten.

2.5 Komponenten zur Aggregierung und Zusammenfassung

Abschließend müssen die gewonnen Informationen sowohl komprimiert als auch in Formate gebracht werden, die für den Endnutzer verständlich und nutzbar sind. Dabei kommen sowohl der Zusammenfassungs-Service, als auch der Ergebnisformatierungs-Service zum Einsatz.

Ersterer hat dabei die Aufgabe die umfangreichen, textuellen Informationen zu kompakten Texten zusammenzufassen. Dafür nutzt der Service ein bestehendes BART-Model, das sowohl für die Abstraktion von langen Texten (Zeitungsartikeln etc.), als auch für die Zusammenfassung von kurzen Texten (Social Media) geeignet ist. BART-Modelle sind transformer-basiert und nutzen neuronale Netzwerke. Dieser Service ermöglicht es dem Nutzer auch die Länge der Zusammenfassung zu bestimmen. Als Input benötigt der Service den unstrukturierten Text, sowie eine minimale und maximale Länge

der Zusammenfassung. Der Text wird dann vom Modell im Hintergrund zusammengefasst und der Service retourniert die Zusammenfassung.
Der Ergebnisformatierungs-Service hingegen generiert drei Ergebnisse, die dem Nutzer zur Verfügung stehen: (1) Eine GeoJSON-Datei die mittels Virtual Reality analysiert werden kann, (2) einen PDF-Bericht mit wesentlichen Informationen und (3) einen Bericht mit allen Fehlermeldungen des Systems. Für die erste Aufgabe muss der Report Generator die aus den Newsitems extrahierten Events in eine GeoJSON Struktur formatieren, damit diese in die VR eingelesen und visualisiert werden können. Für die Erstellung der Berichte nutzt der Service die Python-Bibliothek WeasyPrint[47], die eine flexible Gestaltung der Berichtsvorlagen mittels HTML und CSS-Bausteinen ermöglicht, wodurch moderne Webstandards eingehalten werden. Basierend auf der vorab definierten Berichtsstruktur generiert der Report Generator PDF-Berichte, die die entsprechenden Inhalte enthalten.

[47] Weasyprint, "WeasyPrint - The Awesome Document Factory.," 2023, https://weasyprint.org/.

Teil C: Das GeoCROW-System

1 Datenanalyse

1.1 Analyse eines Newsitems

Im GeoCROW-System beschreibt der Begriff Newsitem eine gecrawlte Dateninstanz. Der Typ dieser Dateninstanz kann ein einzelner Zeitungsartikel, ein Blog-Post oder ein Social Media Post sein. Das entscheidende Attribut eines Newsitems ist sein Inhalt, weil auf diesem das Hauptaugenmerk der Analyse liegt. Allerdings wird die Analyse auch durch optionale gecrawlte Attribute, wie zum Beispiel dem Zeitstempel (Zeitpunkt des Uploads des Newsitems) oder dem Titel, bereichert. Im GeoCROW-System werden die Newsitems, bestehend aus ihrem Inhalt und den optionalen Attributen, zunächst einzeln analysiert und wertvolle Informationen extrahiert.
Grundsätzlich ist anzumerken, dass das GeoCROW-System die gewonnenen Informationen nach jedem Analyseschritt in den GeoCROW-Wissensgraphen speichert. Von diesem bezieht es auch die benötigten Daten für den nächsten Analyseschritt. Mit Blick auf ein einzelnes Newsitem gibt es zunächst 6 entscheidende Analyseschritte:

- Übersetzen des Inhalts und des Titels ins Englische
- Extrahieren temporaler Informationen aus dem Inhalt und dem Titel
- Extrahieren räumlicher Informationen aus dem Inhalt und dem Titel
- Extrahieren von Ereignistypen aus dem Inhalt und dem Titel
- Zusammenfassen des Inhalts
- Das Formen von Ereignissen aus den gewonnenen temporalen, räumlichen Informationen und den Ereignistypen

Zu Beginn der Verarbeitung des Newsitems wird der Inhalt und der Titel ins Englische übersetzt. Der übersetzte Inhalt wird mittels eines Sprachmodells zusammengefasst, um die Dichte der Informationen zu erhöhen und langwierige Texte bei der Visualisierung zu vermeiden. In einem weiteren Schritt werden aus den übersetzten Inhalten und Titeln zeitliche Komponenten extrahiert. Die gewonnen Zeitannotationen bestehen dabei aus der Position des Anfangscharakters und des Endcharakters im geparsten Text, sowie dem

extrahierten Zeitstempel. Gleichzeitig mit der Analyse temporaler Informationen wird eine Analyse der räumlichen Informationen durchgeführt. Dabei werden Ortsnamen aus dem Titel und dem Inhalt extrahiert und anschließend georeferenziert, d.h. es werden Koordinaten zugewiesen. Aus dem Titel und dem Inhalt werden außerdem jene Ereignistypen extrahiert, die zuvor in der GeoCROW-Taxonomie definiert wurden.
Der entscheidende Schritt bei der Verarbeitung eines Newsitems ist nun die Extraktion von Ereignissen aus den gewonnenen temporalen und räumlichen Informationen, sowie den Ereignistypen. Denn während ein Ereignistyp lediglich die Art des Ereignisses beschreibt, dass im Text beschrieben wird, ist ein Ereignis eine Kombination der gefundenen Informationen in diesem Text. D.h. ein Ereignis besitzt eine zeitliche Komponente, eine räumliche Komponente und einen spezifischen Ereignistyp. Dabei können auch multiple Ereignisse im gleichen Newsitem enthalten sein, wenn entsprechend viele Annotationen vorhanden sind. Der Ansatz, der im GeoCROW-System zur Extraktion der Ereignisse verwendet wird, basiert auf der Annahme, dass Annotationen, die sich auf das gleiche Ereignis beziehen, auch im Text nahe beieinander liegen. D.h., Annotationen deren Position im Text näher beieinander liegen werden zu einem Ereignis geformt.

1.2 Die GeoCROW-Ontologie

Die GeoCROW-Ontologie beschreibt die Datenstruktur und Zusammenhänge der extrahierten Aspekte der Newsitems. Sie bietet die Grundlage für die Datenspeicherung, die Interoperabilität und Transparenz des Systems.

Die Ontologie stellt das Event, also ein Geschehen oder Ereignis, als zentrale Entität in den Vordergrund. Um eine holistische Darstellung zu ermöglichen, definiert die Ontologie weiters verwandte Konzepte sowie deren Beziehungen: EventType, Time, Place, and Actor. Grundlage für die Modellierung dieses Teils der Ontologie stellt die Simple Event Model Ontology[48]

[48] "Simple Event Modelling Ontology," accessed August 23, 2024, https://semantic-web.cs.vu.nl/2009/11/sem/.

dar, welche jedoch erweitert wurde, um den spezifischen Anforderungen dieses Anwendungsfalls gerecht zu werden.

Neben der Darstellung der Ereignisse selbst, modelliert die Ontologie weiters die Verbindung zwischen Ereignissen und Textressourcen. EventReports, die schriftliche Berichterstattung über ein Ereignis, sind mit NewsItems wie etwa Nachrichtenartikeln oder Posts in sozialen Medien verknüpft. Diese EventReports dienen als Brücke zwischen der strukturierten Welt der Events und dem unstrukturierten Bereich der natürlichen Sprache. Um die Extraktion relevanter Informationen aus Textressourcen zu erleichtern, enthält die Ontologie Elemente aus der NIF-Ontologie[49], einem Rahmenwerk, das speziell für Annotationen im Bereich der natürlichen Sprachverarbeitung (NLP) entwickelt wurde. Durch die Verknüpfung von NewsItems zu Kontexten und EventReports zu einer Sammlung an zusammengehörigen zeitlichen, räumlichen, und typbasierten Annotationen, ermöglicht die Ontologie die präzise Identifizierung und Extraktion von ereignisbezogenen Details aus Textinhalten, und unterstützt die Rückführbarkeit der finalen Ergebnisse auf die zugrundeliegenden Daten, was die Transparenz und Überprüfbarkeit des Systems verbessert.

[49] Sebastian Hellmann, Martin Brümmer, and Markus Ackermann, "NLP Interchange Format (NIF)," 2015, https://nif.readthedocs.io/en/latest/.

