Richtlinie zum „Software-Sprint“

EveryoneCounts – Das Social Distancing Dashboard

Schlussbericht

Zuwendungsempfänger:

EveryoneCounts GbR

DLR PT- Berlin
Eing.am:
31. Aug. 2020
Eingangsnr.: 44/1816

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen **01IS20S25** gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Bisher fehlten verlässliche, vergleichbare und aktuelle Daten zur Umsetzung von Social Distancing bzw. dem Aktivitätsniveau im öffentlichen Raum. Diese Information ist für gesellschaftliche Akteure (Bürger:innen und Politiker:innen) nützlich, um informierte Entscheidungen treffen zu können. Auch Wissenschaftler:innen und Journalist:innen profitieren von einer soliden Datengrundlage für Forschung und Berichterstattung. Als Basis dafür dienten bisher teilweise schwer vergleichbare Datenquellen oder sogar nur anekdotische Evidenz.

Um dieses Problem zu lösen wurde im Zuge des WirVsVirus-Hackathons der Prototyp eines Online-Dashboards entwickelt, auf dem verschiedene öffentliche Datenquellen zur Aktivität im öffentlichen Raum zusammengeführt und auf Landkreisebene visualisiert wurden. Im Zuge der „Software Sprint“-Förderung sollte auf dem Prototypen aufgebaut und ein verbessertes Dashboard entwickelt werden.

Ursprünglich geplante Meilensteine gemäß Projektantrag waren:

- In der ersten Phase, soll die Lauffähigkeit des Produktes des Projekts weiter optimiert werden. Besonderer Fokus liegt dabei auf der Anbindung weiterer Datenquellen.
- In der zweiten Phase, wird die Analyse der Daten vorangetrieben. Dazu werden Korrelationen zwischen den einzelnen Datensätzen überprüft und grafisch aufgearbeitet. Gewonnene Erkenntnisse werden in einem Blogsegment auf der Website veröffentlicht.

Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint“

Für das Projekt wurden drei relevante Zielgruppen identifiziert:

1. Bürger:innen
2. Politiker:innen und Entscheidungsträger:innen
3. Wissenschaftler:innen und Journalist:innen

Die erste Gruppe der Bürger:innen möchte eine Übersicht bekommen wie groß die Aktivität in ihrer direkten Umgebung aber auch an anderen Orten (Urlaubsregionen oder Wohnorten von Bekannten) ist und wie sich diese entwickelt. Dies kann für zeitnahe Entscheidungen genutzt werden („Fahre ich an Ort X in den Urlaub oder sind da zu viele Menschen und damit das Risiko für mich persönlich zu hoch?“). Zudem können die Zahlen genutzt werden um subjektive Einschätzungen und Berichte zu überprüfen („Ein Freund meint, der Strand sei total überlaufen. Stimmt das?“). Dies ist die Hauptzielgruppe des Dashboards, auf dem die Aktivitätsdaten aufbereitet dargestellt und über eine einfache Metrik (7-Tage-Trend) verglichen werden können.

Die zweite Gruppe, die Politiker:innen, möchte politische und Verwaltungsmaßnahmen auf ihre Effektivität hin beurteilen. Aktivitätsmessungen können hier ein wertvoller Datensatz sein. Das Dashboard liefert einen schnellen Überblick. Daneben können über eine Software-Schnittstelle (API) auch die Rohdaten des Projekts abgefragt werden, welche für tiefer gehende Analysen weiterverarbeitet werden können.

Die Mitglieder der dritte Gruppe, die Wissenschaftler:innen und Journalist:innen, möchten in den Aktivitätsdaten Zusammenhänge entdecken und sie mit anderen Datenquellen verknüpfen (beispielsweise Covid-19-Infektionszahlen). Dazu können auch sie die API nutzen um direkt Zugriff auf die Rohdaten zu erhalten.

