

# Linked COVID-19 Data – Semantische Modellierung von Linked Geodata

Florian Thiery

## Zusammenfassung

Dieser Beitrag beschäftigt sich mit semantischer Modellierung und Linked Geodata im Kontext der COVID-19-Pandemie. Diese neuen Technologien interoperabler und verlinkter Datenhaltung werden in der Geo-Branche immer populärer und weltweit sowohl von administrativen Einrichtungen wie auch von unabhängigen freiwilligen Communities als Volunteered Geographic Information (VGI) eingesetzt. Im Zuge der COVID-19-Pandemie ist die Veröffentlichung dieser (Geo-) Daten von zentraler Bedeutung. Dieser Beitrag zeigt Hintergründe und Möglichkeiten semantischer und Linked Open Technologies mit Wikidata auf. Als Ergebnis dieses Prozesses wird eine Ontologie vorgestellt, die verschiedene Datenqualitäten und Datenquellen homogenisiert. Ergebnisse dieser COVID-19-Datenmodellierungen können als Linked COVID-19 Data in einem Explorer oder mit Hilfe eines QGIS-Plugins visualisiert werden.

## Summary

*This article deals with semantic modelling and Linked Geodata in context of the COVID-19 pandemic. These new technologies of interoperable and Linked Data management are becoming more and more popular in the geo sector and are used worldwide, both by administrative institutions and by independent voluntary communities as Volunteered Geographic Information (VGI). In the course of the COVID-19 pandemic, publishing (geo)data is of central importance. This article shows the background and possibilities of semantic and Linked Open Data technologies using Wikidata. As a result of this process, an ontology is presented, that homogenizes different data qualities and data sources. Results of this COVID-19 data modelling can be visualized as Linked COVID-19 data in an explorer or by using a QGIS plugin.*

**Schlüsselwörter:** Linked Data, Semantische Modellierung, COVID-19, Wikidata, SPARQL Unicorn

## 1 Motivation

Seitdem im Dezember 2019 erste Fälle der Atemwegserkrankung COVID-19 in der chinesischen Millionenstadt Wuhan erkannt wurden, entwickelte sich die ab Januar 2020 in China als Epidemie einzustufende Erkrankung zu einer weltweiten Pandemie. Im Januar 2020 wurde daher eine gesundheitliche Notlage von internationaler Tragweite (GNIT) von der Weltgesundheitsorganisation (WHO) ausgerufen. Mögliche Erstüberträger des Coronavirus auf den Menschen könnten Fledermäuse oder Flughunde

sein, die in Regionen Asiens von Menschen verzehrt und auf Märkten verkauft werden (Zhou et al. 2020).

In Deutschland ist die Pandemie seit Januar 2020 präsent. Hier sind insbesondere die Fälle des Automobilzulieferers Webasto als eine der ersten Quellen zu erwähnen (Brack 2020) (vgl. Kap. 2). Umso wichtiger ist die transparente Veröffentlichung von verlässlichen und wissenschaftlich erhobenen Daten zu akuten Infiziertenzahlen, Neuinfektionen und lokalen Phänomenen. Im Rahmen dessen spielen in einer aufgeklärten, auf Fakten basierenden agierenden, wissenschaftlich geprägten Gesellschaft offene, frei zugängliche und interoperable Schnittstellen der Gesundheitsinstitutionen und anderer universitärer oder außeruniversitärer Einrichtungen eine Rolle (vgl. Kap. 5). Eine große Herausforderung bei der Analyse dieser sehr heterogenen virologischen Daten ist die Visualisierung der Daten und dabei die Offenlegung der Algorithmen und Hintergründe zur Visualisierung. Hier sind insbesondere Geodäten gefragt, da in der COVID-19-Pandemie räumliche Verteilungen wichtige Rollen spielen und auch politisch genutzt werden.

Eine Technik und Strategie, verteilte und verlinkte (räumliche) Daten interoperabel im World Wide Web bereitzustellen, ist das Semantic Web und sogenannte Linked Open Data (LOD) (vgl. Kap. 3). Diese technologischen Trends als ein Zukunftsthema der Geodäsie zu betrachten, wird auch im GDI-DE-Leitfaden zu Geodatendiensten im Internet aufgegriffen (BKG 2019). Darauf aufbauend gibt es bereits interessante Ansätze der Arbeitsgruppe »Research Squirrel Engineers« zur semantischen Modellierung und Bereitstellung von räumlichen Daten als LOD, insbesondere im Bereich des kulturellen Erbes unter Einbeziehung von Community-Daten, sogenannten Volunteered Geographic Information (VGI), wie Wikidata (vgl. Kap. 4) und OpenStreetMap.