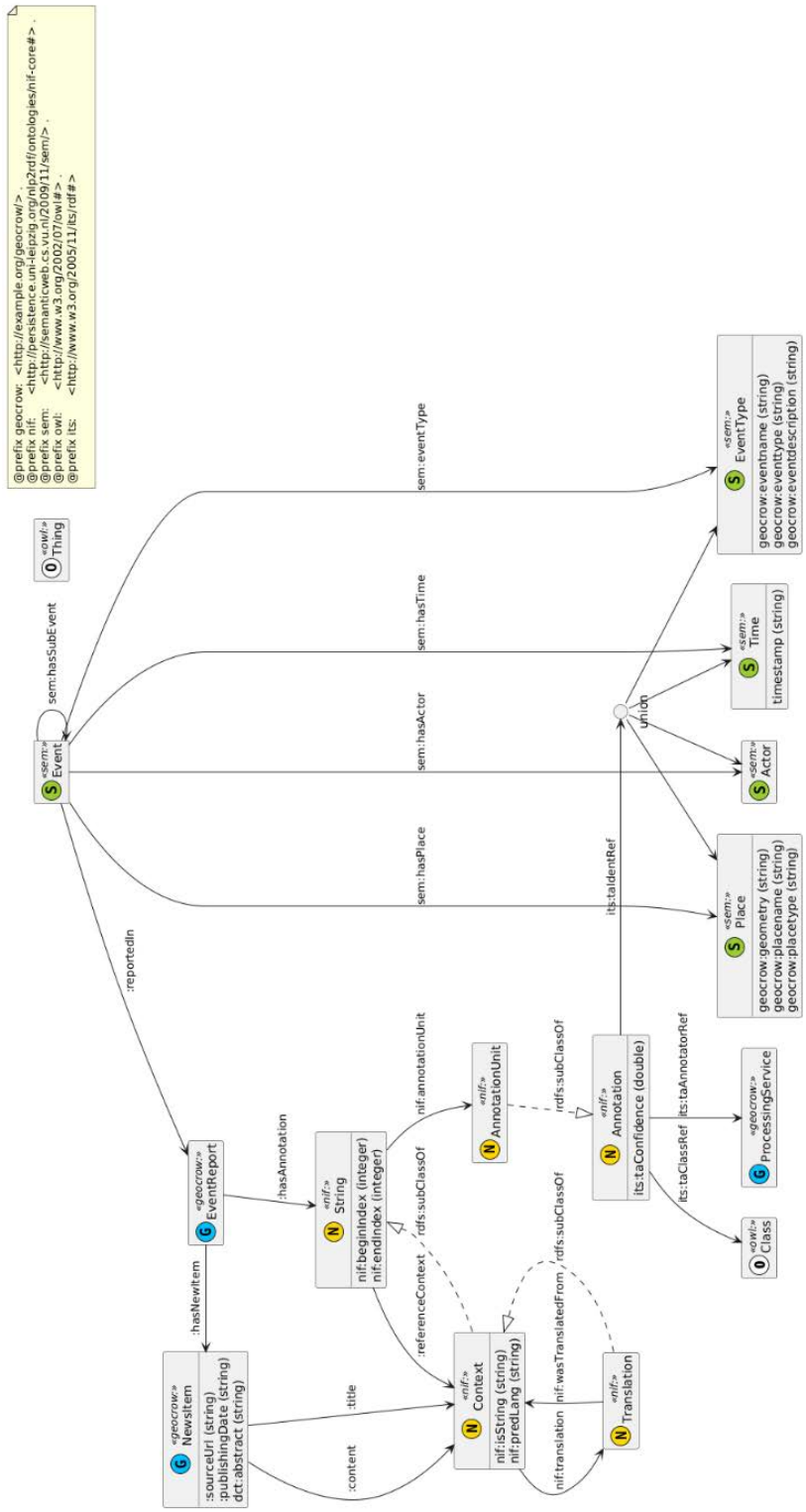

Abbildung 3: Die GeoCROW-Ontologie, die auf der Simple Event Ontologie und NIF-Ontologie basiert und diese um verschiedene projektrelevante Konzepte erweitert.

1.3 Der GeoCROW-Wissensgraph

Der Wissensgraph nutzt die zuvor beschriebene Ontologie als Grundlage, um Informationen über extrahierte Ereignisse in strukturierter Form zu speichern. Dabei wird nicht nur das Endergebnis des Extraktionsprozesses im Wissensgraphen abgelegt, sondern dieser wird nach jedem einzelnen Verarbeitungsschritt kontinuierlich mit den entsprechenden Triples aktualisiert. Dies bringt mehrere entscheidende Vorteile mit sich.

Einerseits erhöht diese Vorgehensweise die Robustheit des Gesamtsystems. Durch die fortlaufende Aktualisierung des Wissensgraphen ist gewährleistet, dass der Ausfall eines Teilsystems die Funktionsfähigkeit und den Datenzugriff der anderen Teilsysteme nicht beeinträchtigt. Das System bleibt somit widerstandsfähiger gegenüber Störungen.

Andererseits verbessert diese Methodik die Nachvollziehbarkeit und Wartbarkeit des Systems erheblich. Fehlerhafte Extraktionen lassen sich durch die detaillierte und schrittweise Dokumentation der Verarbeitungsschritte leichter zu ihrem Ursprung zurückverfolgen. Dies ermöglicht eine gezielte Fehlerbehebung und vereinfacht die kontinuierliche Weiterentwicklung und Optimierung des Systems.

Für den Endanwender ergibt sich daraus zudem ein entscheidender Vorteil in Bezug auf die Transparenz. Da die Originalquellen parallel zu den verarbeiteten Daten bereitgestellt werden, kann der Nutzer die Herkunft und den Kontext der angezeigten Ereignisse besser nachvollziehen. Dies fördert ein tieferes Verständnis der präsentierten Informationen und ermöglicht eine fundiertere Einschätzung der Daten.

2 Einordnung der Kernfunktionen des GeoCROW-Systems

2.1 Extraktion der räumlichen Komponenten

Es gibt viele Studien, die sich mit der Extraktion von räumlichen Komponenten aus unstrukturierten Texten beschäftigen[50, 51, 52, 53, 54]. Der Prozess zur Extrahierung räumlicher Komponenten (Geoparsing) besteht im Allgemeinen aus zwei Schritten, (1) der Erkennung von Ortsbeschreibungen im Text und (2) der Georeferenzierung der erkannten Ortsbeschreibungen, d.h. der Zuweisung von Koordinaten und räumlicher Repräsentationen.
Beispielsweise analysieren Gelernter et al.[55] verschiedene Ansätze zur Identifikation von Orten in Social Media Nachrichten im Katastrophenkontext. Dabei zeigen sie auch Forschungslücken im Bereich des Geoparsing auf. In einer weiteren Studie präsentieren sie einen Geoparser, der speziell dafür entwickelt wurde Ortsbezeichnungen aus Mikrotexten (z.B. Twitter-Texten) zu extrahieren. Dabei nutzen sie Heuristiken, Named Entity Recognition und Machine Learning. Ein besonderer Ansatz für Mikrotexte wird benötigt, weil Ortsbezeichnungen in Ihnen oftmals abgekürzt oder falsch geschrieben sind[56].

---

[50] Jimin Wang, Yingjie Hu, and Kenneth Joseph, "NeuroTPR: A Neuro- net Toponym Recognition Model for Extracting Locations from Social Media Messages," *Transactions in GIS* 24, no. 3 (June 2020): 719–35, https://doi.org/10.1111/tgis.12627.

[51] Cillian Berragan et al., "Transformer Based Named Entity Recognition for Place Name Extraction from Unstructured Text," *International Journal of Geographical Information Science* 37, no. 4 (April 3, 2023): 747–66, https://doi.org/10.1080/13658816.2022.2133125.

[52] Praval Sharma et al., "A Spatially-Aware Algorithm for Location Extraction from Structured Documents," *GeoInformatica* 27, no. 4 (October 2023): 645–79, https://doi.org/10.1007/s10707-022-00482-1.

[53] Liufeng Tao et al., "Geographic Named Entity Recognition by Employing Natural Language Processing and an Improved BERT Model," *ISPRS International Journal of Geo-Information* 11, no. 12 (November 28, 2022): 598, https://doi.org/10.3390/ijgi11120598.

[54] Morteza Karimzadeh et al., "GeoTxt: A Scalable Geoparsing System for Unstructured Text Geolocation," *Transactions in GIS* 23, no. 1 (February 2019): 118–36, https://doi.org/10.1111/tgis.12510.

[55] Judith Gelernter and Nikolai Mushegian, "Geo- parsing Messages from Microtext," *Transactions in GIS* 15, no. 6 (December 2011): 753–73, https://doi.org/10.1111/j.1467-9671.2011.01294.x.

[56] Judith Gelernter and Shilpa Balaji, "An Algorithm for Local Geoparsing of Microtext," *GeoInformatica* 17, no. 4 (October 2013): 635–67, https://doi.org/10.1007/s10707-012-0173-8.