Das Projekt ist ganz im Sinne des „Software Sprint“-Initiative quelloffen und mit einem Fokus auf einen Nutzen für die Allgemeinheit ausgerichtet. Der Quelltext für alle Komponenten (Datensammlung und -verarbeitung sowie die Visualisierung auf dem Dashboard) ist auf Github öffentlich einsehbar (https://github.com/socialdistancingdashboard) und das Dashboard selbst ist öffentlich (https://everyonecounts.de). Die gesammelten Rohdaten werden – soweit eine Erlaubnis der ursprünglichen Datenquelle vorhanden ist – als Open Data behandelt: Auf sie kann unkompliziert über die Software-Schnittstelle zugegriffen werden. Zugang zur API wird auf Nachfrage bereit gestellt.

Ausführliche Darstellung der Ergebnisse

Der erste Meilenstein aus dem Projektantrag war die „Lauffähigkeit des Produktes des Projekts weiter [zu] optimier[en]“. Hierzu wurden folgende Maßnahmen durchgeführt:

- Einbindung weiterer Datenquellen
- Entwicklung eines neue Dashboards
- Verbesserung der Personenzählung auf Webcams
- Entwicklung einer Software-Schnittstelle (API)

Im Folgenden wird detailliert auf diese Maßnahmen eingegangen.

Einbindung weiterer Datenquellen

Der Prototyp des Dashboards aus dem WirVsVirus-Hackathon umfasste folgende Datenquellen:

- Passantenfrequenz in Innenstädten (bereitgestellt von der Firma Hystreet)
- Fahrradfrequenz in Innnenstädten (öffentlich, abgerufen über die Eco-Compteur-Infrastruktur)
- Passanten auf öffentlichen Plätzen (öffentliche Webcams, eigene Zählung über Bilderkennungsalgorithmen)
- Popularität von ÖPNV-Haltestellen (bereitgestellt von Google)
- Straßenverkehrsaufkommen („Traffic Index", öffentlich bereitgestellt von TomTom)
- Luftqualität („Air Quality Index", öffentlich bereitgestellt vom „World Air Quality Index"-Projekt)

Diese Datenquellen wurden automatisiert in regelmäßigen Abständen (stündlich oder täglich) abgefragt und die erhaltenen Daten zur Weiterverarbeitung in einem Amazon S3 Bucket gespeichert. Dieser Prozess wurde während des Förderzeitraums beibehalten und optimiert.

Aufgrund Inkompatibilität mit den Nutzungsbedingungen (Terms of Service) musste im Zuge der Weiterentwicklung des Projekts auf die Daten von Google verzichtet werden. Die kostenlose Bereitstellung von Straßenverkehrsdaten von TomTom wurde während des Förderzeitraums ersatzlos eingestellt, wodurch auf diesen Datensatz verzichtet werden musste.

Alle anderen bestehenden Datenquellen konnten in das neue Dashboard übernommen werden. Die Erkennungsrate bei der Zählung der Passanten auf Webcams wurde zudem verbessert (siehe unten).

Im Projektantrag wurden über Mobilfunkmasten gesammelte Bewegungsdaten als mögliche neue Datenquellen identifiziert. Solche Daten werden von den Firmen Teralytics und Motionlogic gesammelt und kommerziell vertrieben. Im Zuge der Covid-19-Situation spendeten beide Firmen dem RKI diese Daten. Nach erfolgreicher Kontaktaufnahme zu beiden Firmen konnte jedoch keine Vereinbarung zur Weitergabe oder nichtkommerziellen Nutzung der Daten im Projekt EveryoneCounts geschlossen werden.

Des Weiteren wurde der Mobilitäts-Daten-Marktplatz (MDM) des BMVI als interessante weitere Datenquelle identifiziert. In einem ersten Schritt wurden Daten der Verkehrsmessquerschnitte der Straßenbauverwaltung Baden-Württemberg (Landesstelle für Straßentechnik) in das Projekt integriert. Dies bedeutet, dass die Daten aus dem MDM ebenfalls in regelmäßigen Abständen automatisiert abgefragt werden und zur Weiterverarbeitung dem bestehenden Datensatz hinzugefügt werden. Dies ermöglicht eine Anzeige über das neue Dashboard (siehe unten) und einer Ausgabe über die Software-Schnittstelle (siehe unten). Bei den Verkehrsmessquerschnitten der Straßenbauverwaltung Baden-Württemberg handelt es sich um eine Zählung der PKW und LKW auf Autobahnen sowie deren Ein- und Ausfahrten. Diese Kennzahl, insbesondere die Anzahl der PKW, erlaubt Rückschlüsse auf die regionale Aktivität und Mobilität und passt deswegen zu den Projektzielen. Insgesamt beinhaltet dieser Datensatz 1300 Messstationen in Baden-Württemberg. Die entsprechenden Daten sind auch für Hessen und Nordrhein-Westfalen im MDM verfügbar und werden demnächst in das Projekt übernommen.