Die Bundesregierung hat zur Einbeziehung der Bürger in diesen Prozess im März 2020 einen #WirVsVirus Hackathon ins Leben gerufen. In Anlehnung an den #WirVsVirus Hackathon und unter Einbeziehung von offenen Schnittstellen, interoperablen Daten, Linked Open Data und der Community basierten, freien Wissensdatenbank Wikidata beschreibt dieser Artikel in Kap. 6 den Aufbau einer semantischen Modellierung von COVID-19-Daten als Linked Open Data in einer Ontologie und die Transformation von COVID-19-Daten der Johns Hopkins University (<http://coronavirus.jhu.edu>) und des Robert Koch-Instituts (<http://corona.rki.de>). Die Idee der Research Squirrel Engineers ist es dabei, COVID-19-Daten als Linked Data zu veröffentlichen und im Sinne des SPARQL Unicorn Prinzips (Thiery et al. 2020b)

Nutzern ohne Kenntnisse von Ontologien und SPARQL zur Verfügung zu stellen. Ergebnisse dieser Bereitstellung werden in Kap. 7 als Linked COVID-19 Data Explorer und in Kap. 8 als Linked COVID-19 Data in QGIS vorgestellt und in Kap. 9 mit einem Ausblick abgeschlossen.

## 2 COVID-19

Die durch ein Coronavirus verursachte Lungenerkrankung wird nach Angaben der Weltgesundheitsorganisation (WHO) COVID-19 (Abk. engl. Corona Virus Disease 2019) genannt. Der dafür verantwortliche Erreger ist Sars-CoV-2. Der Ursprung von COVID-19 liegt nach aktuellen wissenschaftlichen Erkenntnissen in China und breitet sich seitdem in einer großen Geschwindigkeit als Pandemie aus (Andersen et al. 2020). In Deutschland »erfasst [das Robert Koch-Institut (RKI)] kontinuierlich die aktuelle COVID-19-Lage, bewertet alle Informationen, schätzt das Risiko für die Bevölkerung in Deutschland ein und stellt Empfehlungen für die Fachöffentlichkeit zur Verfügung« (RKI 2020a). Auf Basis dieser Quelle und anderer Veröffentlichungen aktualisiert auch die Wikipedia-Community Fakten, Daten und räumliche Ausbreitungen weltweit und dokumentiert dabei die unterschiedlichen Maßnahmen zur Eindämmung der Ausbreitung und zu Lockerungsmaßnahmen (Wikipedia 2020).

Im Zuge der Veröffentlichungen des Robert Koch-Instituts sind bestimmte Berechnungen und Zahlenwerte zu einem Entscheidungskriterium der politisch Wirkenen geworden (Wöhrbach 2020). Diese sind als räumlich sehr unterschiedlich zu betrachten und daher besonders aus einer geodätischen Sicht interessant und können zum Beispiel in Karten (vgl. Kap. 7) dargestellt werden. Zu diesen Zahlenwerten gehören die Verdopplungszeit, die Reproduktionszahl R, die Anzahl der Neuinfektionen pro 100.000 Einwohner in einem Zeitraum von sieben Tagen und Berechnungen des Nowcastings. Zu Beginn der Pandemie wurde vor allem die Verdopplungszeit (die Anzahl der Tage, in der sich die Anzahl der Fälle verdoppelt) genutzt. Hierbei wird von einem exponentiellen Wachstum ausgegangen. Wie bei allen Methoden sind insbesondere die Verfügbarkeit der Daten durch mangelnde Tests und die Qualität der Daten als Probleme zu sehen. Die Reproduktionszahl R (wie viele Personen steckt eine infizierte Person statistisch an) und das Nowcasting (Modellrechnung zur Schätzung der aktuellen Infektionsentwicklung) folgt einer Modellrechnung, die jedoch mit einem hohen Fehlerintervall behaftet ist und ihre Schwächen mit der fehlenden Testrate ausdrückt (RKI 2020b).

## 3 Linked Open Geodata

Im Rahmen der rasanten Entwicklung der Geodäsie zu einer »Geodäsie 4.0« im Zuge des Prozesses »Geodäsie 2030 – Arbeiten 4.0« hat dies Auswirkungen auf das digitale und vernetzte Arbeiten des Berufsstandes, welches auch im Arbeitskreis 1 »Beruf« des DVW adressiert wird. Abb. 1 zeigt, dass eine Geodäsie 4.0 (knowledge era) eine Geodäsie 3.0 (semantic era) voraussetzt, welche auf semantischer Modellierung, Linked Data, Standards und Interoperabilität basiert (Thiery 2019a).

