Richtlinie zum „Software-Sprint“

Pluragraph

Schlussbericht

Zuwendungsempfänger:

Martin Fuchs, Jona Hölderle, Niels Richter GbR

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS17S39 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

In den vergangenen sieben Jahren haben unzählige wissenschaftliche Forschungsprojekte, Daten-Journalismus-Projekte und NGOs unsere Daten genutzt, um die Gesellschaft zu analysieren, zu verstehen und weiterzuentwickeln. Dabei wurden bisher nicht alle Daten öffentlich zugänglich gemacht und externer Zugriff war nicht möglich. Um den „Datenschatz“ möglichst umfangreich zur Verfügung zu stellen wollten wir in der Projektlaufzeit neue Zugriffe über eine Web-API, sowie umfangreiche Export- und Download-Möglichkeiten geschaffen. Hierfür war der Aufbau eines skalierbaren Datenmodells vonnöten. Zudem mussten die bestehenden Daten konsolidiert und überprüft werden. Als größter Meilenstein war dann die Bereitstellung einer API (sowie verschiedener Export-Möglichkeiten) geplant sowie eine neue Visualisierung der noch nicht veröffentlichten Daten.

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

Durch die zur Verfügungstellung von Daten, Analysen & Interpretationen der Social-Media-Nutzung von nicht-kommerziellen Organisationen ermöglichen wir die stetige Verbesserung dieser im Sinne einer demokratischen Gesellschaft. Zudem helfen unsere Daten aktuelle Phänomene wie Shitstorms, Hatespeech, Social Bots oder Fakenews zu verstehen und Lösungsmöglichkeiten für die potentielle Bedrohung demokratischer Strukturen zu entwickeln.

Vertreter von NGOs, Parteien, Journalismus und Forschung profitieren von den APIs indem sie diese nutzen können um die Daten bei uns abzufragen und eigene Datenbestände zu erweitern oder Visualisierungen zu erzeugen. Durch die Nutzung des JSON-Standards ist auch die Integration in die unterschiedlichsten Programmiersprachen gewährleistet.

Durch den direkten Export der Daten in das CSV oder XLS Format ist auch für Nutzer ohne Programmierkenntnisse die Weiterverarbeitung unserer Daten gewährleistet. So kann jeder Excel-Nutzer sich einen Export zusammenstellen und diesen in Excel nach belieben kürzen, erweitern, durchsuchen und weiterverarbeiten.

Ausführliche Darstellung der Ergebnisse

Im Rahmen des Projektzeitraums konnte der Bestand an gemeinnützigen Organisationen, Kultureinrichtungen, Verwaltungseinheiten und politischen Akteure auf 17.000 Organisationen mit über 40.000 Social-Media-Profilen ausgeweitet werden. Alle Organisationen sind in 180 Kategorien kategorisiert und entsprechen den qualitativen Vorgaben.

