Richtlinie zum „Software-Sprint“

EmissionsAPI – Emissions Daten API basierend auf Satellitendaten

Schlussbericht

Zuwendungsempfänger: Brockmeyer Haardiek Kerssen Kerssen Kiesow Rosemann GbR

Timo Nogueira Brockmeyer

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS19S07 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Am Anfang unserer Idee stand die Erkenntnis, das politische Instrumente zur Umsetzung von Klimazielen in ihrer Effektivität nur für einen sehr kleinen Personenkreis überprüfbar sind. Dementsprechend bleibt die öffentliche Diskussion für viele Menschen abstrakt und schwer greifbar. Um einem breiteren Publikum die Möglichkeit zu geben, sich eine durch Fakten gestützte Meinung zu bilden, sehen wir es als notwendig an, einen einfachen Zugang zu den Datenquellen zu schaffen. Dieser Zugang kann konkret durch interaktive Visualisierungen oder einfach abrufbare Statistiken geschaffen werden. Im Laufe der Recherchen zu unserem Projekt haben wir dabei schnell erkannt, dass die Daten die wir von der ESA nutzen wollten, zwar öffentlich zugänglich sind, aber nicht auf einfache Art durch standard Webtechnologien zu verarbeiten sind. Daher haben wir uns entschieden diese fehlende digitale Infrastuktur zu erschaffen, die die wissenschaftlichen Daten aufbereitet und durch ein einfaches Application Programming Interface (API) in einer automatisierbaren, dynamischen Weise ausliefern kann. Diese Schnittstelle kann dann wiederum von einem verbreiterten Personenkreis aus Web-Entwicklern und Data Scientists genutzt werden, um eigene Projekte umzusetzen. Beispielhaft könnte man die Effektivität von Umweltvorschriften - wie Dieselfahrverbote in Innenstädten oder neue Schwefelgrenzwerte in der Schiffahrt - durch einen Vorher/Nacher-Vergleich auf einer Landkarte bewerten um dadurch Argumente mit Daten zu überprüfen.

In der anfänglichen Planung haben wir die im Folgenden aufgeführten Meilensteine definiert. In der Arbeitsweise haben wir dabei ein vertikales Prototyping angestrebt, so dass immer an allen Meilensteinen gleichzeit gearbeitet wird.

Meilenstein 1: Datenaufbereitung

- Entwicklung der Algorithmen zur automatischen Aufbereitung der Daten
- Entwicklung der Rechenmodelle zur effizienten Datenreduzierung

Meilenstein 2: Datenmodell

- Einbindung der Sentinel-API
- Caching der Daten
- Datenstrukturen innerhalb des Dienstes

Meilenstein 3: Infrastruktur

- Berechnung und Datenfluss
- automatisches und nachhaltiges Deployment

Meilenstein 4: API

- Definition der API Endpunkte
- Anbindung an Module des Backends
- Beispiel-App einer Visualisierung in OpenStreetMap

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

Die primäre Zielgruppe für unser Projekt sind Entwickler und Data Scientists, die Visualisierungen oder Analysen basierend auf den Emissionsdaten des Sentinel-5P Programmes entwickeln möchten. Die Eintsiegshürde bei diesen Daten ist jedoch aus zwei Gründen sehr hoch: erstens, benötigen die wissenschaftlichen Daten eine ausführliche Auseinandersetzung mit der Dokumentation durch die ESA um sie korrekt verarbeiten zu können und zweitens, ist eine aufwendige Infrastuktur notwendig um sie dynamisch darzustellen. Es fehlt also eine technische Lösung, die diese Daten auf eine an die in unserer Zielgruppe vorherrschenden Technologien angepasste Art und Weise ausliefert. Unser Projekt schafft diese Lösung und ermöglicht es unserer Zielgruppe schneller und ohne tiefgehendes Fachwissen zu arbeiten. Zusätzlich hoffen wir auf einen Multiplikatoreffekt, indem die Daten einer breiteren Gruppe als bisher zugänglich werden. In diesem Sinne fällt unser Projekt klar unter den Public Interest Tech Begriff, da es eine Lücke in der öffentlichen, digitalen Infrastruktur füllt.

Ausführliche Darstellung der Ergebnisse

Am Ende unserer Förderung steht ein prototypischer Dienst, der ein Datenprodukt (Kohlenmonoxid) der Sentinel-5P Mission in einem einfach in andere Dienste integrierbarem Format (JSON) ausliefert: **https://api.emissions-api.org/ui/**. In diesem Sinne konnten auch alle Meilensteine erfolgreich abgeschlossen werden. Eine Einschränkung von den ursprünglichen Zielen ist lediglich, dass wir uns auf ein Datenprodukt der Sentinel-5P Mission beschränken, statt schon während des Förderzeitraumes alle Datenprodukte anzubieten. Dieser Umstand soll allerdings zeitnah behoben werden (siehe Punkt Weiterentwicklungen). In den folgenden Absätzen werden wir die Ergebnisse nach Themengebieten geordnet darstellen.