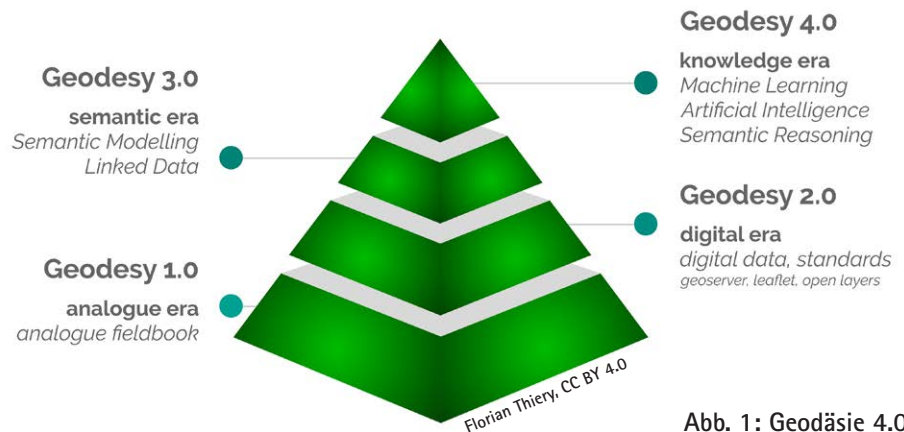


Abb. 1: Geodäsie 4.0

Arbeiten 4.0 in einem Web 4.0 zum Erreichen einer Geodäsie 4.0 setzt das Veröffentlichen von Linked Data (LD), Linked Open Data (LOD) and Linked Open Usable Data (LOUD) voraus (Berners-Lee et al. 2001). Heute bietet das Web Geodäten und Wissenschaftlern die Möglichkeit, ihre Geodaten zu teilen und ermöglicht ihnen, am wissenschaftlichen und politischen Diskurs teilzunehmen. Ein Großteil dieser gemeinsam genutzten Daten ist jedoch nicht auffindbar oder zugänglich, was zu modernen unbekanntem Datendracen führt. Dies ist die moderne Interpretation antiker Kartographen, dass Gebiete, die unbekannt waren, mit dem Ausdruck »Hic sunt dracones« (dt.: Hier sind Drachen) beschrieben wurden. Oft fehlen diesen Datendracen Verbindungen zu anderen Datensätzen, d.h. sie sind nicht interoperabel und daher nicht nutzbar und wiederverwendbar. Um diesen Mangel zu beheben, können eine Reihe von Techniken, Standards und Empfehlungen verwendet werden: Semantic Web und verknüpfte (offene) Daten, die FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) und LOUD-Daten (Thiery 2019b). Tim Berners-Lee führte das Konzept des Semantic Web ein, indem er die Ideen von Open Data, semantisch beschriebene Ressourcen und Links sowie verwendbare (maschinenlesbare) Schnittstellen und Anwendungen zur Erstellung eines riesigen globalen Graphen verwendete (Thiery et al. 2019). Linked Open Data (LOD) sind dabei verknüpfte Daten, die unter einer offenen Lizenz veröffentlicht werden (Berners-Lee 2006). Darüber hinaus muss LOD für Wissenschaftler und Programmierer verwendbar sein, um die gesamte LOD-Leistung voll ausnutzen zu können.

Das Befolgen der LOUD-Prinzipien wird LOD noch FAIRer machen.

Einige Verwaltungsbehörden sowie community-driven, ehrenamtlich betriebene Wissensdatenbanken bieten bereits Geodaten als LOD an, die mit verschiedenen Ressourcen im Internet verknüpft sind. Sie wachsen ständig, da sie mit Quelleninformationen angereichert und mit verwandtem Material in anderen offiziellen Datenbanken verknüpft werden (Thiery et al. 2020a). Besonders ist hier das Projekt »Intelligente Datenerfassung, Haltung und Bereitstellung innerhalb der Öffentlichen Verwaltung« zu erwähnen, dessen Ziel es ist »eine Linked Data Infrastruktur am Bundesamt für Kartographie und Geodäsie anhand von einigen ausgewählten Datenbeständen aufzusetzen und zu integrieren. Hierbei sollen Ontologien für Datenstandards definiert und Best Practices für die semantische Integration in der Praxis ausgetestet werden« (Homburg et al. 2019). Das White Paper »OGC Benefits of Representing Spatial Data Using Semantic and Graph Technologies« zeigt Vorteile bei der Darstellung von Geodaten mit Hilfe von semantischen Technologien auf und wird als Basis für einen neuen GeoSPARQL-Standard dienen (Abhayaratna et al. 2020).

#### 4 Wikidata

Wikidata ist eine sekundäre Datenbank für strukturierte Daten, die 2012 erstellt wurde (Vrandečić und Krötzsch 2014). Sie ist eine freie und offene Wissensdatenbank, in der jeder Nutzer ehrenamtlich und freiwillig Daten hinzufügen und bearbeiten kann. Wikidata ist damit, wie auch OpenStreetMap, Teil der VGI-Initiative. Wikidata ist der zentrale Speicher für strukturierte Daten von Wikimedia-Projekten, z.B. Wikipedia und Wiktionary. Die in Wikidata enthaltenen Daten sind unter einer freien Lizenz (CC0) verfügbar, mehrsprachig, für Mensch und Maschine zugänglich (GUI, API, SPARQL), in Standardformaten exportierbar (z.B. JSON, RDF, SPARQL) und mit anderen offenen Datensätzen in der Linked Data Cloud verlinkt. Das Datenmodell von Wikidata enthält Elemente (z.B. label, description, alias, identifier) und Aussagen (z.B. property, value, qualifier, reference), siehe Trognitz und Thiery (2019).