Auch die Analyse anderer Arbeiten hat gezeigt, dass für den ersten Schritt des Geoparsings typischerweise Named Entity Recognition (NER) genutzt wird. Bei NER werden Sprachmodelle genutzt, um bestimmte Bestandteile des Textes zu annotieren (hier Ortsbezeichnungen). Die extrahierten Ortsbezeichnungen können dann in einem zweiten Schritt disambiguiert und georeferenziert werden. Beim Disambiguieren geht es darum für einen detektierten Ortskandidaten festzustellen, ob es sich wirklich um einen Ortsnamen handelt und wenn ja um welchen Ort. Beispielsweise kann der Begriff „Paris" im Text einen Namen darstellen oder das Paris in Frankreich meinen oder das Paris in den USA bezeichnen. Ambiguitäten wie diese müssen folglich zunächst behoben werden. Nachdem festgestellt wurde, dass es sich um einen Ortsnamen handelt und um welchen, muss eine räumliche Repräsentation bzw. Koordinate bestimmt werden. Generell wird dieser zweite Schritt (Georeferenzierung) oftmals mittels Gazetteer-Matching gelöst. D.h. die identifizierten Ortsbezeichnungen werden mit Ortsverzeichnis abgeglichen. Dieses Ortsverzeichnis enthält für jeden enthaltenen Ort Koordinaten und diese werden den Annotationen zugewiesen[57, 58, 59].
Ein eben solcher Ansatz kommt im GeoCROW-System zum Einsatz. Der Algorithmus Mordecai3 wird genutzt, um Ortsbezeichnungen zunächst in den unstrukturierten Texten zu identifizieren und mittels Gazetteer-Matching zu georeferenzieren. Zur Disambiguierung werden potenzielle Orte aus dem Ortsverzeichnis Geonames abgefragt. Es wurde Geonames als Ortsverzeichnis ausgewählt, da es open-source ist, über 25 Millionen geographische Namen enthält und eine globale Abdeckung hat. Die potenziellen Kandidaten, die aus dem Ortsverzeichnis abgefragt wurden, werden durch ein neuronales Netzwerk gereiht und jenes mit dem besten Wert wird schließlich zur Georeferenzierung genutzt[60].

---

[57] Karimzadeh et al., "GeoTxt."

[58] Stuart E. Middleton et al., "Location Extraction from Social Media: Geoparsing, Location Disambiguation, and Geotagging," *ACM Transactions on Information Systems* 36, no. 4 (October 31, 2018): 1–27, https://doi.org/10.1145/3202662.

[59] Elise Acheson and Ross S. Purves, "Extracting and Modeling Geographic Information from Scientific Articles," ed. Song Gao, *PLOS ONE* 16, no. 1 (January 6, 2021): e0244918, https://doi.org/10.1371/journal.pone.0244918.

[60] Andrew Halterman, "Mordecai 3: A Neural Geoparser and Event Geocoder" (arXiv, March 23, 2023), http://arxiv.org/abs/2303.13675.

2.2 Extraktion von temporären Komponenten

Das Parsen von Datumsangaben ist eine seit langem etablierte Herausforderung, die zur Entwicklung zahlreicher spezialisierter Werkzeuge und Bibliotheken geführt hat. Nach einer umfassenden Analyse verschiedener solcher Bibliotheken, darunter dateutil[61], dateparser[62] und datefinder[63], wurde für das GeoCROW-System das von der Universität Heidelberg entwickelten Java-Tool HeidelTime[64] ausgewählt.
HeidelTime unterstützt eine breite Palette an Sprachen und Datumsformaten und ist in der Lage, die Art der Annotation—sei es ein vollständiges Datum, ein relatives Datum oder eine andere Zeitangabe—präzise zu erfassen und diese direkt im Text zu annotieren. Relative Zeitangaben spielen in diesem Anwendungsfall eine besonders wichtige Rolle, da sie in Nachrichtenbeiträgen deutlich häufiger vorkommen als absolute Datumsangaben. Beispielsweise wird häufig auf Formulierungen wie „letzter Donnerstag" oder „in den vergangenen Monaten" Bezug genommen. Um diese relativen Zeitangaben präzise zu interpretieren, wird das Publikationsdatum des jeweiligen Nachrichtenartikels als Referenzpunkt herangezogen.

2.3 Extraktion von Ereignistypen

Das Herzstück der Extraktion von Ereignis Typen ist der PoolParty Extraktor[65], der auf einer sorgfältig ausgearbeiteten Taxonomie aufbaut.
Diese Taxonomie (siehe Abbildung 4) dient als umfassende hierarchische Wissensbasis aller relevanten Ereignistypen, die in einem kooperativen, iterativen Prozess unter Einbeziehung der relevanten Interessengruppen entwickelt wurde. Um wertvolle kontextbezogene Informationen zu liefern und

---

[61] Niemeyer Niemeyer et al., "Dateutil," accessed August 23, 2024, https://github.com/dateutil/dateutil.
[62] "Dateparser," accessed August 23, 2024, https://github.com/scrapinghub/dateparser.
[63] "Datefinder - Extract Dates from Text," accessed August 23, 2024, https://github.com/akoumjian/datefinder.
[64] Strötgen and Gertz, "A Baseline Temporal Tagger for All Languages."
[65] Semantic Web Company, "PoolParty Extractor," accessed August 23, 2024, https://www.poolparty.biz/poolparty-extractor.

die Interoperabilität mit anderen Datenquellen zu erleichtern wurden Verbindungen zu relevanten Konzepten aus Wikidata[66] integriert. Darüber hinaus wurde die Taxonomie um Konzepte und Labels aus EventWiki[67] erweitert.
Um eine effektive Identifizierung zu gewährleisten, ist jeder Ereignistyp sowohl mit englischen als auch mit deutschen Bezeichnungen und Synonymen versehen. Weiters beinhaltet jedes Konzept eine Beschreibung des jeweiligen Ereignistypen, um die Abgrenzung zu anderen Konzepten klarer zu machen, sowie die Disambiguierung zu erleichtern.
Diese Taxonomie vermittelt dem PoolParty Extractor ein umfassendes Verständnis von Ereignistypen, sowie deren Beziehungen, wodurch er in der Lage ist, Ereignisse im Text genau zu erkennen und zu kategorisieren, auch wenn sie in verschiedenen sprachlichen Formen oder Kontexten ausgedrückt werden.
Einer der Vorteile dieses Ansatzes ist seine große Flexibilität und Anpassungsfähigkeit. Wenn ein neues Konzept oder ein neuer Ereignistyp, welcher noch nicht abgebildet ist, an Relevanz gewinnt, kann dieser nahtlos in die Taxonomie integriert werden, ohne dass ein umfangreiches Neutraining oder eine Neukalibrierung des Modells erforderlich ist. Diese Flexibilität stellt sicher, dass der Prozess der Ereignistyp-Extraktion aktuell bleibt und auf sich entwickelnde Informationslandschaften reagieren kann.

[66] "Wikidata," accessed August 23, 2024, https://www.wikidata.org/wiki/Wikidata:Main_Page.
[67] Tao Ge et al., "EventWiki: A Knowledge Base of Major Events," n.d.

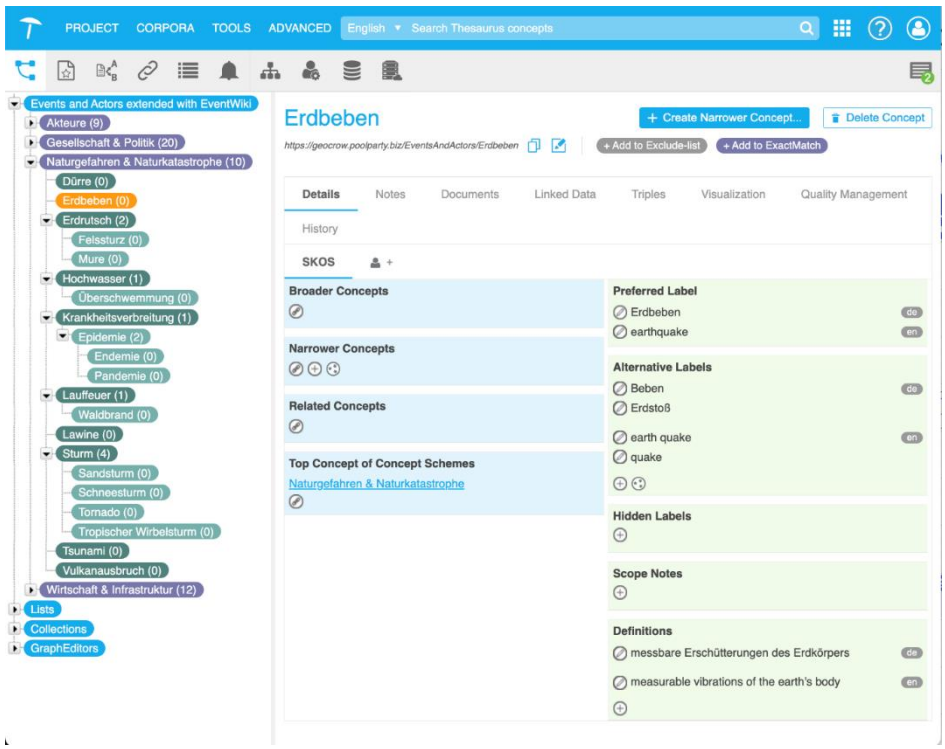

Abbildung 4: Die GeoCROW-Taxonomie im Interface des PoolParty-Thesaurus-Managers.

