



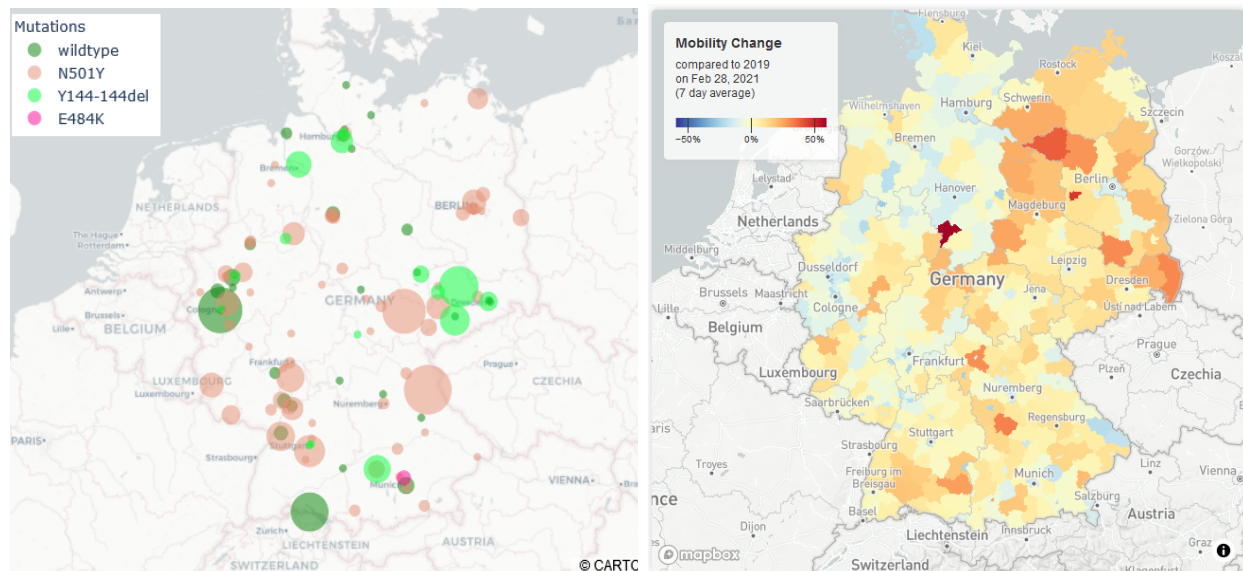
Entwicklung einer Plattform zur Modellierung der Ausbreitung von SARS-CoV-2-Mutationen



Projektbeschreibung:

In diesem Bachelorprojekt soll eine Plattform für eine Ausbreitungsanalyse mit dem Fokus auf Mutationsdaten in Deutschland entwickelt und prototypisch umgesetzt werden. Die Idee ist es, den Ursprung verschiedener Mutationen in Deutschland zu identifizieren und auch deren Ausbreitungsmuster zu visualisieren, um die großflächige Ausbreitung weiterer Mutationen in Zukunft besser modellieren und verhindern zu können. Zu den Datenquellen gehören genomische Mutationen (vom [Robert-Koch-Institut](#), wie in [covradar.net](#) dargestellt) und Mobilitätsmuster innerhalb Deutschlands sowie aus anderen Ländern nach Deutschland. Mobilitätsdaten von Mobiltelefonen ermöglichen es, die Bewegung der Bevölkerung zu erfassen, um Ereignisse zu korrelieren und Verhaltensänderungen und die frühzeitige Erkennung von möglichen Ausbrüchen vorherzusagen.

Die Infrastrukturkomponente dieses Projekts besteht aus der Erstellung eines prototypischen Datenintegrations- und Analyseframeworks im Rahmen der bei [Data4Life](#) verfügbaren Infrastruktur. Data4Life hat es sich zur Aufgabe gemacht, die Gesundheitsversorgung zu verbessern durch die Entwicklung und den Betrieb einer IT-Infrastruktur, die zur sicheren Bereitstellung ("Datenspende"), Analyse und Verwaltung von Gesundheitsdaten dient, sowie die Anwendung von Algorithmen zur Erforschung von Krankheiten und der Auswertung von Behandlungsmethoden ermöglicht. Die Integration neuer Datenquellen in die bestehende Infrastruktur von Data4Life ist dabei unerlässlich. Ihr werdet mit Data4Life zusammenarbeiten, um einen Prototyp für die Automatisierung dieses Prozesses zu entwickeln.



Darstellungen der Mutationshäufigkeit und Mobilitätsänderung an verschiedenen Orten im Februar 2021. (Bilder entnommen aus: [covradar](#) (links) und [covid-19-mobility](#) (rechts). Anmerkung: Die auf der Mutationskarte (links) markierten Orte sind die Standorte der Sequenzierlabore und nicht die Standorte der Patienten, da diese Informationen vertraulich sind.)

Die unten aufgeführten Anforderungen sind im Prototyp zu implementieren:

- Konvertierung der neu aufbereiteten Datenquellen in das Data4Life-Framework;
- Prototypischer Entwurf einer Datenintegrationsarchitektur;
- Software-Implementierung unter Verwendung einer agilen Entwicklungsmethodik;
- Analyse der Interoperabilität zwischen den Daten unter Verwendung von statistischen und Machine-Learning Methoden für die Ausbreitung von SARS-CoV-2-Mutationen;
- Visualisierung der Ausbreitung von Mutationen.

Im Rahmen des Bachelor Projekts kann die technische Infrastruktur der Data4Life Analyse-Plattform genutzt und entsprechend der Anforderungen erweitert werden:

- Selbst betriebene Private Cloud mit Datacentern in Deutschland;
- Verwendung von SAP HANA für die Datenanalyse;
- Microservice Architektur unter Nutzung von Docker und Kubernetes;
- Typescript und Python als zentrale Programmiersprachen.

Projektpartner:

Die [D4L data4life gGmbH](#) ist eine Ende 2017 gegründete gemeinnützige Organisation mit rund 100 Mitarbeitern an Standorten in Europa und Singapur, die aus einem Forschungsprojekt (unter dem Namen Gesundheitscloud) des Hasso-Plattner-Institut für Digital Engineering gGmbH hervorgegangen ist und von der Hasso Plattner Foundation finanziert wird. Data4Life ist im Juli 2019 gemäß ISO 27001 nach dem IT-Grundschutz vom Bundesamt für Sicherheit in der Informationstechnik zertifiziert worden.

Betreuer aus dem Lehrstuhl für Datenanalyse und Computational Statistik:

Ferdous Nasri - Ferdous.Nasri@hpi.de

Fábio Miranda - Fabio.MalcherMiranda@hpi.de

Alice Wittig - Alice.Wittig@hpi.de

Dr. Athar Khodabakhsh - Athar.Khodabakhsh@hpi.de

Prof. Dr. Bernhard Renard - Bernhard.Renard@hpi.de

Bitte zögert nicht, uns eine E-Mail zu schicken, wenn Ihr Fragen habt oder ein Zoom-Meeting vereinbaren möchtet.

Information zu Pandemie-bedingten Vorgehen:

Die Gesundheit aller Beteiligten hat oberste Priorität. Wir werden die universitären und staatlichen Vorgaben beachten und die Möglichkeiten von Vor-Ort-Treffen laufend nach der aktuellen Infektionslage beurteilen. Die DACS-Gruppe und Data4Life verfügen über Büroräume in Potsdam und es besteht die Möglichkeit, Euch dort Besprechungs- und Arbeitsräume anzubieten.