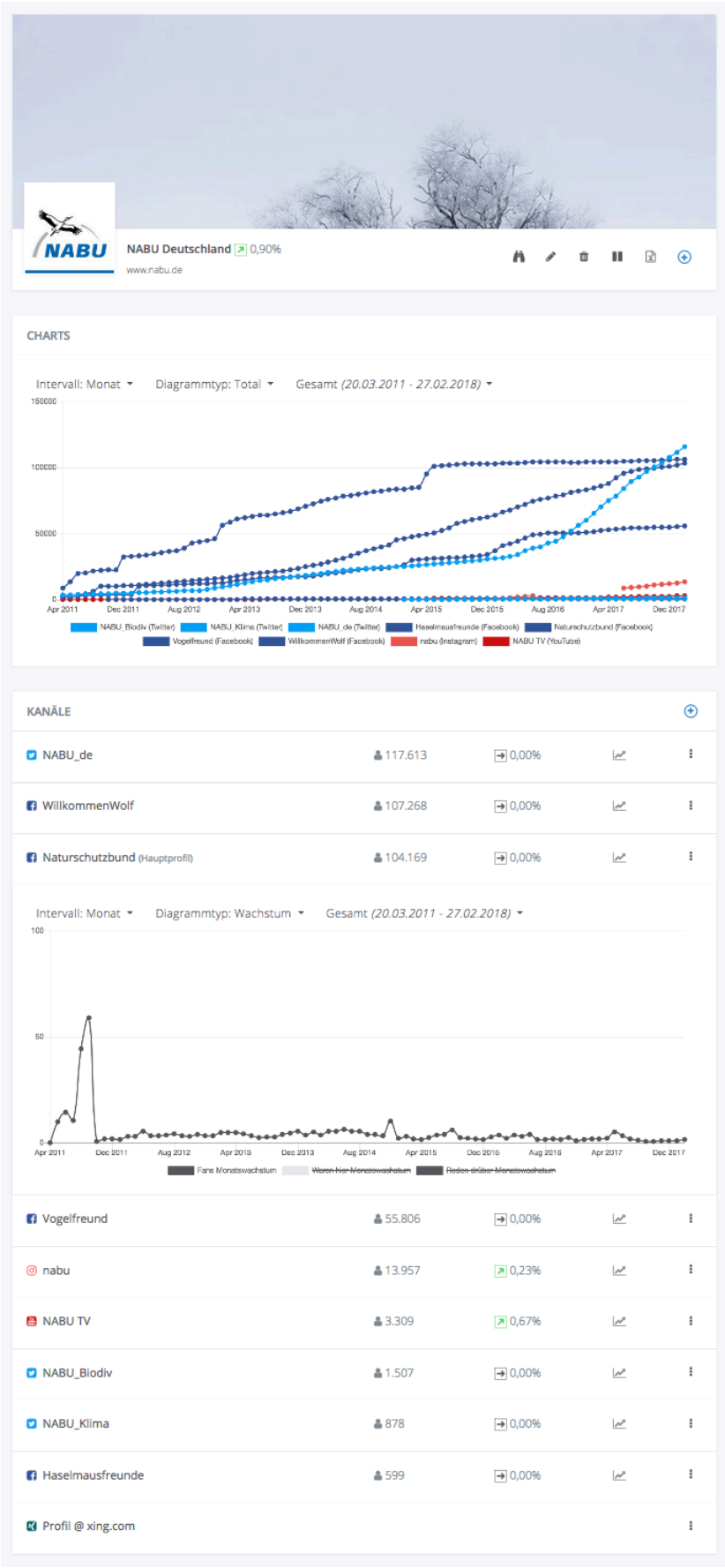
Als Vorbereitung auf die Bereitstellung der Daten wurden die rund 50.000.000 Datenpunkte in ein neues Datenmodell übertragen und Durchschnittswerte (letzte Woche, Monat, Quartal, Jahr, Insgesamt) sowie entsprechende Wachstumswerte zur schnelleren Abfrage vorberechnet.

Auf Basis dieser Daten wurde eine API auf Basis von GraphQL entwickelt. Unter https://pages.pluragraph.de/api/ ist die Dokumentation der API verfügbar. Mittels der API kann auf alle zur Verfügung stehenden Daten zugegriffen werden. Als ersten Beispielcase wurde über die API Wikidata angebunden. 1.395 Organisationen und Personen wurden mit der jeweiligen WikidataID verknüpft und konnten dadurch abgeglichen werden.

Um auch für weniger technisch versierte Nutzer die Daten aufzubereiten gibt es nun alle Daten in einer neuen Darstellung auf https://www.pluragraph.de. Zudem können Datensätze direkt als CSV für Excel und Co. heruntergeladen werden.