Geodaten können in Wikidata z.B. mit der property »P625« (coordinate location) oder »P5140« (coordinates of geographic center), also WGS84-Punktdaten, hinterlegt werden. Administrative Einheiten wie Staaten und Gebietseinheiten sind zudem auch mit OpenStreetMap (P402) und anderen Gazetteers (z.B. Geonames P1566) verknüpft. COVID-19-Daten werden von der Wikidata-Community ebenfalls gesammelt und in einem Wiki-Projekt zur Verfügung gestellt (Wikidata 2020). Beispiele hierfür sind Daten zur weltweiten Ausbreitung in »Q81068910« und zur Ausbreitung in Deutschland in »Q83889294«.

Aus der Erkenntnis heraus, dass Volunteers (administrative oder private Initiativen) Open Data veröffentlichen und freiwillige Initiativen der Gemeinschaft zur Sammlung von Daten wie Wikidata unterstützen wollen, ist das SPARQL Unicorn entstanden. Denn es mangelt an benutzerfreundlichen, leicht zu bedienenden und offen verfügbaren Werkzeugen, insbesondere für Linked-Open-Data-Technologien und -Repositorien sowie für Wikidata selbst. Zur Überwindung dieses Engpasses soll das SPARQL Unicorn helfen. Dabei werden die Research Squirrel Engineers Aktivitäten (z.B. Werkzeuge, Projekte) nach den SPARQL Unicorn principles entwickeln und pflegen, andere verwandte Anwendungen entwickeln sowie Outreach- und Support-Aktivitäten generieren, um das SPARQL Unicorn in Forschung und Lehre einzubinden (Thiery et al. 2020b). Am Problem der fehlenden Verfügbarkeit von Tools für Semantic Web und Geodaten setzt das SPARQLing Unicorn QGIS Plugin an, welches im Sinne des SPARQL Unicorns die Ausführung von Linked Data Anfragen in (Geo)SPARQL an ausgewählte Triplestores und geofähige SPARQL Endpoints ermöglicht und die Ergebnisse der Anfragen somit in QGIS für die Geocommunity aufbereitet (Thiery und Homburg 2020).

#### 5 Open Access und COVID-19

Die hohe Dynamik und die drängende Suche nach Lösungen in der COVID-19-Pandemie erzeugen eine hohe Transparenz und eine Vielzahl an frei nutzbaren Daten und Schnittstellen. Insbesondere die Johns Hopkins University (JHU), das Robert Koch-Institut (RKI) und das European Centre for Disease Prevention and Control (ECDC) stellen ihre Daten mit APIs und Dashboards ins World Wide Web, sodass Bürger und Wissenschaftler frei auf diese Daten zugreifen können. Für Deutschland sei hier der NPGEO Corona Hub 2020 genannt, für weltweite Daten kann das COVID-19-Dashboard des Center for Systems Science and Engineering (CSSE) an der JHU genutzt werden. Beide basieren auf der Technologie der Firma Esri, das Dashboard des RKI wurde mit Hilfe des *Esri Disaster Response Programs* umgesetzt. Auf der Wissensplattform Zenodo werden Forschungsdaten und Veröffentlichungen in der »Coronavirus Disease Research Community – COVID-19« gesammelt. Eine Übersicht über COVID-19 in Wikidata gibt Colavizza (2020). Statistische Auswertungen der RKI-Daten können z.B. mit dem Software Framework R im Tool »covid19germany« (Schmid 2020) erzeugt werden, welches aus dem #WirVsVirus Hackathon entstanden ist. Zusätzliche Daten sind auch über das COVID-19 Data Portal und fairsharing.org abrufbar. Ontologien zur Beschreibung der COVID-19-Daten sind ebenfalls schon veröffentlicht: Johns Hopkins Daily COVID-19 CSV Dumps Ontology (JHU 2020) und CODO (Dutta 2020).

## 6 Linked COVID-19 Data

Linked COVID-19 Data (<http://covid19data.link>) ist eine Linked Data Plattform für COVID-19-Daten. Ziel ist es, verschiedene heterogene, frei verfügbare Schnittstellen des Robert Koch-Instituts (RKI), der Johns Hopkins University (JHU) und des European Centre for Disease Prevention and Control (ECDC) durch Python-Skripte (DOI: 10.5281/zenodo.3763504) in Linked Data zu transformieren.