Der Quelltext für die automatisierte Abfrage der Daten aus dem Bestand des MDM ist öffentlich einsehbar unter https://github.com/socialdistancingdashboard/SDD-MDM.

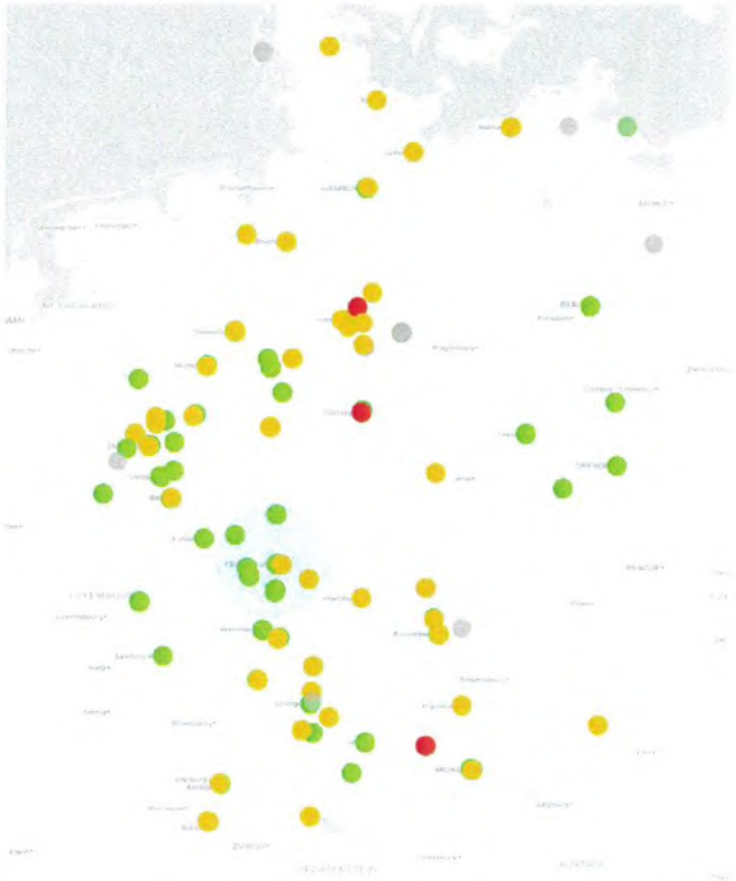
Entwicklung eines neue Dashboards

Der ursprüngliche Prototyp des Dashboards zeigte die Aktivitätsdaten auf Landkreisebene aggregiert und normiert auf ein „prä-Corona"-Aktivitäts-Niveau an. Dieser Ansatz stößt an seine Grenzen wenn wenige Datenpunkte den Score für große Landkreise bestimmen oder wenn keine historischen Daten vorliegen. Um die Informativität, Anschaulichkeit und somit den Mehrwert des Dashboards zu erhöhen wurde eine komplette Neuentwicklung durchgeführt.

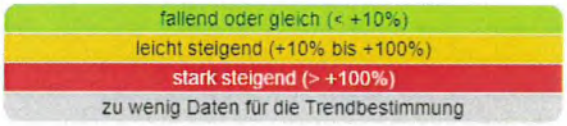
Die Designziele waren

- einzelne Messpunkte auf einer interaktiven Karte anzuzeigen,
- einen 7-Tage-Trend für jeden Messpunkt anzuzeigen,
- den Durchschnittstrend in einer bestimmten Region bestimmen, und
- weiterhin die Möglichkeit zu bieten, auf Landkreise oder Bundesländer zu aggregieren.