2.4 Zusammenfassung zu Ereignissen und Events

Nachdem aus den Newsitems die räumlichen und zeitlichen Komponenten, sowie die Ereignistypen extrahiert wurden, werden diese zu Ereignissen und Events aggregiert. Dieser Schritt birgt einige Vorteile für die Visualisierung in der Virtual Reality, auf die in diesem Kapitel eingegangen wird. Zunächst müssen wir jedoch festlegen was der Unterschied zwischen einem (1) *Ereignistyp*, einem (2) *Ereignis* und einem (3) *Event* ist.

(1) Unter einem Ereignistyp verstehen wir die Annotation eines Ereignistypen in einem Newsitem. Ein Newsitem kann folglich mehrere Ereignistypen (z.B.: Erdbeben) besitzen, ebenso wie es mehrere Ortsbeschreibungen (z.B.:

Istanbul) und Zeitstempel (z.B.: 050000ZFEB22) beinhalten kann.
(2) Innerhalb eines Newsitems können aus zusammengehörenden Ereignistypen, Ortsbeschreibungen und Zeitstempeln Ereignisse gebildet werden. Ein Ereignis ist folglich ein Vorfall, der zu einer bestimmten Zeit an einem bestimmten Ort stattgefunden hat und einen bestimmten Ereignistyp hat. Zur Bestimmung welche Ereignistypen, Zeitstempel und Ortsbeschreibungen innerhalb eines Newsitems ein Ereignis bilden, wird angenommen, dass diese im Text nahe beieinander liegen müssen. Folglich werden jene Ereignistypen, Zeitstempel und Ortsbeschreibungen zu einem Ereignis zusammengefasst, die die geringste Distanz im Text aufweisen.
(3) Multiple Ereignisse aus verschiedenen Newsitems können wiederum zu Events zusammengefasst werden. Damit Ereignisse zu einem Event aggregiert werden können, müssen sie Ähnlichkeiten hinsichtlich ihrer Ortsbeschreibung, ihres Zeitstempels und ihres Ereignistyps aufweisen.
Durch die Aggregation der gewonnenen Informationen zu Ereignissen und Events werden neue Möglichkeiten zur Visualisierung in der Virtual Reality erschlossen. So können nun die verschiedenen Ereignisse nach Ereignistyp klassifiziert und entsprechend visualisiert werden. Außerdem können Ereignisse entsprechend ihrer Georeferenzierung räumlich präsentiert werden. Auch eine räumliche Clusterung (z.B. nach Ereignistyp) ist möglich. Durch die gewonnenen temporalen Informationen kann die zeitliche Entwicklung der Ereignisse verfolgt werden. Die Zusammenfassung zu Events ermöglicht es wiederum größere Zusammenhänge zwischen einzelnen Ereignissen hervorzuheben.

3 Visualisierung der extrahierten Informationen

Dieser Abschnitt beschreibt die immersive Visualisierung unter Anwendung von Virtual Reality (VR), die im GeoCROW-System durchgeführt wird. Die Techniken und Methoden zur Visualisierung georeferenzierter Daten in VR, sowie die Interaktion mit der Visualisierung, werden diskutiert.

Die virtuelle Umgebung besteht aus zwei Teilen, namentlich der "Roomscale"- und der "Lifescale"-Umgebung. Die Anforderungsermittlung

mit dem Bedarfsträger ergab die Notwendigkeit der Modi "Tabletop", "Flugmodus" und "Ego-Perspektive". Die in VR realisierte Aufteilung in "Roomscale" und "Lifescale" implementiert diese Anforderung: Beide Umgebungen enthalten eine geografische Karte, unterscheiden sich jedoch stark durch die Skalierung dieser: Während die Karte in der Roomscale-Umgebung ein verkleinertes Modell der echten Welt darstellen soll, stellt die Karte in der Lifescale-Umgebung durch 1:1-Skalierung reale Größenverhältnisse korrekt dar. Die Lifescale-Umgebung kann sowohl zu Fuß ("Ego-Perspektive") als auch fliegend ("Flugmodus") erkundet werden.

Beide Umgebungen enthalten eine geografische Karte, unterscheiden sich jedoch stark durch die Skalierung dieser: Während die Karte in der Roomscale-Umgebung ein verkleinertes Modell der echten Welt darstellen soll, stellt die Karte in der Lifescale-Umgebung durch 1:1-Skalierung reale Größenverhältnisse korrekt dar.

Im Zentrum der Roomscale-Umgebung befindet sich ein ca. ein Meter hoher Kartentisch. Auf diesem wird eine verkleinerte, interaktive geografische Karte dargestellt. Die Visualisierung des GeoCROW Knowledge Graph erfolgt auf dieser Karte. Die gesamte Umgebung ist in Abbildung 2 dargestellt. Eine Ansicht aus Sicht des Benutzers findet sich in Abbildung 3. Beim Start der Anwendung befindet sich der Benutzer in der Roomscale-Umgebung.

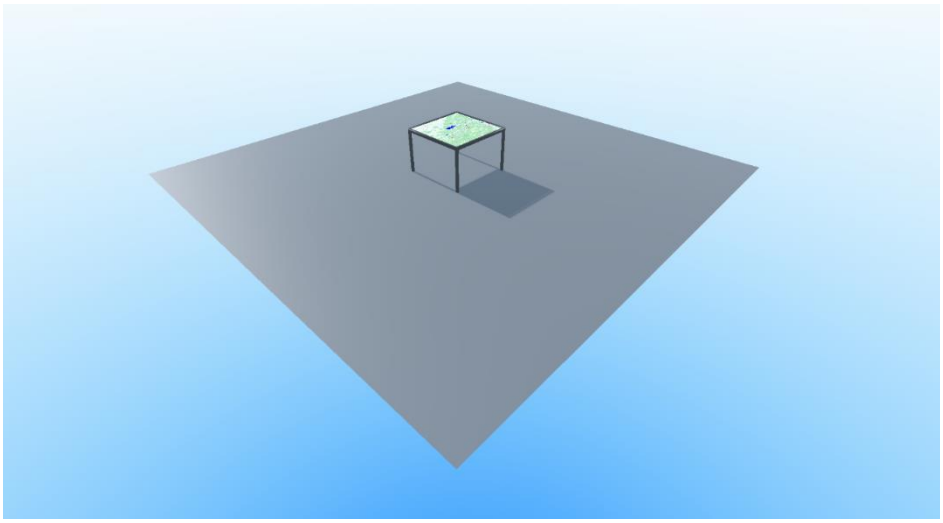

Abbildung 5: Roomscale-Umgebung.

Die Lifescale-Umgebung stellt einen ausgewählten Ausschnitt der geografischen Karte in Originalgröße dar. Daher sind alle dargestellten Objekte in dieser Umgebung so groß wie in der physischen Welt. Da die visualisierte Fläche daraus resultierend sehr groß ist, und die Größe eines Raumes um ein weites übersteigt, sind Benutzer in dieser Umgebung auf VR- Fortbewegungsmethoden angewiesen, um die dargestellten Inhalte sinnvoll zu erkunden. Auch diese Karte enthält die Visualisierungen des Knowledge Graph. Eine Vogelperspektive der Lifescale-Umgebung ist in Abbildung 4 dargestellt. Abbildung 5 zeigt die Lifescale-Umgebung aus der Sicht des Benutzers.

Visualisierung des Knowledge Graph

In der Definition des Knowledge Graph gibt es drei verschiedene Ereignistypen: Point, LineString und MultiPolygon.

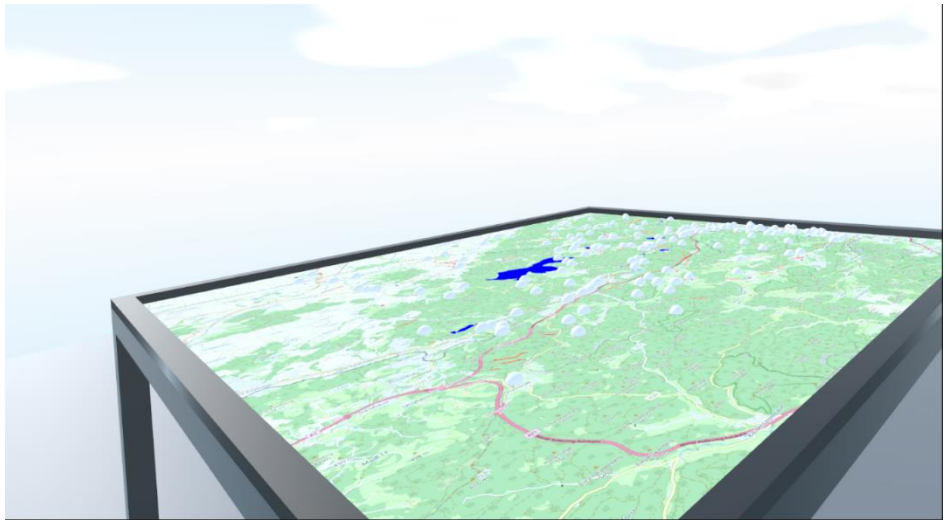

Abbildung 6: Roomscale-Umgebung aus Benutzersicht.

