

Informationsvisualisierung von Pressemitteilungen auf Basis von Open Source und Open Data – Am Beispiel von Pressemeldungen der Polizei Mainz

Pascal NEIS

Hochschule Mainz · pascal.neis@hs-mainz.de

Zusammenfassung

In den letzten Jahren werden zunehmend Verwaltungsdaten der Bundesregierung oder auch auf kommunaler Ebene in Deutschland als Open Data im Internet bereitgestellt. Der Inhalt, aber auch das Format, kann dabei leider recht unterschiedlich sein. Oft eignen sich freigegebene Daten direkt für eine Weiterverarbeitung oder Visualisierung, vereinzelt müssen jedoch Verfahren angewendet werden, um eine angedachte Analyse oder Darstellung umzusetzen. Die in diesem Beitrag vorgestellte Methodik verwendet Open Data und Open Source Software, um öffentlich zugängliche Pressemitteilungen auf einer Karte abbilden zu können. Insgesamt zeigen die Ergebnisse der untersuchten Pressemitteilungen des Polizeipräsidiums der Landeshauptstadt Mainz wie gut die vorgestellte Methodik funktioniert und welches Potential in den dadurch neu gewonnen Informationen liegt.

1 Einführung

In Deutschland werden seit dem Jahr 2015 Verwaltungsdaten und -dokumente von kommunalen, Länder- und Bundeseinrichtungen auf einem Datenportal zugänglich gemacht (GOVDATA 2014). Damit verfolgen die Betreiber des Datenportals vor allem folgende Ziele: Der Zugriff auf die Daten wird allen Interessierten aus Wirtschaft, Verwaltung und Bevölkerung erleichtert, die Daten liegen in maschinenlesbarer Form vor und können ohne größere Hindernisse weiterverarbeitet werden. Welche Informationen und Daten auf dem Portal bereitgestellt werden, ist oft den Verwaltungen selbst überlassen. Auch die Anzahl der zur Verfügung gestellten Datensätze kann je nach Bereitsteller (bzw. Organisation) sehr stark variieren. Im Datenportal für Deutschland (GOVDATA 2020) treten besonders Datenanbieter hervor, die Open Data und die damit verbundene Transparenz und Innovation unterstützen. Dazu zählen beispielsweise das Land Schleswig-Holstein, das Open Government Portal Rheinland-Pfalz, Open.NRW, der Freistaat Sachsen oder Berlin Open Data (vgl. GOVDATA 2020).

Neben einer Vielzahl von Datensätzen in den Kategorien Bevölkerung und Gesellschaft, Umwelt oder Verkehr, sind auch Daten unter Justiz, Rechtssystem und öffentliche Sicherheit im Portal abrufbar. Bei einem dieser Datensätze handelt es sich um die Polizeiliche Kriminalstatistik (PKS). Die PKS wird in Deutschland vom Bundeskriminalamt (BKA) herausgegeben und "enthält die der Polizei bekannt gewordenen rechtswidrigen Straftaten einschließlich der mit Strafe bedrohten Versuche, die Anzahl der ermittelten Tatverdächtigen und eine Reihe weiterer Angaben zu Fällen, Opfern oder Tatverdächtigen" (PKS 2020). In den Statistiken sind jeweils nur die Landesdaten von den 16 Landeskriminalämtern zu finden. Genauere Angaben über die geographische Position des Deliktes sind in der PKS nicht zu finden.

Auch andere Datensätze mit Informationen von Polizeimeldungen sind im Datenportal für Deutschland derzeit nicht enthalten. Dabei können detaillierte Informationen zu den Einsätzen der Polizei und deren Bereitstellung als Open Data, unter Einhaltung deren Prinzipien (u.a. keine personenbezogenen Daten oder Daten, die dem Datenschutz unterliegen), die Grundlage für die Verbrechenskartierung (engl. crime mapping) bilden. Nach HADAMITZKY (2015) dienen insbesondere diese Kartierungen als ein nützliches Mittel zur Vorbeugung und Bekämpfung von Kriminalität.

Um Kriminalität bzw. Verbrechen oder ggf. Vergehen besser zu verstehen und zu bekämpfen, spielt der Ort mit seiner geographischen Position eine entscheidende Rolle (CHAINÉY & RATCLIFFE 2013). Auf Basis dieser Verknüpfung mit einer räumlichen Information werden neben einer Visualisierung auf einer Karte und Ermittlung von Hotspots (ECK et al. 2005), vor allem weitere räumliche Analysen durchgeführt (SANTOS 2016). Die daraus entstehenden Erkenntnisse und Ergebnisse können wiederum zur Prävention (WEISBURD & MCEWEN 1998) und Vorhersage (ALMANIE et al. 2015) verwendet werden. Dabei wäre es beispielsweise auch möglich weitere Datensätze, wie demografische oder Daten von Mobilfunkanbietern, miteinander zu kombinieren und für weiterführende Untersuchungen zu analysieren (BOGOMOLOV et al. 2014).