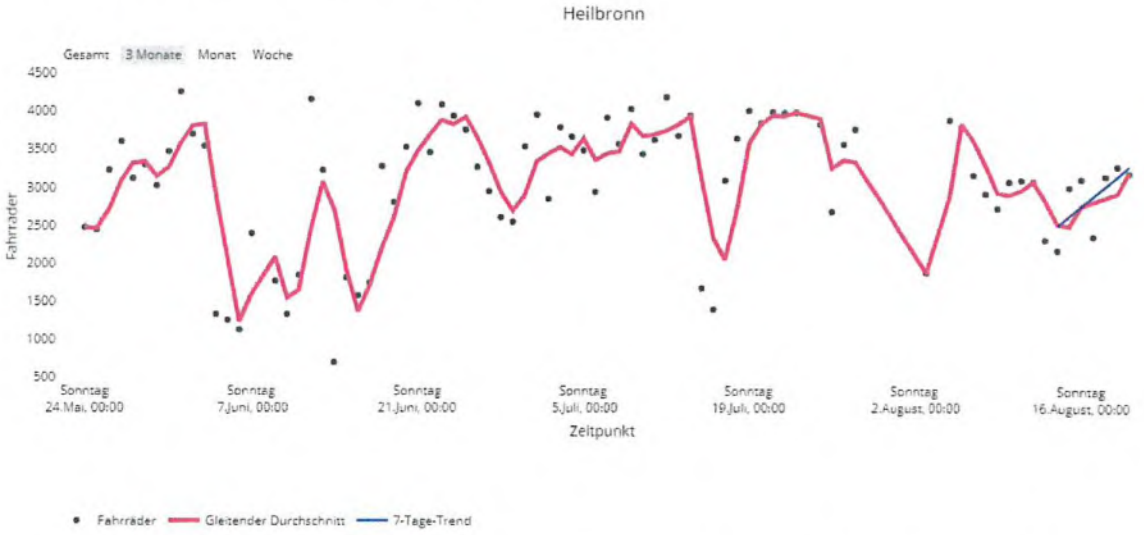
Das Live-Dashboard ist unter https://everyonecounts.de zu erreichen und kann mit jedem modernen Browser angezeigt werden. Die Hauptansicht des Dashboards ist eine Karte Deutschlands auf der die einzelnen Messquellen eingezeichnet sind:



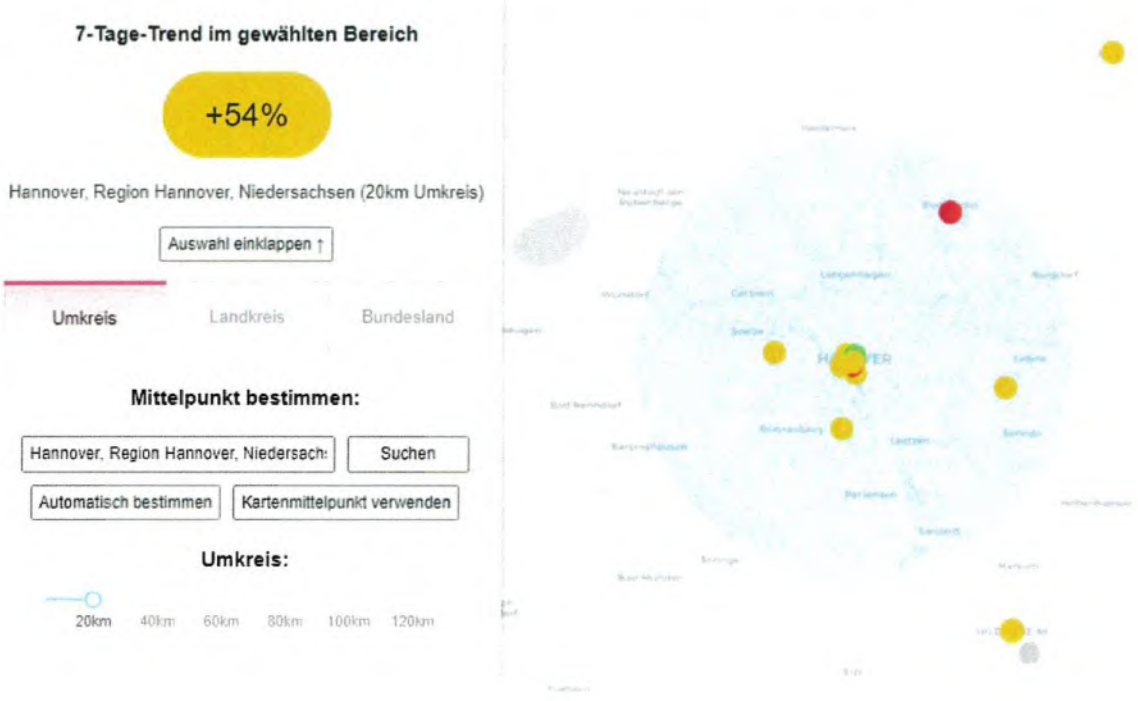
Diese sind entsprechend des Trends der letzten sieben Tage nach einem Ampel-System eingefärbt.