Wikidata verknüpft sind, z. B. »*world:Germany*« mit »*wikidata:Q183*«. Über diese Verknüpfung sind durch Techniken des Semantic Web Anreicherungen der Daten um weitere Inhalte, wie Einwohnerzahlen oder weitere Koordinaten, möglich. Die in diesem Datensatz modellierten Geodaten sind im GeoSPARQL-Standard modelliert.

Die transformierten Linked COVID-19-Daten sind durch eine Schnittstelle, einen SPARQL Endpoint, verfügbar und können so in verschiedene Anwendungen,

```
for c in countriesJHU:
    cstring = str(c)
    cstring = cstring.replace(" ", "_").replace("(", "").replace(")", "").replace("'", "").replace(":", "").replace(".", "")
    thiscountry = dataJHU[c]
    for item in thiscountry:
        i = i+1
        m = hashlib.md5()
        m.update(cstring + str(item['date']) + "JHU")
        UUID = str(int(m.hexdigest(), 16))[0:16]
        dstrArr = str(item['date']).split("-")
        dstr = str(dstrArr[0]).zfill(2) + "-" + str(dstrArr[1]).zfill(2) + "-" + str(dstrArr[2]).zfill(2) + "T00:00:00.000Z"
        lines0.append("covid19:" + UUID + " " + "rdf:type" + " covid19:JHU_Dataset .")
        if cstring == "US":
            cstring = "United_States"
        lines0.append("covid19:" + UUID + " " + "covid19:country" + " world:" + cstring + ".")
        lines0.append("covid19:" + UUID + " " + "covid19:date" + " " + "'" + dstr + "'" + ".")
        lines0.append("covid19:" + UUID + " " + "covid19:confirmed" + " " + "'" + str(item['confirmed']) + "'" + ".")
        lines0.append("covid19:" + UUID + " " + "covid19:deaths" + " " + "'" + str(item['deaths']) + "'" + ".")
        lines0.append("covid19:" + UUID + " " + "covid19:recovered" + " " + "'" + str(item['recovered']) + "'" + ".")
        lines0.append("")
```

Abb. 2: Datentransformation JHU-Daten

Die besondere Herausforderung besteht darin, dass die Daten des RKI, der JHU und des ECDC in unterschiedlichen Formaten und verschiedenen Aggregationsformaten vorliegen. Zur Vereinheitlichung der resultierenden Daten (veröffentlicht unter Thiery 2020a, 2020b) wurde eine COVID-19-Ontologie (DOI: 10.5281/zenodo.3757828) entwickelt. Beispiele für Datentransformationen sind Abb. 2 und Abb. 3 zu entnehmen.

Hierbei ist wichtig zu erwähnen, dass die COVID-19-Ontologie Attribute (properties) für bestätigte Fälle (confirmed), Todesfälle (deaths) und Genesungen (recovered) vorsieht. Je nach Datenanbieter können diese Daten jedoch qualitätsmäßig variieren oder unterschiedliche Bedeutungen haben. So gibt es im Datensatz des RKI mehrere Zweitstempel, wie das Meldedatum und das Erkrankungsdatum. In der Anwendung wird das Erkrankungsdatum zur Generierung des Zeitstempels genutzt. Die Datentransformation stellt hier den Ist-Zustand der einzelnen Quellen dar und tätigt keinerlei inhaltliche Analyse und Interpretation.

Zur Verknüpfung der einzelnen Datensätze mit räumlichen Informationen werden die properties »*covid19:country*«, »*covid19:bundesland*« und »*covid19:landkreis*« genutzt. Die verknüpften Ressourcen verweisen auf den RSE Geometry Datensatz (DOI: 10.5281/zenodo.3876257). Dieser Geometriedatensatz ist in die Linked Data Cloud eingebunden, da einzelne Länder und Regionen direkt mit

wie den Linked COVID-19 Data Explorer (Kap. 7) und das SPARQLing Unicorn QGIS Plugin (Kap. 8), eingebunden werden.

## 7 Linked COVID-19 Data Explorer

Auf Basis des in Kap. 6 beschriebenen Workflows zur Generierung von georeferenzierten COVID-19-Daten bietet der Linked COVID-19 Data Explorer eine Kombination aus Karten und Diagrammauswertungen an. Durch die Verknüpfung der einzelnen Datensätze mit Wikidata können z. B. auch Statistiken unter Einbeziehung der Einwohnerzahlen erstellt werden. Abb. 3 zeigt die aktuellen Fallzahlen zum 02.06.2020 nach Angaben der Johns Hopkins University. In Abb. 4 ist der prozentuale Anteil der Todesfälle im Bezug zur Einwohnerzahl (Daten aus Wikidata) zum Stichtag 02.06.2020 aufgezeigt. In Abb. 5 kann der unterschiedliche (exponentielle) Anstieg in den einzelnen Ländern nachvollzogen werden. All diese Visualisierungen basieren auf SPARQL-Anfragen an den COVID-19 Data SPARQL Endpoint (<https://digits.mainzed.org/covid19>).