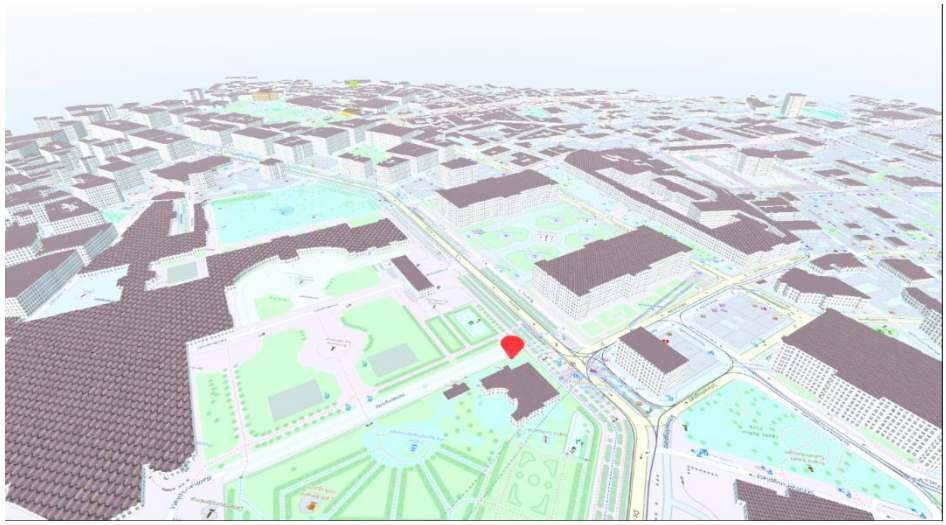

Abbildung 7: Lifescale-Umgebung im Flugmodus. Roter Marker: Benutzerposition in Abbildung 5.

Ein Point-Feature stellt dabei ein punktförmiges Ereignis dar, dessen Ort sich auf einen einzigen geografischen Punkt beschränkt. Beispielhaft hierfür wäre der Einsturz eines Gebäudes, bei dem der Punkt der Adresse des Gebäudes entspricht.

LineStrings sind linienförmige Ereignisse. Bei diesen bezieht sich der Ereignisort auf eine geografische Linie, die mehrere Kurven haben kann. Ein Beispiel für ein solches Ereignis wäre ein Stau, bei dem die Linie dem Verlauf der Straße, auf der sich dieser befindet, folgt.

MultiPolygons stellen flächenförmige Ereignisse dar. Dabei bezieht sich der Ereignisort auf eine Fläche, die mehrere Ecken und eine zufällige Form haben kann. Waldbrände oder andere Großereignisse können durch diese Polygone repräsentiert werden.

Die Visualisierung dieser Ereignistypen findet in der virtuellen Umgebung sowohl in der Roomscale-, als auch in der Lifescale-Umgebung statt und ist das Kernelement der VR-Anwendung.

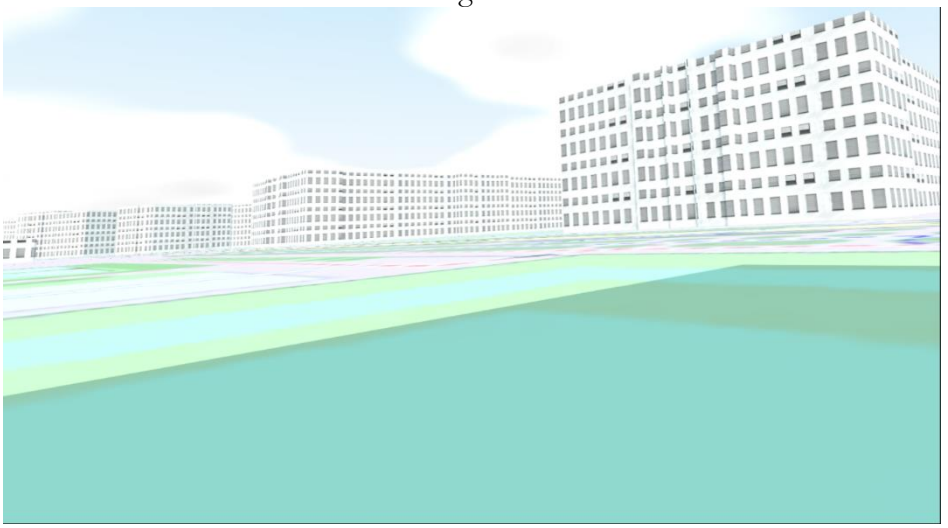

Abbildung 8: Lifescale-Umgebung aus Benutzersicht.

In der Roomscale-Umgebung findet die Visualisierung auf dem Kartentisch statt. Point-Features werden dabei durch Halbkugeln auf der Karte dargestellt. Die Visualisierung von LineString-Features erfolgt durch farbige Linien. MultiPolygon-Features werden durch farbige, erhabene Polygone repräsentiert. Um eine Überladung des Bildschirms zu vermeiden, kommt hier Clustering zum Einsatz. Dabei werden mehrere Features zu einem Cluster zusammengefasst und dieser durch eine Säule dargestellt. Abbildung 6 zeigt die Visualisierung einer größeren Menge von Features verschiedener Typen

und Clustern aus der Anwenderperspektive.

Abbildung 9: Ereignisvisualisierung im Roomscale-Modus aus Anwendersicht.

Die Darstellung in der Lifescale-Umgebung verwendet zusätzlich zu der Visualisierung durch Polygone und Linien den Clustern ähnliche Säulen. Diese sind aus großer Entfernung sichtbar und überragen auch etwaige im Sichtfeld befindliche Gebäude, sodass die Orientierung und Zielfindung erleichtert wird. Abbildung 7 und 8 zeigen Visualisierungen im Lifescale-Modus aus der Ego-Perspektive und dem Flugmodus.

POIs und Gebäude

Points of Interest (POIs) werden in der virtuellen Umgebung im Roomscale-Modus durch semitransparente Objekte dargestellt. Punktförmige POIs wie Bushaltestellen oder Fahrradparkplätze werden durch Säulen visualisiert, während linienförmige POIs durch Linien und flächenförmige POIs durch Umrisslinien dargestellt werden. Dabei können die folgenden POI-Typen an- oder abgewählt werden:

- Fahrradparkplatz
- Bushaltestelle
- Baum
- Krankenhaus

- Schule
- Flughafen
- Brücke
- Fluss
- Eisenbahnlinie

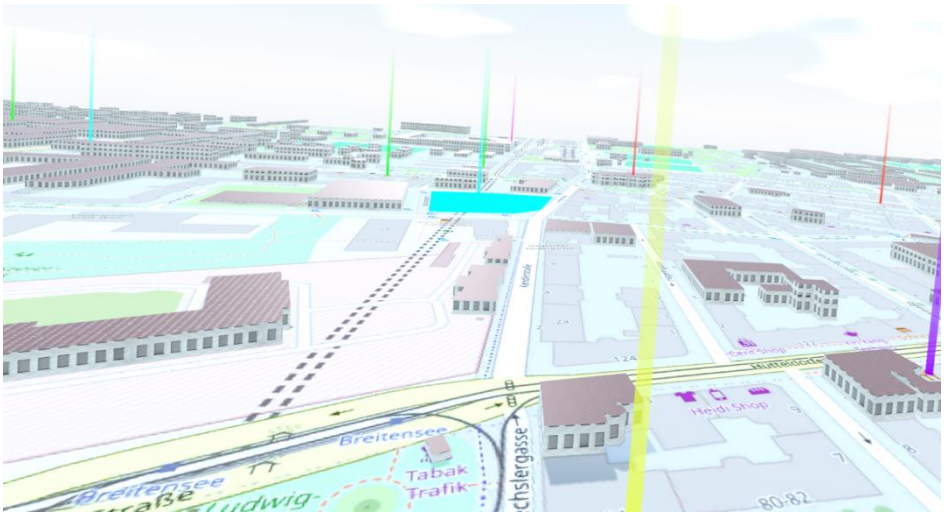

Abbildung 10: Ereignisvisualisierung im Lifescale-Modus im Flugmodus.