Im Folgenden wird eine Methodik vorgestellt, wie freizugängliche Pressemitteilungen von Polizeidienststellen aus dem Internet mit Hilfe von Open Source und Open Data georeferenziert werden. Damit können Informationen einer Polizeimeldung, die vorher lediglich in Textform vorgelegen haben, nicht nur räumlich dargestellt werden, sondern ebenfalls die Grundlage für weitere Verknüpfungen und Analysen mit zusätzlichen Datenquellen bilden.

Der Beitrag gliedert sich wie folgt: Abschnitt 2 stellt die entwickelte Methodik vor. In Abschnitt 3 wird die Umsetzung vorgestellt und in Abschnitt 4 die Ergebnisse diskutiert. Ein Fazit & Ausblick schließt den Beitrag ab.

2 Methodik

2.1 Verfügbare Informationen

Pressemitteilungen und -meldungen der verschiedenen Blaulichtorganisation (wie z. B. Polizei, Feuerwehr oder Technisches Hilfswerk) werden in Deutschland über das presseportal.de¹ veröffentlicht. Das Portal beinhaltet verschiedene Auswahlmöglichkeiten wodurch z.B. nach Dienststelle, Bundesland oder den Zeitraum der Veröffentlichung gefiltert werden kann. Für die hier beschriebene Methodik wurden die Meldungen vom Polizeipräsidium Mainz (Pressestelle) für den Zeitraum 01.01.2019 bis 08.09.2019 verwendet. Die jeweiligen „Pressemeldungen der Polizei Rheinland-Pfalz sind unter Nennung der Quelle zur Veröffentlichung frei.“². Ein Zugriff auf die veröffentlichten Meldungen ist entweder über einen Webbrowser oder über eine Programmierschnittstelle (API) mit Schlüssel möglich. Im erwähnten Zeitraum wurden somit 991 Meldungen des Mainzer Polizeipräsidiums heruntergeladen und verwendet.

Jede veröffentlichte Meldung enthält die folgenden Informationen:

- Titel der Meldung

¹ <https://www.presseportal.de/blaulicht>

² Online abgerufen am 01.11.2020: <https://www.presseportal.de/blaulicht/pm/117708/4384726>

- Mitteilungszeitpunkt
- Textuelle Beschreibung der Meldung (Fließtext)
- Eine gesonderte Angabe von Stichworten zu den Orten, die die Meldung betreffen
- Eine Liste von Stichworten (Tags) zu Themen, bezogen auf den Inhalt der Meldung

Die textuelle Beschreibung der Mitteilung kann dabei im Fließtext weitere Informationen zur genaueren Angabe des Zeitpunktes, Beschreibung des Deliktes (z. B. Übertretung, Vergehen oder Verbrechen) oder zur geographischen Position in Form von einer Adressangabe (i. d. R. ohne Hausnummer) enthalten. In der folgenden Abbildung 1 sind die eben beschriebenen Details wie „Orte in dieser Meldung“ (1), „Themen in dieser Meldung“ (2) und die Adressangabe (3), auf die sich die Meldung bezieht, zu finden.

25.09.2019 – 10:06
 Polizeipräsidium Mainz
POL-PPMZ: Mainz - Einbruchversuch in Baucontainer



Mainz (ots)

Dienstag, 24.09.2019, 08:00 Uhr (Mitteilungszeitpunkt)

In der Nacht von Montag auf Dienstag versuche**3** bekannte Täter sich gewaltsam Zutritt zu mehreren Baucontainern zu verschaffen. Die (container befinden an einer Baustelle in de**3** **Neuen Mainzer Straße in Mainz-Hechtsheim**. Letztendlich gelingt es den Tätern nicht, die Container aufzubrechen. Ob etwas entwendet wurde ist derzeit unklar. Vor den Containern können im Rahmen der Anzeigenaufnahme eine Brechstange und ein Hammer festgestellt und sichergestellt werden. Die Ermittlungen dauern an.

Wer sachdienliche Hinweise zu dem genannten Fall geben kann, wird gebeten sich mit der Polizeiinspektion Mainz 3 unter der Rufnummer 06131/65-4310 in Verbindung zu setzen. Hinweise können auch per E-Mail unter pimainz3@polizei.rlp.de an die Polizei übermittelt werden.