Die Karte ist interaktiv navigierbar (scrollen und zoomen). Durch Klick auf einzelne Datenpunkte kann der Verlauf der Messwerte angezeigt werden.



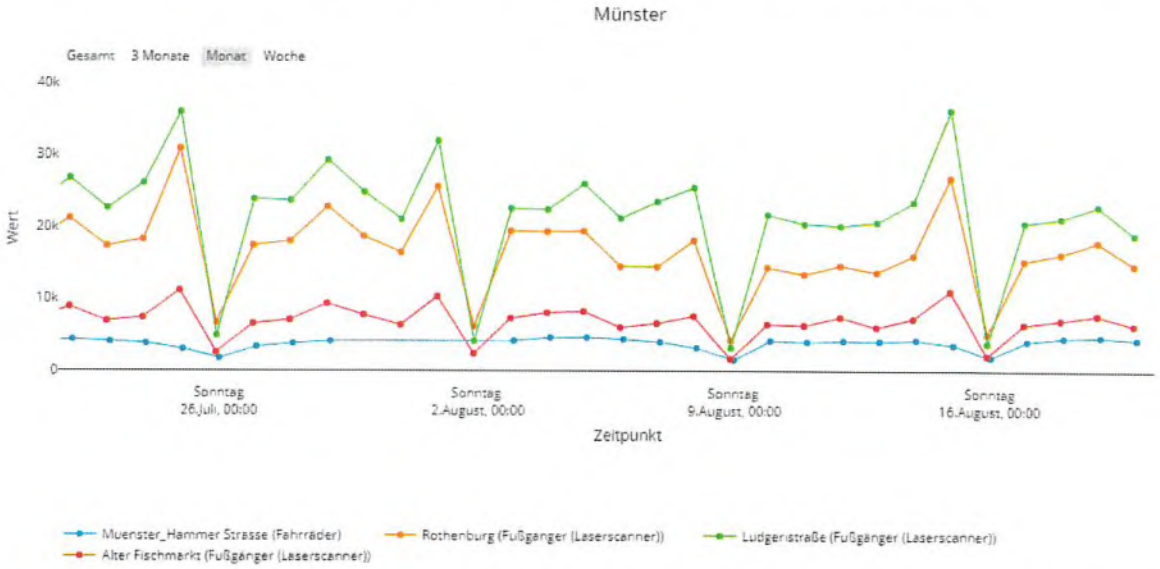
Um Bürger:innen eine einfache Kennzahl der Aktivität in einer Region zu geben können die Trend-Werte in einem bestimmten Gebiet gemittelt werden. Dies kann Beispielsweise ein bestimmter Landkreis sein, aber auch ein bestimmtes Gebiet um die Position des Nutzers (bestimmt durch den genutzten Webbrowser) oder eine frei wählbare Position herum. Das folgende Beispiel zeigt den durchschnittlichen 7-Tage-Trend innerhalb eines 20km-Radius um Hannover herum:



Alternativ zur Anzeige einzelner Punkte kann auch auf eine Landkreis- oder Bundesland-Ansicht gewechselt werden. Dabei werden alle 7-Tage-Trends aller Messpunkte innerhalb der Landkreise/Bundesländer gemittelt und wie oben farbcodiert angezeigt.