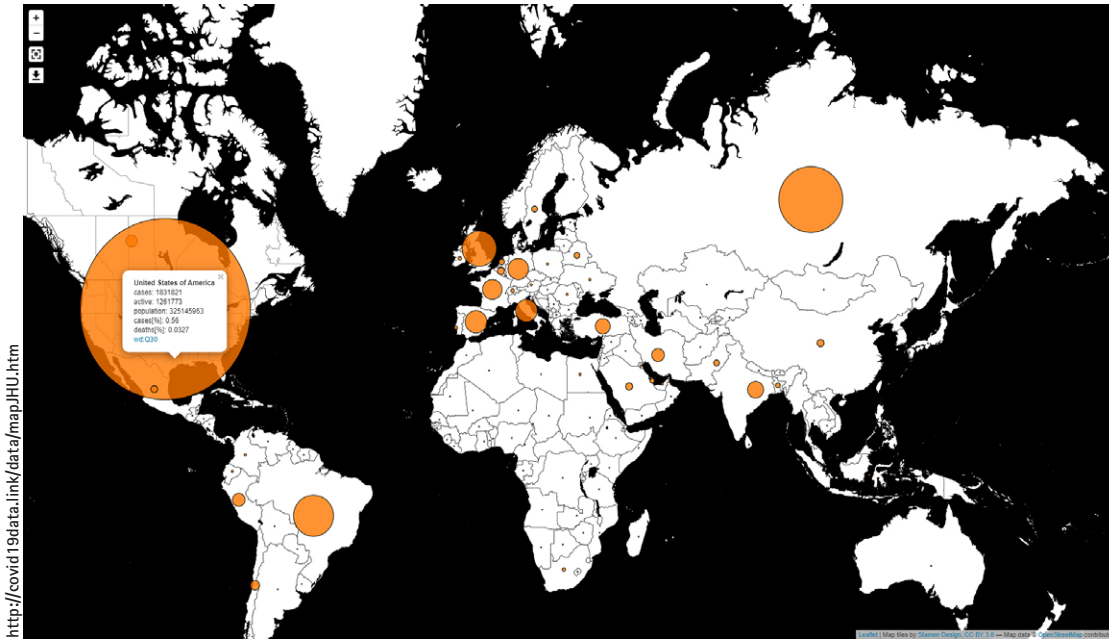


Abb. 3: COVID-19-Fälle nach Daten der Johns Hopkins University (Stand: 02.06.2020)

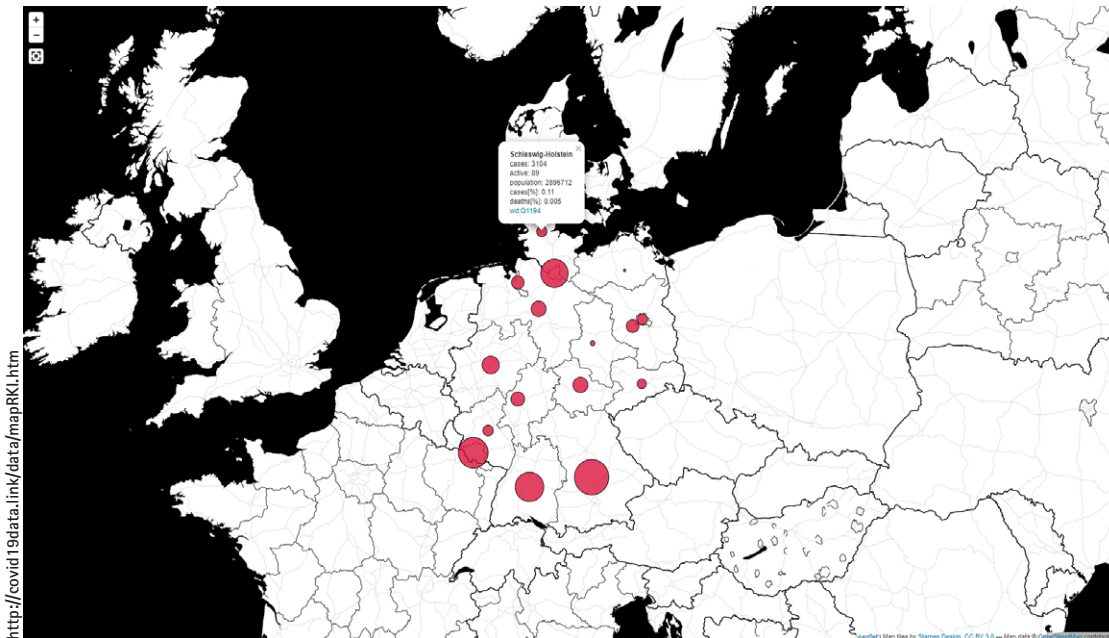


Abb. 4: COVID-19-Todesfälle prozentual zur Einwohnerzahl nach Daten des Robert Koch-Instituts (Stand: 02.06.2020)