Rückfragen bitte an:

Polizeipräsidium Mainz
 Pressestelle

Abb. 1: Beispiel einer Meldung vom Polizeipräsidium Mainz im Presseportal (Quelle³)

Im Beispiel einer Polizeimeldung aus dem Presseportal (Abbildung 1) ist bereits ersichtlich, dass die „Orte der Meldung“ (1) nicht zur genauen Verortung der Meldung auf einer Karte verwendet werden können. Bei den Orten handelt es sich lediglich um die Angabe des Bundeslandes, der Stadt bzw. des Ortsbezirkes. Nur im Fließtext (3) erfolgt eine exaktere Angabe in Form einer Adresse, auf welche geographische Position (oder Bereich) sich die Meldung bezieht.

2.2 Ermittlung der geographischen Position/Bezug der Mitteilung

Um die Adresse aus einer Pressemitteilung zu extrahieren und diese zu geokodieren (also in eine Positionsangabe umzuwandeln) existieren unterschiedliche Verfahren. Beispielsweise könnte auf Basis von Natural Language Processing und Parsing (MELO & MARTINS 2017)

³ Online abgerufen am 01.11.2020: <https://www.presseportal.de/blaulicht/pm/117708/4384726>

versucht werden die Adresse im Fließtext zu finden und diese anschließend mit einem beliebigen Geocoding-Service in ein Koordinatenpaar als Punkt oder Linien- bzw. Flächengeometrie, umzuwandeln.

Im Rahmen dieses Beitrags wurde jedoch ein anderer Ansatz entwickelt und im Folgenden untersucht. Auf Basis der freiverfügbaren und durch Freiwillige zusammengetragene räumliche Informationen des OpenStreetMap (OSM) Projektes⁴, wurde ein eigenes Adressverzeichnis bestehend aus Planungsbereichen, Ortsbezirke, Stadtbezirke und Straßennamen für das Untersuchungsgebiet der Landeshauptstadt Mainz erstellt. Das Verzeichnis wurde allerdings nicht auf Basis von verfügbaren Extrakten aus der Datensammlung des Projektes, sondern stattdessen über die Overpass API⁵ erstellt. Durch ein verfügbares Web-Frontend (Overpass Turbo⁶) konnten somit komfortabel die Daten des OSM Projektes über eine eigene Query Language abgefragt und zur Erstellung des beschriebenen Adressverzeichnisses genutzt werden. Bei der eigentlichen Ermittlung der geographischen Position bzw. des Bezugs einer Pressemitteilung wird im Fließtext nach einer Übereinstimmung eines Eintrags aus dem Adressverzeichnis gesucht. Sollten mehrere Adressangaben im Fließtext vorkommen, werden diese geographischen Positionen ebenfalls der Mitteilung zugeschrieben. Die folgende Tabelle 1 zeigt die Einzelschritte des Arbeitsablauf vom Abruf der Mitteilungen aus dem Presseportal (Zeile 2) bis zur Ermittlung des geographischen Bezugs einer Pressemitteilung (Zeile 5) und dem Speichern der Ergebnisse (Zeile 9).

Tabelle 1: Arbeitsablauf bei der Informationsvisualisierung (Pseudocode)

Zeile	Anweisung
1	program Informationsvisualisierung
2	ladeMitteilungenPresseportal()
3	generiereAdressverzeichnisOpenStreetMap()
4	foreach mitteilung
5	if fließtext contains any Adressverzeichniseintrag then
6	Adresseintrag (Position) zu Mitteilung speichern
7	end if
8	end foreach
9	Speicher Ergebnisse in *.geosjon & *.json Datei
10	erstelleTagSammlung()
11	Speicher TagSammlung als Bild & *.json Datei
12	end program

In der Tabelle 1 ist in Zeile 10 & 11 auch die Erstellung einer Sammlung und Ermittlung der Häufigkeit der Tags von den geladenen Mitteilungen zu sehen. Über die generierte Tag-Cloud können zusätzliche Informationen über die angegebenen Tags und deren Häufigkeit in den untersuchten Mitteilungen entnommen werden. Die folgende Abbildung 2 zeigt beispielhaft wie solch ein Ergebnis einer Tag-Cloud für die Mitteilungen einer Polizeidienststelle aussehen könnte.