Bei Klick auf einen Landkreis oder ein Bundesland werden die Graphen aller Messpunkte in einer interaktiven Grafik zusammengefasst, wie am Beispiel des Landkreises Münster illustriert:



Das neue Dashboard wurde zum Ende der Förderperiode (Ende Juli 2020) veröffentlicht. Es basiert technologisch auf dem Python-Framework „Dash" und läuft auf einer Amazon EC2 Instanz. Der Quelltext ist offen einsehbar unter https://github.com/socialdistancingdashboard/frontendv2.

Verbesserung der Personenzählung auf Webcams

Eine Datenquelle des Projekts sind öffentliche Webcams, auf denen mittels Methoden der maschinellen Objekterkennung Personen gezählt werden. Ziel ist auch hier Passanten-Ströme im öffentlichen Raum zu erfassen. In diesem Fall wurde ein mathematisches Modell (ein so genanntes „neuronales Netzwerk") verwendet, welches Objekte auf Bildern erkennen kann. Die Erkennungsrate hängt dabei stark von der Größe des zu erkennenden Objektes im Bild und der Auflösung ab.

Die Betreiber der öffentlich zugänglichen Webcams richten sich nach den gesetzlichen Vorgaben, welche besagen, dass nicht die Möglichkeit bestehen darf, einen Menschen als bestimmte Person zu identifizieren. Dies wird von den Betreibern meist durch eine erhöhte Installationshöhe der Kamera

und eine geringe Auflösung gewährleistet. Infolgedessen sind Personen und Objekte nur wenige Pixel groß, wodurch die Erkennungsrate mit universellen mathematischen Modellen sinkt. Das anfangs verwendete, auf den COCO-Datensatz (https://cocodataset.org) vor-trainierte „Faster R-CNN" Model hatte für den vorliegenden Datensatz an Webcam-Aufnahmen eine sehr niedrige Erkennungsrate. Diese Beobachtung wurde qualitativ überprüft.

Um die Erkennungsrate von Personen auf Webcams zu verbessern wurde ein eigenes mathematisches Modell auf den konkreten Anwendungsfall hin optimiert. Das neue Modell basiert auf einem vor-trainierten Modell, welches mit selbst annotierten Webcam-Bildern nach-trainiert wurde. Für die Erstellung und das Nachtraining des neuen Modells wurde Microsoft Azure Custom Vision genutzt. Durch das optimierte Modell wurde die Erkennungsrate von Personen auf Webcams deutlich verbessert.

Entwicklung einer Software-Schnittstelle (API)

Öffentlich relevante Daten müssen anderen Entwicklern verfügbar gemacht werden um die gesellschaftliche Entwicklung in Deutschland voranzubringen. Die Entwicklung einer standardisierten und einfach zu bedienenden Software-Schnittstelle (API) ist deshalb ein zentraler Bestandteil des Projektes. Beispiele für bisherige Projekte für öffentlich zugängliche Daten sind auf https://www.govdata.de zu finden. Dort sind öffentlich verfügbare Datensätze verlinkt. Ziel der Open-Data-Initiative ist es, durch die Bereitstellung von strukturierten Inhalten mehr Transparenz und Zusammenarbeit zu ermöglichen. Diese Daten können jedoch nur umständlich in andere Applikationen eingebunden werden. Die API von EveryoneCounts orientiert sich an diesem Ansatz, bringt allerdings alle Daten auf ein einheitliches Format. Dies garantiert eine bessere Nutzbarkeit und reduziert den anfänglichen Arbeitsaufwand für Nutzer pro Datenquelle.

Die API wurde auf der Basis einer REST-API umgesetzt, was dem aktuellen Stand der Technik entspricht. REST-APIs verwenden eine einheitliche Schnittstelle und geben die Daten einheitlich zurück. Nutzer der API können alle Datenquellen mit dem selben Schema abrufen. Das Schema wird dem Nutzer mit Hilfe der API Dokumentation erklärt.