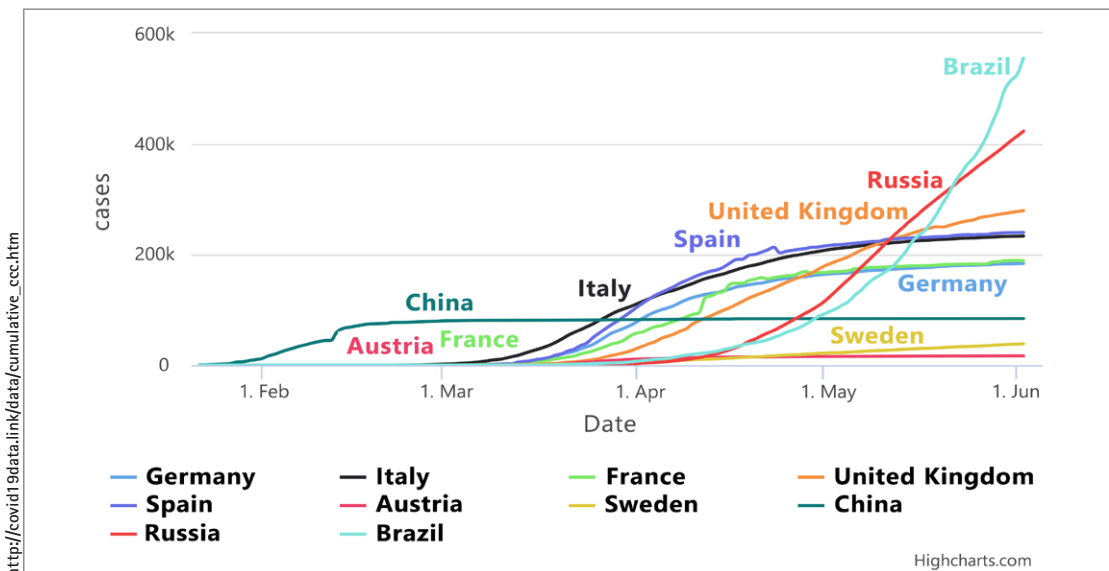


Abb. 5: COVID-19-Fälle nach Daten der Johns Hopkins University (Stand: 02.06.2020)

## 8 Linked COVID-19 Data in QGIS

Das SPARQLing Unicorn QGIS Plugin (QGIS 2020) ab Version 0.10 unterstützt die Einbindung des COVID-19 Triplestores. Somit kann mit einer einfachen SPARQL-Abfrage ein QGIS Layer erzeugt werden. Abb. 6 zeigt eine thematische Auswertung der von der JHU ermittelten Fallzahlen im Mai 2020 in einer logarithmischen Skala. In Abb. 7 sind das SPARQLing Unicorn QGIS Plugin und die erforderliche SPARQL Query zu sehen.

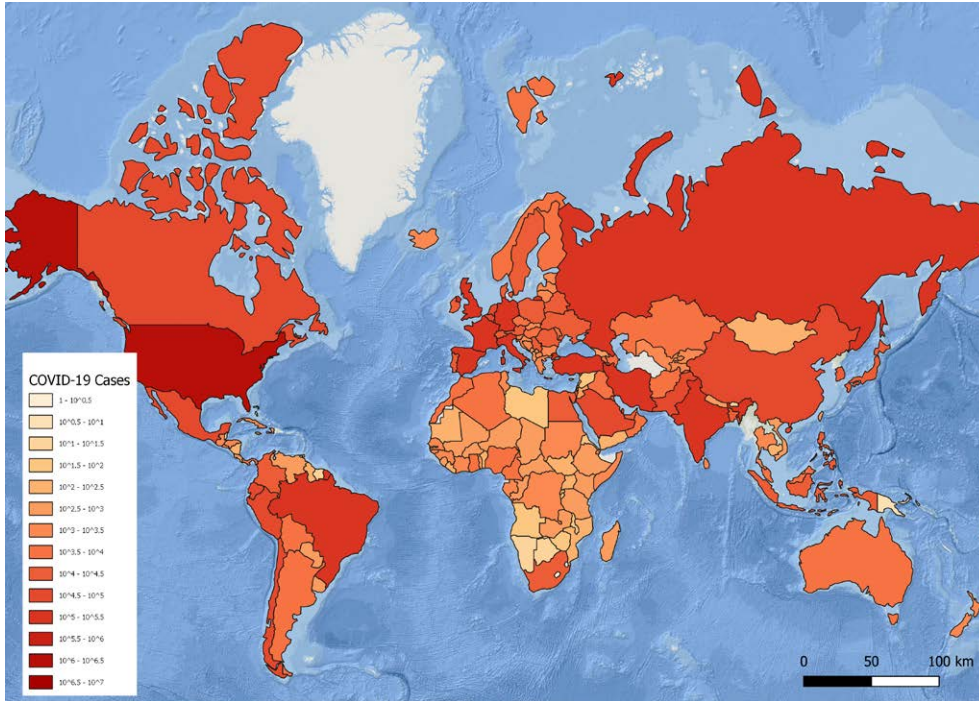


Abb. 6: Fälle nach Daten der Johns Hopkins University (Stand: Mai 2020)