⁴ Online abgerufen am 01.11.2020: <https://www.openstreetmap.org>

⁵ Online abgerufen am 01.11.2020: <http://overpass-api.de>

⁶ Online abgerufen am 01.11.2020: <https://overpass-turbo.eu>



Abb. 2:
Beispiel einer Tag-Cloud von den untersuchten Mitteilungen eines Polizeipräsidiums

3 Visualisierung der Pressemitteilungen

Zur Visualisierung der georeferenzierten Pressemitteilungen wurde ein Prototyp als Webclient implementiert. Als Grundlage diente dafür die freiverfügbare JavaScript-Bibliothek Leaflet⁷. Die interaktive WebGIS-Anwendung zeigt, wie in folgender Abbildung 3 zusehen, die visualisierten Pressemitteilungen im Untersuchungsgebiet der Landeshauptstadt Mainz.

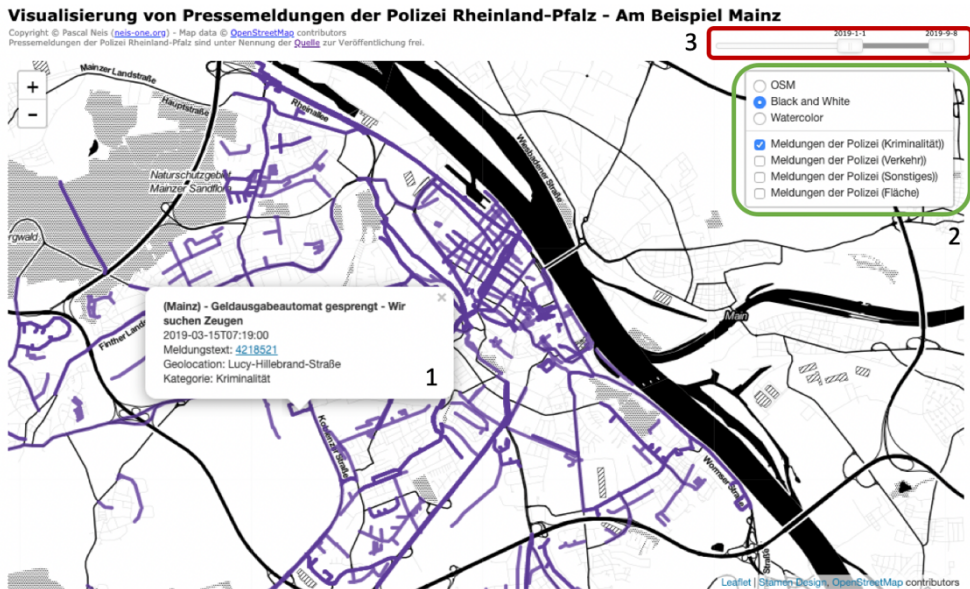


Abb. 3: Darstellung der Mitteilungen der Polizei im Webclient (Basiskarte und Daten von OpenStreetMap, Pressemitteilungen der Polizei Rheinland-Pfalz bzw. Mainz)

Neben der räumlichen Darstellung, auf welchen Bereich sich eine Mitteilung bezieht, können weitere Informationen (siehe Abbildung 3 (1)), wie die erwähnte Uhrzeit der Mitteilung oder Kategorie individuell als Popup angezeigt werden. Weiterhin bietet der Webclient die Möglichkeit die verschiedenen Kategorien bzw. angegebenen Themen einzeln als Layer ein- oder

⁷ Online abgerufen am 01.11.2020: <https://leafletjs.com>

auszublenden (siehe Abbildung 3 (2)). Die Uhrzeit der Mitteilung bietet zudem die Option durch einen Zeitschieberegler (siehe Abbildung 3 (3)) die Visualisierung auf einen bestimmten Zeitpunkt oder vorgegebenen Zeitraum einzuschränken.

4 Diskussion der Ergebnisse

Mit der vorgestellten Methodik konnten für den Zeitraum vom 01.01.2019 bis 08.09.2019 aus dem Presseportal 991 Meldungen extrahiert werden. Von diesen 991 Meldungen wurden ohne größere Schwierigkeiten 800 Meldungen (81%) georeferenziert. Für eine genauere Validierung der Ergebnisse wurden 120 zufällig gewählte Meldungen aus den georeferenzierten Meldungen ausgewählt und manuell geprüft. Diese Prüfung ergab, dass bei dem beschriebenen Verfahren 87% der Meldungen der korrekten geographischen Position oder Bereich zugewiesen werden konnten. Neben der Lage wurde auch die Kategorisierung über die angegebenen Themen in einer Meldung kontrolliert. Die manuelle Prüfung hierbei ergab, dass 89% der Meldung inhaltlich richtig zugeordnet wurden.