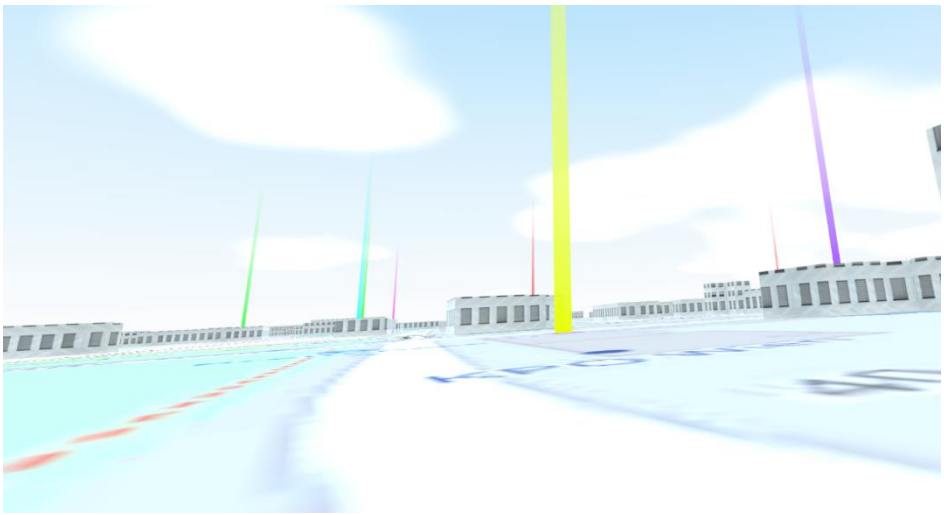

Abbildung 11: Ereignisvisualisierung im Lifescale-Modus aus der Ego-Perspektive.

Darüber hinaus können in der Roomscale-Ansicht Gebäude dargestellt werden. Diese werden in der Lifescale-Ansicht standardmäßig zur Verbesserung des geografischen Verständnisses angezeigt, in der Roomscale-Ansicht sind sie optional. Die Gebäude werden durch ein vereinfachtes Modell, das den in der Karte eingezeichneten „Fußabdruck“ realer Gebäude um die von der Karte bereitgestellten Gebäudehöhendaten extrahiert und mit einer Textur versehen. Ein Beispiel für die POI-Visualisierung mit Gebäuden findet sich in Abbildung 9.

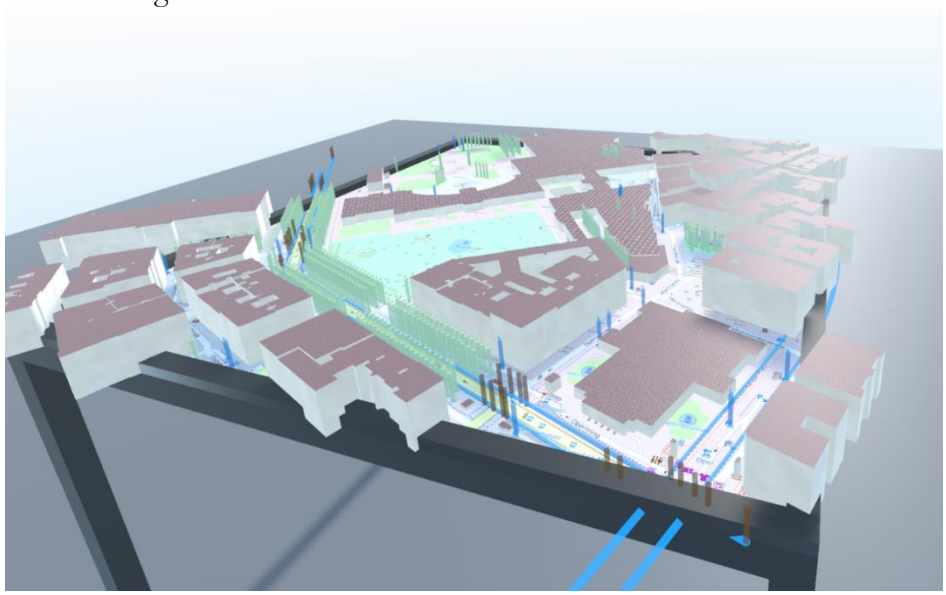

Abbildung 12: POIs und Gebäude im Roomscale-Modus aus Anwendersicht.

Interaktion

Die Karte kann im Roomscale-Modus mit von Smartphone-Anwendungen bekannten Gesten gesteuert werden. Dazu zeigt der Anwender mit dem VR-Controller auf die Karte und hält den Grip-Button des Controllers (siehe Abbildung 10). Nun kann die Karte an der Hand gezogen werden und so der dargestellte geografische Ort verändert werden. Beim Loslassen des Grip-Buttons endet die Interaktion und die Karte bleibt am ausgewählten Ort

Um die Karte zu zoomen, wird mit dem Controller auf den Punkt, auf den gezoomt werden soll, gezeigt und währenddessen der Thumbstick nach vorn

(hineinzoomen) oder zurück (herauszoomen) bewegt.

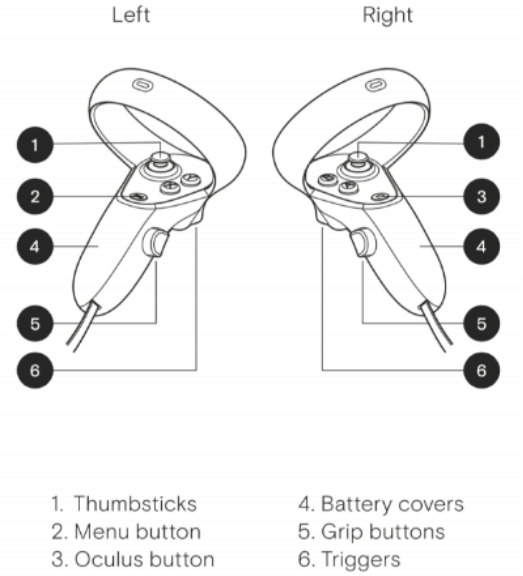

Abbildung 13: Bedienelemente des VR-Controllers.

Zum Wechsel zwischen Roomscale- und Lifescale-Modus muss der Anwender auf der Karte einen Marker platzieren (siehe Abbildung 11), der auf die Stelle gesetzt wird, an der er sich im Lifescale-Modus befinden möchte. Hat der Anwender diesen Marker gesetzt, so gelangt er durch Knopfdruck auf dem VR-Controller in den Lifescale-Modus. Hier kann derselbe Knopf wieder gedrückt werden, um zurück in den Roomscale-Modus zu kommen.

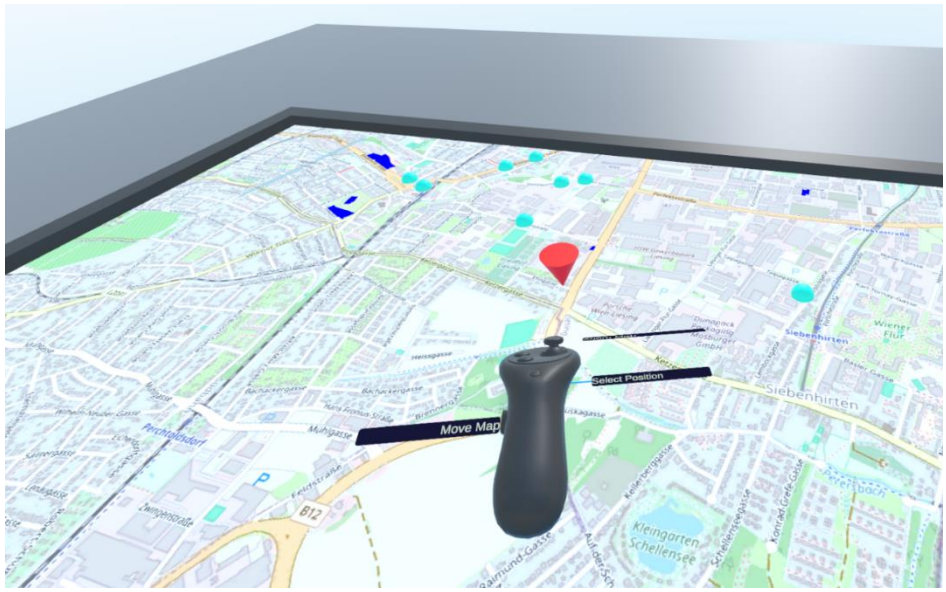

Abbildung 14: Platzierung des Markers für den Moduswechsel

Im Roomscale-Modus besteht die Möglichkeit der Auswahl von Ereignissen zur zusätzlichen Anzeige von Details. Hierfür zeigt der Anwender mit dem Controller auf die Visualisierung (Halbkugel, Linie, Polygon) eines Ereignisses und wählt dieses mit der Trigger-Taste aus. Nun öffnet sich ein kleines Menü (siehe Abbildung 12), das verschiedene Metadaten zum Ereignis anzeigen kann.