Die Betreuung durch die Open Knowledge Foundation hat uns ermöglicht aus dem engen Planungsgerüst etwas herauszutreten und neben der API auch die visuellen Darstellungen für nicht-technische Nutzer zu verbessern. Hierbei war auch das UI-Coaching sehr gut, welches viele neue Ideen initiiert hat. Diese konnten im Rahmen des Projektzeitraums allerdings nur Ansatzweise umgesetzt werden. Im Anschluss an den Projektzeitraum soll deshalb insbesondere die Möglichkeit für Nutzer der Plattform eigene Einträge anzulegen bzw. bestehende zu erweitern verbessert werden.

Beispielhafter Screenshot einer Organisation mit Wachstumsdaten im zeitlichen Verlauf der letzten Jahre:



Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Die Web-Oberfläche von Pluragaph wird hauptsächlich von den Social Media nutzenden Institutionen genutzt um sich selber vergleichen zu können und auf neue Entwicklungen aufmerksam zu werden. Für Journalisten, Wisenschaftler und Datenprojekte ist es durch die autoamtisierte Abfrage der neuen API jetzt möglich direkten Zugriff auf alle aktuellen und historischen Daten der Organisationen, Politiker und Institutionen zu bekommen. Dadurch wird es möglich aktuelle Phänomene wie Shitstorms, Hatespeech, Social Bots oder Fakenews zu verstehen und Lösungsmöglichkeiten für die potentielle Bedrohung demokratischer Strukturen zu entwickeln.

Die Arbeit an den Schnittstellen und mit einzelnen Akteuren aus der Wissenschaft hat gezeigt, dass eine API ein guter erster Schritt ist, die visuelle Aufbereitung auf der Website aber nicht vernachlässigt werden darf. Hier will das Team weiter ehrenatmlich weiterarbeiten. Ebenso hat die Einbindung von WIkidata gezeigt wie wichtig es ist verschiedene Datensilos zu verbinden. Hier ist jedoch manuelle Nacharbeit notwendig, weil die Daten in Wikidata nicht alle aktuell sind.

Das Team hat sich durch das Projekt sehr intensiv mit Auswertungsmöglichkeiten öffentlicher Social-Media-Daten beschäftigt was auch beruflich weiter genutzt werden soll. Auch die technisch nicht so versierten Team-Mitglieder haben durch die einfach verständliche API einen neuen Zugriff auf die Daten bekommen.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Im Bereich Qualitätssicherung gab es die anfängliche Hoffnung technische Wege zu finden, wie Organisationen kategorisiert und verifiziert werden können. Diese Hoffnung hat sich leider früh zerschlagen, sodass hier weiter manuelle Qualitätsicherung betrieben werden muss. Diese wird aber durch Reports und automatisierte Warnungen (z.B. bei zu vielen bzw. ungewöhnlichen Kategorien) vereinfacht.

Im Rahmen des UI-Coachings wurden Schwachstellen in der Web-Oberfläche aufgedeckt. Diese waren nicht Teil des Projektes, sollen aber in Zukunft verbessert werden.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Interessierte können sich unter https://www.pluragraph.de über den aktuellen Stand des Projektes Informieren und auf alle Daten zugreifen. Der Zugriff über die API ist unter https://pages.pluragraph.de/api/ dokumentiert.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Die Aufwandsplanung hat sich von den zur Verfügung gestellten Zeitressourcen als gut bewiesen. Allerdings wurden durch familiäre Umstände (alle drei Team-Mitglieder haben im letzten Jahr ein Kind bekommen) die Arbeitszeiten nicht so kontinuierlich durchgeführt wie geplant. Viel Arbeit hat sich geballt auf einzelne Tage verschoben was eine gemeinsame Bearbeitung der Probleme manchmal etwas schwierig gemacht hat.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Das UI-Coaching hat noch einmal einen grundsätzlichen Fragenkatalog für die Nutzung der Weboberfläche eröffnet. Hier werden nach Abschluss der Förderung noch weitere ehrenamtliche Arbeiten anfallen um auch die Weboberfläche zufriedenstellend umzusetzen.