4.1 Einflussfaktoren bei der Ermittlung der geographischen Position

Bei der manuellen Prüfung der Ergebnisse wurden verschiedene Faktoren ermittelt, die einen direkten Einfluss auf die Ergebnisse haben können. Interessantweise sind in verschiedenen Pressemitteilungen Rechtschreibfehler in den Adressen enthalten, weswegen nicht der richtige Eintrag aus dem Adressverzeichnis zugewiesen werden konnte. Auch eine verkürzte Schreibweise der Adresse, wie z. B. „Str.“ anstatt „Straße“, kann ein Problem verursachen. Dies betrifft allerdings nicht nur die Pressemitteilung (Fließtext), sondern auch die verwendeten Geoinformationen des OSM Projektes, die zur Geokodierung verwendet wurden. Weitere Probleme gab es, wenn mehrere Adressen innerhalb einer Mitteilung im Fließtext erschienen sind. Daneben kam es sehr vereinzelt zu der Schwierigkeit, dass eine Meldung sich auf ein Delikt bezieht, welches nicht innerhalb der administrativen Grenze der Landeshauptstadt Mainz stattgefunden hat. Im Rahmen der hier untersuchten Pressemitteilungen gab es ebenfalls Mitteilungen, die gar keinen Raumbezug beinhalteten. Gerade die beiden zuletzt genannten Punkte beeinflussten die gewonnenen Ergebnisse der vorgestellten Methodik im Untersuchungsgebiet.

5 Fazit & Ausblick

Durch die Nutzung von Open Data und Open Source Software wurde sehr gut gezeigt, wie mit der vorgestellten Methodik öffentlich zugängliche Pressemitteilungen georeferenziert und auf einer Karte visualisiert werden können. Der gezeigte Prototyp ist allerdings nur eine erste Möglichkeit, die neu angereicherten Informationen zu nutzen. Weitere Visualisierungen in Form einer Heatmap oder zur Anzeige von Hotspots wären denkbar. Das vermutlich größte Potential liegt aber in der Verknüpfung und Analyse mit weiteren verfügbaren Informationen, wie bereits in der Einführung beschrieben wurde.

Derzeit wird der Mitteilungszeitpunkt für die Visualisierung der Pressemeldung verwendet, hier wäre es aber ebenfalls sinnvoll, die Angaben im Fließtext zu extrahieren. Für die Kate-

gorisierung wurde auf die Tags bzw. „Themen in dieser Meldung“ zurückgegriffen, allerdings könnten hier ebenfalls der Fließtext für weitere Informationen oder auch zur Verifikation von anderen Inhalten verwendet werden.

Literatur

- BOGOMOLOV, A., LEPRI, B., STAIANO, J., OLIVER, N., PIANESI, F., & PENTLAND, A. (2014): Once upon a crime: towards crime prediction from demographics and mobile data. In: Proceedings of the 16th international conference on multimodal interaction, 427-434.
- CHANEY, S., & RATCLIFFE, J. (2013): GIS and crime mapping. John Wiley & Sons.
- HADAMITZKY, G. (2015): Crime Mapping – Digitale Kriminalitätskartierung in Zeiten der Risikogesellschaft. Verlag für Polizeiwissenschaft, Frankfurt am Main.
- ECK, J., CHANEY, S., CAMERON, J., LEITNER, M. & WILSON, R. (2005): Mapping crime: understanding hot spots. Research report. National Institute of Justice, US.
- GOVDATA (2014): Hintergrundinformationen zu GovData. www.govdata.de/documents/10156/18448/GovData_Hintergrundinfos.pdf (30.07.2020)
- GOVDATA (2020): Daten. www.govdata.de/web/guest/daten (30.07.2020)
- MELO, F. & MARTINS, B. (2017): Automated geocoding of textual documents: A survey of current approaches. In: Transactions in GIS.
- PKS (2020): PKS – Bedeutung, Inhalt, Aussagekraft. www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/bedeutungInhaltAussagekraft.html (30.07.2020)
- RUSSNÁK, J., ONDREJKA, P., HERMAN, L., KUBÍČEK, P. & MERTEL, A. (2016): Visualization and spatial analysis of police open data as a part of community policing in the city of Pardubice (Czech Republic). In: Annals of GIS, Volume 22, 2016 - Issue 3, 187-201
- SANTOS, R. B. (2016). Crime analysis with crime mapping. Sage publications.
- WEISBURD, D., & MCEWEN, T. (1998). Crime mapping and crime prevention (No. 8). In: New York: Criminal Justice Press.