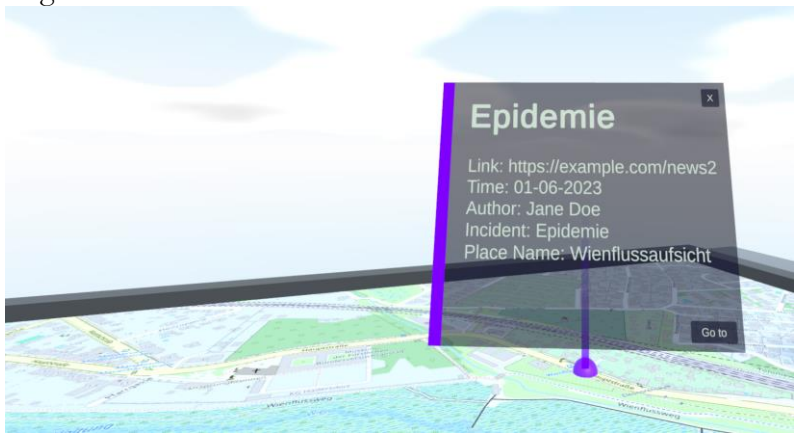

Abbildung 15: Darstellung von Details eines Ereignisses

Konklusion

Das GeoCROW-Projekt hat sich als ein innovatives und bedeutendes Vorhaben etabliert, das semantische Technologien, Crowdsourcing und Virtual Reality (VR) kombiniert, um die situative Wahrnehmung in humanitären Kriseneinsätzen entscheidend zu verbessern. In einer Welt, in der Umwelt- und Naturkatastrophen, politische Unruhen und humanitäre Krisen zunehmen, bietet GeoCROW einen technologischen Rahmen, der den Entscheidungsprozess für Einsatzkräfte optimiert. Durch die Zusammenführung von Daten aus unterschiedlichsten Quellen wie sozialen Medien, Nachrichtenplattformen und geographischen Datenbanken entsteht ein Bild der Lage vor Ort, das durch immersive VR-Technologien visualisiert wird.
Das Herzstück von GeoCROW ist eine modular aufgebaute Systemarchitektur, die verschiedene technologische Fortschritte integriert, um die Bedürfnisse von Einsatzkräften optimal zu bedienen. Dabei kann das GeoCROW-System große Mengen an unstrukturierten Daten aus verschiedenen Quellen effizient analysieren. Die gewonnenen Informationen werden in einem Wissensgraphen gespeichert, der auf einer eigens entwickelten Ontologie basiert. Der Wissensgraph ermöglicht eine strukturierte Speicherung und eine schnelle Abfrage von Ereignissen.
Ein weiterer zentraler Aspekt ist die Integration von Virtual Reality. Die VR-Umgebung ermöglicht es Einsatzkräften, die Lage interaktiv und immersiv zu erkunden. Dabei erhalten Nutzer im Roomscale-Modus eine Überblicksperspektive, während sie im Lifescale-Modus die Möglichkeit haben einzelne Regionen in Originalgröße zu erkunden. Diese immersive hilft Entscheidungsträgern komplexe Informationen schnell zu erfassen und besser zu verstehen.
Die Stärke von GeoCROW liegt in seiner Fähigkeit, humanitäre und militärische Einsatzkräfte gezielt bei der Lagebeurteilung und Einsatzplanung zu unterstützen. Durch die Analyse verschiedenster Datenquellen ermöglicht das System die schnelle Erfassung der Lage. Die beiden Fallbeispiele Niger und Türkei zeigen, dass das System in verschiedenen Szenarien eingesetzt werden kann. Dabei ist vor allem die frühzeitige Identifikation von Risiken und Gefahren ist entscheidend, um die Sicherheit von Einsatzkräften zu gewährleisten. GeoCROW stellt sicher, dass relevante Informationen bereits in der Vorbereitungsphase gesammelt und analysiert werden. Dies erleichtert

die Einsatzplanung und ermöglicht eine rasche und fundierte Entscheidungsfindung.
Allerdings bringt das GeoCROW-System auch ethische und rechtliche Herausforderungen mit sich. So wirft die Verarbeitung von großen Mengen personenbezogener Daten, insbesondere aus sozialen Medien, wichtige Fragen hinsichtlich des Datenschutzes und der Grundrechte auf. Das Projekt legt großen Wert darauf, die Anonymität der Daten zu wahren und die Privatsphäre der betroffenen Personen zu schützen. Durch eine konsequente Anwendung von Datenschutz-Folgenabschätzungen und die Implementierung von „Privacy by Design"-Prinzipien werden mögliche Grundrechtseingriffe minimiert.
Ein weiterer wichtiger Aspekt ist die ethische Verantwortung im Umgang mit neuen Technologien. GeoCROW verfolgt das Ziel, den Einsatz von VR und semantischen Technologien so zu gestalten, dass es ethischen Standards entspricht. Dies umfasst sowohl die Verhinderung eines Missbrauchs der gewonnenen Daten als auch die Sicherstellung, dass die Technologie zur Förderung humanitärer Ziele eingesetzt wird.
GeoCROW zeigt, wie semantische Technologien und Virtual Reality die Arbeit in humanitären Kriseneinsätzen unterstützen können. Die Kombination aus automatisierter Datenauswertung und immersiver Visualisierung bietet eine effektive Hilfe bei der Lagebewertung und Einsatzplanung. Besonders die Fallstudien in Niger und der Türkei belegen den praktischen Nutzen des Systems in unterschiedlichen Krisensituationen.
Trotz der erzielten Fortschritte gibt es noch viele Herausforderungen. Die Systemleistung muss in weiteren Einsätzen geprüft und verbessert werden, um sicherzustellen, dass GeoCROW den Anforderungen in verschiedenen Kontexten gerecht wird. Auch die Einhaltung ethischer und rechtlicher Vorgaben erfordert kontinuierliche Aufmerksamkeit, insbesondere bei der Nutzung personenbezogener Daten.
In Zukunft könnte das System weiter ausgebaut und in internationale Katastrophenhilfe-Strukturen integriert werden. GeoCROW hat eine solide Basis geschaffen, auf der aufgebaut werden kann, um die Unterstützung für Einsatzkräfte noch effizienter zu gestalten.

Literaturverzeichnis

Acheson, Elise, and Ross S. Purves. "Extracting and Modeling Geographic Information from Scientific Articles." Edited by Song Gao. *PLOS ONE* 16, no. 1 (January 6, 2021): e0244918. https://doi.org/10.1371/journal.pone.0244918.

Alexander, Larry, and Michael Moore. "Deontological Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/.

Bacher, Victoria. "Regulierungsbedarf Beim Einsatz von Algorithmen in Sozialen Medien Zum Schutz Der Meinungsfreiheit." *ELSA Austria Law Review* 7, no. 1 (2022): 64–69. https://doi.org/10.33196/ealr202201006401.

BBC. "Niger Media Guide," 2023. https://www.bbc.com/news/world-africa-13943665.

———. "Turkey Media Guide," 2023. https://www.bbc.com/news/world-europe-17992011.

Bentham, J. *An Introduction to the Principles of Morals and Legislation*. Dover Publications, 2012. https://books.google.at/books?id=GcnssaU-1wgC.

Berragan, Cillian, Alex Singleton, Alessia Calafiore, and Jeremy Morley. "Transformer Based Named Entity Recognition for Place Name Extraction from Unstructured Text." *International Journal of Geographical Information Science* 37, no. 4 (April 3, 2023): 747–66. https://doi.org/10.1080/13658816.2022.2133125.

Datareportal. "Digital 2023: Turkey," 2023. https://datareportal.com/reports/digital-2023-turkey.

"Datefinder - Extract Dates from Text." Accessed August 23, 2024. https://github.com/akoumjian/datefinder.

"Dateparser." Accessed August 23, 2024. https://github.com/scrapinghub/dateparser.

Davis, K. *Ethics of Big Data: Balancing Risk and Innovation*. O'Reilly Media, 2012. https://books.google.at/books?id=SG5F3vxBRu4C.

Driver, Julia. "The History of Utilitarianism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2022. Metaphysics Research Lab, Stanford University, 2022.

https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history/.

FastAPI. "FastAPI," 2024. https://fastapi.tiangolo.com/.

Floridi, Luciano. "The Method of Levels of Abstraction." *Minds and Machines* 18, no. 3 (September 1, 2008): 303–29. https://doi.org/10.1007/s11023-008-9113-7.

Floridi, Luciano, and Mariarosaria Taddeo. "What Is Data Ethics?" *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, no. 2083 (December 28, 2016): 20160360. https://doi.org/10.1098/rsta.2016.0360.

Ge, Tao, Lei Cui, Baobao Chang, Zhifang Sui, Furu Wei, and Ming Zhou. "EventWiki: A Knowledge Base of Major Events," n.d.

Gelernter, Judith, and Shilpa Balaji. "An Algorithm for Local Geoparsing of Microtext." *GeoInformatica* 17, no. 4 (October 2013): 635–67. https://doi.org/10.1007/s10707-012-0173-8.

Gelernter, Judith, and Nikolai Mushegian. "Geo-parsing Messages from Microtext." *Transactions in GIS* 15, no. 6 (December 2011): 753–73. https://doi.org/10.1111/j.1467-9671.2011.01294.x.

GeoNames. "GeoNames," 2023. https://www.geonames.org/.

Halterman, Andrew. "Mordecai 3: A Neural Geoparser and Event Geocoder." arXiv, March 23, 2023. http://arxiv.org/abs/2303.13675.

Hellmann, Sebastian, Martin Brümmer, and Markus Ackermann. "NLP Interchange Format (NIF)," 2015. https://nif.readthedocs.io/en/latest/.