Zur Authentifizierung benötigt jeder Nutzer einen API Schlüssel, welcher jeweils einem Nutzungsplan zugeordnet wird. Durch diese Regelung sind die Daten frei verfügbar, jedoch sind die freien Aufrufe der Nutzer pro Monat auf ein sinnvolles Maß limitiert um Ressourcenverschwendung vorzubeugen. Eine denkbare Möglichkeit, das Projekt langfristig kostendeckend zu betreiben, läge darin, dass Nutzer dieses Limit bei Bedarf gegen eine geringe Gebühr erhöhen lassen können. Dieser Ansatz wird aktuell aber noch nicht verfolgt.

Die API ist unter https://blog.everyonecounts.de/api-dokumentation/ dokumentiert.

Analyse der Daten im eigenen Blog

Als zweiter Meilenstein des Projektantrags war die Veröffentlichung von Analysen in einem eigenen Blog geplant. Der Blog wurde im Förderzeitraum eingerichtet und ist unter https://blog.everyonecounts.de erreichbar. Neben allgemeinen Information über das Projekt und die Datenquellen können hier Updates und Analysen veröffentlicht werden. Ein Visualisierung des Datenbestands wurde am 18.05. veröffentlicht (https://blog.everyonecounts.de/2020/05/18/wo-sind-unsere-messpunkte). Eine tiefer gehende Analyse und Aufarbeitung der Daten konnte im Förderzeitraum leider nicht mehr abgeschlossen werden.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Die Weiterentwicklung des Dashboards ermöglicht es Bürger:innen und Politiker:innen einen schnelleren und genaueren Überblick über aktuelle und lokale Mobilitäts- und Aktivitäts-Entwicklungen zu erhalten. Durch die Optimierung der Webcam-basierten Personenzählung ist auch der Mehrwert dieser Informationen gestiegen. Neben Einblicken in die Aktivität in Innenstädten ermöglicht die Einbindung neuer Daten aus dem MDM zusätzliche Erkenntnisse über Veränderungen im Mobilitätsverhalten auf Bundesautobahnen.

Für Entscheidungsträger:innen und Wissenschaftler:innen steht nun mit der API eine komfortable Möglichkeit zur Verfügung die genutzten Daten in eigene Analysen miteinzubeziehen. Zu den bisherigen Nutzern zählt das Stadtmarketing der Stadt Biberach sowie Forscher des Instituts für Infektionsprävention und Krankenhaushygiene am Universitätsklinikum Freiburg.

Offene, aktuelle und frei verfügbare Daten aus dem öffentlichen Raum sind aktuell nicht einfach für die Öffentlichkeit zugänglich. Dies ist ein generelles Problem in Deutschland. Projekte wie der MDM sind ein guter Anfang um Daten aus der Verwaltung zu veröffentlichen, aber immer noch mit einer signifikanten organisatorischen und technischen Einstiegshürde versehen. Zudem sind diese Daten äußerst lückenhaft, so teilen beispielsweise nur drei Bundesländer Autobahn-Verkehrsmessschnitte über den MDM. Das Projekt EveryoneCounts verfolgt den Ansatz die Einstiegshürde möglichst gering zu halten und allen Menschen im Sinne der digitalen Teilhabe Zugang zu öffentlichen Daten zu ermöglichen.

Der quelloffene Ansatz des Projekts ermöglicht Transparenz bezüglich der Datenverarbeitung und kann so zur unabhängigen Validierung genutzt werden. Durch den modularen Aufbau des Quellcodes können einzelne Komponenten von anderen Projekten genutzt und darauf aufgebaut werden. So existiert beispielsweise Interesse aus der Community an der Nutzung des optimierten Modells für Webcam-basierte Personenzählung. Da der Code und dessen Versionsgeschichte quelloffen auf Github liegen, ist eine solche Nutzung und Weiterentwicklung problemlos und dauerhaft (auch weit über die Lebensdauer des Projekts hinaus) möglich.