Technische Entwicklung

Am Anfang der Projektlaufzeit haben wir einen minimalen Prototypen entwickelt, der bereits alle relevanten technischen Eigenschaften des späteren Dienstes abgebildet hat (vertikales Prototyping). Dieser Prototyp wurde dann als Ausgangsbasis genommen um im weiteren Projektverlauf die verschiedenen technischen Aspekte im Detail zu testen und weiterentwickeln zu können. Als erstes stand dabei das Beschaffen der Daten von der ESA im Mittelpunkt. Dazu musste ein Programm geschaffen werden, dass automatisiert die Daten bei der ESA herunteladen kann. Dazu wurde das Pythonpaket sentinel5dl entwickelt, das auch als eigenständige Software benutzt werden kann. Es wurde über die Plattform PyPI zur einfachen Installation veröffentlich. Diese Software findet bereits auch außerhalb unseres Projektes Anwendung durch Dritte.

Der nächste Fokus war das Design der für den Nutzer relevanten API-Endpunkte. Dabei hat uns das begleitende Coaching durch die Open Knowledge Foundation geholfen unsere Nutzer besser zu verstehen. Der dritte Schritt war die Entwicklung eines Paketes zur Datenvorverarbeitung und -filterung (s5a), das auch über die Plattform PyPI der Öffentlichkeit zur Verfügung gestellt wurde.

Der größte Teil der technischen Entwicklung wurde durch das Testen und die Optimierung der Datenhaltung in Anspruch genommen. Die großen Datenmengen die durch das Zwischenspeichern und Konvertieren des binären Datenformats der Orignaldaten, in ein besser abzufragendes Datenbankformat entstehen, waren wie erwartet eine große Herausforderung. Bei der Bearbeitung dieser Problemstellung kam uns der transparente Umgang mit Wissen in der Open-Source Gemeinschaft zu Gute. Dadurch war es uns möglich, mit Datenbankentwicklern der von uns genutzten Softwareprodukte Kontakt aufzunehmen und mit ihnen über unsere technischen Fragen zu diskutieren. Das hat uns geholfen einige wichtige Optimierungen an den Schnittstellen zu unserer Datenbanklösung vorzunehmen. Letztenendes mussten wir uns dennoch entscheiden, dass die Gesamtdatenmenge aller ESA-Datenprodukte, für uns nicht mit realistisch verfügbaren Hardwarekapazitäten handhabbar ist. Daher haben wir uns schlussendlich für die Projektlaufzeit auf ein einzelnes Datenprodukt (Kohlenmonoxid) konzentriert (siehe auch Ausführungen weiter unten).

Der letzte Punkt der technischen Entwicklung betraf das Aufsetzen des Dienstes auf den Servern. Wir haben dazu zwei automatisierte Set-Ups Entwickelt, eines für Rechenzentren und eines für den Amazon Web Service. Als Host konnten wir sowohl Amazon als auch zwei deutsche Universitäten gewinnen. Dadurch konnten wir die Verfügbarkeit unseres Dienstes eine Laufzeit von mindestens einem Jahr nach Förderende garantieren.

Öffentlichkeitsarbeit

Da ein Open-Source Projekt nur von einer lebendigen Gemeinschaft getragen werden kann, haben wir während des gesamten Förderzeitraums Wert auf eine aktive Außendarstellung gelegt. Gleich zu Beginn des Projektes haben wir daher eine Homepage und einen Twitter-Account angelegt, um mögliche Nutzer für unser Projekt zu interessieren. Im Oktober haben wir erfolgreich die virtuelle Veranstaltung Hacktoberfest auf GitHub genutzt, um durch gezielte Veröffentlichung unserer technischen Fragestellungen externe Entwickler zur Mitarbeit zu animieren. Im November haben wir das Projekt in dem Video-Podcast FLOSS Weekly einem breiten internationalen Publikum vorgestellt. Im Dezember folgte dann ein Lightning Talk auf dem 36C3 in Leipzig und dann im Februar ein Lightning Talk auf der FOSDEM in Brüssel.

Neben diesen Events, haben wir uns bemüht die Hürden für Externe die an unserem Projekt teilnehmen oder die Ergebnisse verwerten möchten, so niedrig wei möglich zu halten. Daher haben wir neben der technischen Entwicklungsarbeit einen Schwerpunkt unserer Arbeit auf eine gute Dokumentation mit Contribution Guidelines und auf das Einhalten von Open-Source Best Practices gelegt. Bei der Verbesserung unserer Außendarstellung war uns auch das durch die Open Knowledge Foundation zur Verfügung gestellte Coaching eine Hilfe.

Data Literacy

Ein weiteres Ergebnis des Projektes, ist die ausführliche Begleitung des Nutzers durch Beispiele und Texte die die Data Literacy im Umgang mit den Sentinel-5P Daten erhöhen sollen. Wir haben im Projektverlauf selber viel über die Daten gelernt. Dieses Wissen stellen wir über Blogartikel und begleitende Texte auf unserer Homepage zur Verfügung, um die Nutzer über die Vorteile aber auch die Limitierungen von satellitengestützten Emissionsmessungen aufzuklären. Wir möchtetn dadurch unsere Nutzer befähigen selber gute Anwendungsfälle für die Daten zu erkennen.