Hursthouse, Rosalind, and Glen Pettigrove. "Virtue Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2023. Metaphysics Research Lab, Stanford University, 2023. https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/.

Karimzadeh, Morteza, Scott Pezanowski, Alan M. MacEachren, and Jan O. Wallgrün. "GeoTxt: A Scalable Geoparsing System for Unstructured Text Geolocation." *Transactions in GIS* 23, no. 1 (February 2019): 118–36. https://doi.org/10.1111/tgis.12510.

Kemp, Simon. "Digital 2023: Niger," 2023. https://datareportal.com/reports/digital-2023-niger.

Konvention zum Schutze der Menschenrechte und Grundfreiheiten, Pub. L. No. BGBl 1958/210 (n.d.).

Lingua. "Lingua-Language-Detector 2.0.2," 2023. https://pypi.org/project/lingua-language-detector/.
Media Landscapes. "Niger," 2024. https://medialandscapes.org/country/niger.
Middleton, Stuart E., Giorgos Kordopatis-Zilos, Symeon Papadopoulos, and Yiannis Kompatsiaris. "Location Extraction from Social Media: Geoparsing, Location Disambiguation, and Geotagging." *ACM Transactions on Information Systems* 36, no. 4 (October 31, 2018): 1–27. https://doi.org/10.1145/3202662.
Niemeyer, Niemeyer, Tomi Pievilainen, Yaron de Leeuw, and Paul Ganssle. "Dateutil." Accessed August 23, 2024. https://github.com/dateutil/dateutil.
Ontotext. "About GraphDB," 2023. https://graphdb.ontotext.com/documentation/10.0/about-graphdb.html.
———. "GraphDB," 2023. https://graphdb.ontotext.com/.
OOSGA. "Social Media in Turkey - 2023 Stats & PLatform Trends," 2023. https://oosga.com/social-media/tur/.
OpenStreetMap. "About OpenStreetMap," 2024. https://www.openstreetmap.org/about.
Pöschl, Magdalena. "Neuvermessung der meinungsfreiheit." In *Tatsachenmitteilungen und werturteile: Freiheit und verantwortung*, edited by Helmut Koziol, 31–59. Jan Sramek Verlag, 2018.
Red Hat. "Was Ist Eine API?," 2023. https://www.redhat.com/de/topics/api/what-are-application-programming-interfaces.
Reporter ohne Grenzen. "World Press Freedom Index," 2024. https://rsf.org/en/index.
Scheichenbauer, and Seidl. "Die Verwaltungsrechtliche Einordnung von Internet-Recherchen Durch Verwaltungs-Behörden." In *Jahrbuch Datenschutzrecht 2022*. 56, n.d.
Scheler, M., and C. Bermes. *Der Formalismus in Der Ethik Und Die Materiale Wertethik: Neuer Versuch Der Grundlegung Eines Ethischen Personalismus*. Philosophische Bibliothek. Felix Meiner Verlag, 2014. https://books.google.at/books?id=OwuKDwAAQBAJ.
Scrapy. "Scrapy," 2024. https://scrapy.org/.
Semantic Web Company. "PoolParty Extractor." Accessed August 23, 2024. https://www.poolparty.biz/poolparty-extractor.
Sharma, Praval, Ashok Samal, Leen-Kiat Soh, and Deepti Joshi. "A Spatially-

Aware Algorithm for Location Extraction from Structured Documents." *GeoInformatica* 27, no. 4 (October 2023): 645–79. https://doi.org/10.1007/s10707-022-00482-1.

"Simple Event Modelling Ontology." Accessed August 23, 2024. https://semanticweb.cs.vu.nl/2009/11/sem/.

SmartBear Software. "Swagger," 2023. https://swagger.io/.

Spindler, and Schuster. *Recht Der Elektronischen Medien*. 34–65, n.d.

Statcounter. "Social Media Stats Niger," 2024. https://gs.statcounter.com/social-media-stats/all/niger/#monthly-201901-202305.

———. "Social Media Stats Turkey," 2024. https://gs.statcounter.com/social-media-stats/all/turkey.

Statistics, UNSECO Institute for; UNESCO Institute for Statistics. "Literacy Rate, Adult Total," 2024. https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=NE.

Strötgen, Jannik, and Michael Gertz. "A Baseline Temporal Tagger for All Languages." In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 541–47. Lisbon, Portugal: Association for Computational Linguistics, 2015. https://doi.org/10.18653/v1/D15-1063.

Tagesschau. "Zahl Der Toten Steigt Auf Mehr Als 50.000," 2023. https://www.tagesschau.de/ausland/asien/erdbeben-tote-tuerkei-syrien-101.html.

Tao, Liufeng, Zhong Xie, Dexin Xu, Kai Ma, Qinjun Qiu, Shengyong Pan, and Bo Huang. "Geographic Named Entity Recognition by Employing Natural Language Processing and an Improved BERT Model." *ISPRS International Journal of Geo-Information* 11, no. 12 (November 28, 2022): 598. https://doi.org/10.3390/ijgi11120598.

The Apache Software Foundation. "Apache Airflow," 2023. https://airflow.apache.org/.

The Linux Foundation. "OpenAPI," 2022. https://www.openapis.org/.

Wang, Jimin, Yingjie Hu, and Kenneth Joseph. "NeuroTPR: A Neuro-net Toponym Recognition Model for Extracting Locations from Social Media Messages." *Transactions in GIS* 24, no. 3 (June 2020): 719–35. https://doi.org/10.1111/tgis.12627.

Weasyprint. "WeasyPrint - The Awesome Document Factory.," 2023. https://weasyprint.org/.

"Wikidata." Accessed August 23, 2024. https://www.wikidata.org/wiki/Wikidata:Main_Page.

Abbildungsverzeichnis

Abbildung 1: Prozessablauf bei humanitären Missionen. 17
Abbildung 2: GeoCROW-Systemarchitektur mit allen Services, Inputs und Outputs, sowie der Orchestrierungskomponente Airflow und der zentralen Datenbank GraphDB. 38
Abbildung 3: Die GeoCROW-Ontologie, die auf der Simple Event Ontologie und NIF-Ontologie basiert und diese um verschiedene projektrelevante Konzepte erweitert. 50
Abbildung 4: Die GeoCROW-Taxonomie im Interface des PoolParty-Thesaurus-Managers. 56
Abbildung 5 : Roomscale-Umgebung. 59
Abbildung 6: Roomscale-Umgebung aus Benutzersicht. 60
Abbildung 7: Lifescale-Umgebung im Flugmodus. Roter Marker: Benutzerposition in Abbildung 5. 60
Abbildung 8: Lifescale-Umgebung aus Benutzersicht. 61
Abbildung 9: Ereignisvisualisierung im Roomscale-Modus aus Anwendersicht. 62
Abbildung 10: Ereignisvisualisierung im Lifescale-Modus im Flugmodus. 63
Abbildung 11: Ereignisvisualisierung im Lifescale-Modus aus der Ego-Perspektive. 63
Abbildung 12: POIs und Gebäude im Roomscale-Modus aus Anwendersicht. 64
Abbildung 13: Bedienelemente des VR-Controllers. 65
Abbildung 14: Platzierung des Markers für den Moduswechsel 66
Abbildung 15: Darstellung von Details eines Ereignisses 66

Autoren / Herausgeber und Autoren

- Johannes Scholz, Paris-Lodron-Universität-Salzburg: Fachbereich Geoinformatik
- Thomas Lampoltshammer, Universität für Weiterbildung Krems: Zentrum für E-Governance
- Heidi Scheichenbauer, Research Institute: Digital Human Rights Center
- Philipp Poindl, Research Institute: Digital Human Rights Center
- Franz Welscher, Paris-Lodron-Universität-Salzburg: Fachbereich Geoinformatik
- Diogo Sasdelli, Universität für Weiterbildung Krems: Zentrum für E-Governance
- Anna Breit, Semantic Web Company
- Artem Revenko, Semantic Web Company
- Tom Lautenbach, Technische Universität Wien
- Clemens Strauß, Institut für Militärisches Geowesen: IKT und Cybersicherheitszentrum, Direktion IKT&Cyber
- Johannes Göllner, Landesverteidigungsakademie/Abteilung: Zentraldokumentation, BMLV

Band 17 / 2024

Im GeoCROW-Projekt werden innovative Ansätze zur Krisenbewältigung vorgestellt. GeoCROW kombiniert semantische Technologien, Crowdsourcing und Virtual Reality, um humanitäre Einsätze zu verbessern. Das System analysiert Daten aus sozialen Medien und anderen Quellen, erstellt umfassende Lagebilder und visualisiert diese in einer VR-Umgebung. Fallstudien in Niger und der Türkei zeigen die Vielseitigkeit des Systems. Der Bericht beleuchtet technologische, ethische und rechtliche Rahmenbedingungen und betont, wie moderne Technologien die Effizienz und Sicherheit humanitärer Einsätze erhöhen können.

ISBN: 978-3-903548-02-2