Eine Fortführung des Projekts wird angestrebt. Die Datenquellen werden um weitere Datensätze aus dem MDM ergänzt. Daneben wird ein einbettbarer Software-Baustein (Widget) entwickelt, mit dem Webseitenbetreiber Daten und Graphen aus dem Projekt komfortabel in eigene Seiten einbauen können. Erste Anfragen hierfür existieren bereits von Betreibern der vom Projekt genutzten öffentlichen Webcams.

Um das Projekt weiterzuentwickeln soll verstärkt der Kontakt zur Nutzergemeinschaft gesucht werden, beispielsweise über ein Feedback-Formular und soziale Medien. Nur so kann sichergestellt werden, dass sich das Projekt auch tatsächlich entlang aktueller Bedarfe weiterentwickelt.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Aufgrund Inkompatibilität mit den Nutzungsbedingungen (Terms of Service) musste im Zuge der Weiterentwicklung des Projekts auf die Daten von Google verzichtet werden. Die kostenlose Bereitstellung von Straßenverkehrsdaten von TomTom wurde während des Förderzeitraums ersatzlos eingestellt, wodurch auf diesen Datensatz verzichtet werden musste. Die Verkehrsdaten konnten aber teilweise durch die neu dazugekommen Datenquelle des MDM mit Daten der Straßenbauverwaltung Baden-Württemberg substituiert werden.

Die über den MDM von der Straßenbauverwaltung Baden-Württemberg bereitgestellten Daten sind seit Mitte August 2020 fehlerhaft und können dadurch vorübergehend nicht für das Projekt genutzt werden. Das Problem liegt bei der Straßenbauverwaltung Baden-Württemberg, welche über den Sachverhalt informiert wurde. Sobald diese den Fehler behoben hat und aktuelle Daten über den MDM ausliefert, werden diese auch wieder heruntergeladen und auf dem Dashboard angezeigt.

Im Projektantrag wurden über Mobilfunkmasten gesammelte Bewegungsdaten als mögliche neue Datenquellen identifiziert. Solche Daten werden von den Firmen Teralytics AG und Motionlogic GmbH gesammelt und kommerziell vertrieben. Im Zuge der Covid-19-Situation spendeten beide Firmen dem RKI diese Daten. Nach Kontaktaufnahme mit Prof. Dirk Brockmann vom RKI wurde dem Projekt eine Nutzung dieser Daten zugesichert unter der Voraussetzung einer Zustimmung durch Motionlogic bzw. Teralytics. Nach Einholen der Zustimmung der Motionlogic GmbH gab es aber Bedenken von Seiten des RKI, da die Daten nur mit Universitäten und Forschungsinstituten geteilt werden dürften und eine Datenweitergabe wurde abgelehnt. Für die Teralytics AG kam eine nicht-kommerzielle Kooperation mit dem Projekt nicht in Frage. Zugang zu den Mobilitätsdaten käuflich zu erwerben war für das Projekt jedoch wirtschaftlich nicht möglich. Es konnten somit keine Mobilfunk-basierten Mobilitätsdaten für das Projekt nutzbar gemacht werden.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Das Dashboard befindet sich unter https://everyonecounts.de.

Der Quelltext und die Versionsgeschichte aller Komponenten des Projekts ist auf Github öffentlich einsehbar: https://github.com/socialdistancingdashboard.

Veröffentlichungen des Projekts finden sich auf dem Blog https://blog.everyonecounts.de. Daneben gibt es dort Informationen über das Projekt, das Team, die Daten sowie eine Sammlung an Presseartikeln mit Bezug zum Projekt.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Das Ausscheiden eines Projektmitarbeiters machte eine Anpassung der Arbeitsplanung während der Förderperiode notwendig. Die Arbeitslast konnte gleichmäßig auf die anderen Teammitglieder verteilt werden.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Für Nutzung im Projekt werden Daten (Passantenfrequenz) von der hystreet.com GmbH bereitgestellt. Dem Projekt wurden Nutzungsrechte für das Dashboard gewährt, allerdings wurde einer Weitergabe der Rohdaten über die API von Seiten der hystreet.com GmbH nicht zugestimmt. Aus diesem Grund wird dieser Teil der Rohdaten auch nicht über die API ausgeliefert.