Abschließend möchten wir bemerken, dass es uns ohne die Hilfestellung der Open Knowledge Foundation schwer gefallen wäre in sechs Monaten den gleichen Impact mit unserem Open-Source Projekt zu erreichen. Die Betreuung, Hilfestellung, Unterstützung in der Öffentlichkeitsarbeit und die Vernetzung mit anderen Public Interest Tech Projekten hat uns motiviert, uns eine äußere Struktur gegeben und uns kreativ bereichert.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Die Zielgruppe und der Nutzen der sich für sie aus unserem Projekt ergibt, besteht wie oben erwähnt in dem vereinfachten Zugriffs auf satellitengestützte Emissionsdaten. Neben neu entwickelten Infrastruktur- und Softwarelösungen, gehören zu dieser Vereinfachung auch die Dokumentation und die Begleittexte zum Umgang mit den Daten. Dadurch dass unsere Ergebnisse alle frei lizenziert sind, sind der Weiterverwendung durch Dritte kaum Grenzen gesetzt. Wir erhoffen uns dadurch einen Multiplikatoreffekt der andere Entwickler befähigt, so wie wir auch bei unseren Entwicklungen von der Vorarbeit anderer Entwickler profitiert haben. Für uns persönlich hat dieses Projekt die Möglichkeit bedeutet, den kompletten Prozess, mit allen Aspekten der Organisation und Durchführung eines Open-Source Projekts, einmal konzentriert zu durchlaufen. Das und die intensive, bewusste Auseinandersetzung mit der Schnittstelle von Technik und Gesellschaft, hat unsere Fähigkeiten und unser Verantwortungsbewusstsein geschärft, um in der Zukunft noch aktiver unsere digitale Gesellschaft mitgestalten zu können.

Für die Zukunft der emissions API planen wir eine Reduzierung der Rohdaten auf unserer Serverseite, so dass alle Datenprodukte der Sentinel-5P Mission berücksichtigt werden können. Details haben wir in einem Text auf unserer Homepage veröffentlicht (https://emissions-api.org/roadmap.html).

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Wie bereits erwähnt, haben wir im Laufe des Projekts beschließen müssen nicht alle Datenprodukte als Rohdaten anzubieten. Sobald wir im Projektverlauf absehen konnten, dass dieses Problem auftreten kann, haben wir einen zweigleisigen Ansatz verfolgt. Wir haben zum Einen versucht mit externen Experten auszutesten, wo wirklich unsere technischen Grenzen liegen. Zum Anderen haben wir die Idee erkundet Daten zu reduzieren. Dieses Vorgehen hat es uns zum Ende des Projekts ermöglicht unsere ursprüngliche Idee als Prototyp zu präsentieren, gleichzeitig aber schon einen konkreten Plan mit technischen Lösungsansätzen für die Zukunft zu haben.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Die erste Anlaufstelle für mögliche Nutzer ist unsere Homepage **https://emissions-api.org/**. Dort finden sich neben Links zu allen weiterführenden Ressourcen um das Projekt auch erklärende Texte zu den Datensätzen, Anwendungsbeispiele und Formulierungen von Motivation und Zukunftszielen. Ebenso ist dort der Link zu der technischen Dokumentation der API zu finden (https://api.emissions-api.org/ui/).

Unsere GitHub-Präsenz (https://github.com/emissions-api) stellt den Quellcode für alle Programmteile öffentlich unter der MIT-Lizenz zur Verfügung. Außerdem finden sich hier die technische Dokumentation (https://docs.emissions-api.org/ und https://sentinel5dl.emissions-api.org/), öffentlich dokumentierte Dikussionen zur Entscheidungsfindung im Projektverlauf und Hinweise wie man selber in unserem Projekt aktiv werden kann. Nutzer die an Neuigkeiten zu unserem Projekt interessiert sind, können unserem Twitter-Account @emissions_api folgen und Aufzeichnungen von öffentlichen Vorträgen über unser Projekt sind unter folgenden Links zu finden:

- https://media.ccc.de/v/36c3-10525-lightning_talks_day_3#t=1146
- https://fosdem.org/2020/schedule/event/emissions_api/
- https://twit.tv/shows/floss-weekly/episodes/555

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Zum Ende des Projektes standen wir vor der Frage ob wir mehr Zeit in technische Perfektionierung investieren oder in breitere Dokumentation und Begleitung der Nutzer. Wir haben uns letztendlich dazu entschieden einen stärkeren Fokus auf die Dokumentation und die Begleittexte zu legen, da wir darin die nachhaltigere Strategie für das Fortbestehen eines Open-Source Projektes sehen.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Da die momentanen satellitengestützten Emissionsdaten sich gut eignen um große räumliche und zeitliche Muster abzubilden, verweisen wir gerne auf die Projekte OpenSenseMap und lufdaten.info, die durch Bodensensoren erfasste Emissionsdaten lokal und zeitlich fein aufgelöst ausliefern können (dafür mit einer geringeren Abdeckung als Satellitendaten). Diese Projekte sind erfreulich komplementär zu unserem und wir stehen in Austausch über zukünftige Entwicklungen.