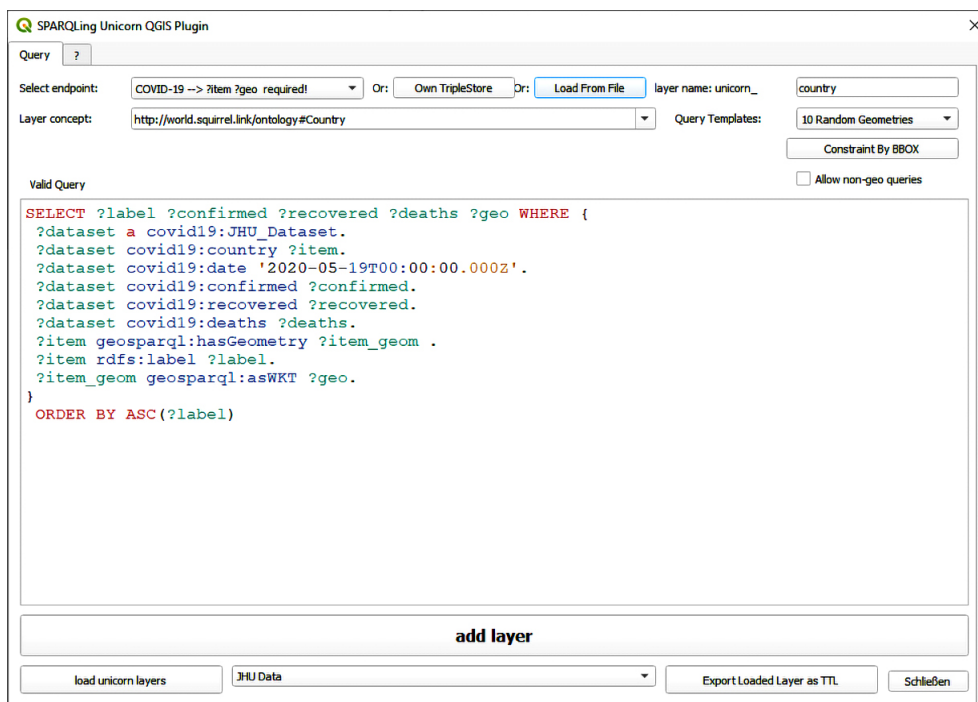


Abb. 7: Beispiel einer SPARQL-Abfrage

## 9 Fazit und Ausblick

Die COVID-19-Pandemie wird uns in Deutschland und auch die gesamte Weltgemeinschaft noch Monate, wenn nicht Jahre beschäftigen. Geodäten können durch ihre Expertise mit räumlichen Daten dazu beitragen, politischen Entscheidungsträgern Informationen zukommen zu lassen. Zurzeit wird dies z.B. mit Hilfe von Produkten der Firma ESRI getan. Darüber hinaus besitzen Geodäten Expertise in Datenmodellierungen und Interoperabilität,

sodass wir auch hier unseren Beitrag leisten können. Geodäten verfügen zudem über Techniken im Umgang mit Datenqualitäten und Dateninhomogenitäten (vgl. Homburg et al. 2018). Im Linked COVID-19 Data Projekt mündet dies im Versuch einer Homogenisierung der COVID-19-Daten in einer Ontologie. Weitergehende Analysen, die auf unterschiedlichen Quellen fußen, müssen diese Datenqualität stark im Blick behalten.

Im Falle der Linked COVID-19-Daten wird daher versucht, eine objektive Ist-Zustands-Sicht als Daten bereitzustellen ohne Wertung der Datenqualität, jedoch mit Verweis auf die Urheber der Daten. Linked COVID-19 Data Modellierungen sind ein erster Schritt, COVID-19-Geodaten als GeoSPARQL zu modellieren und so die Geocommunity von semantischen Technologien zu überzeugen. Volunteered Geographic Information und semantische Geodatenmodellierung werden in der Geocommunity zukünftig eine immer größere Rolle spielen. Das wird insbesondere deutlich, da das Bundesamt für

Kartographie und Geoinformation in Linked Data Forschungsinitiativen setzt. Hürden zur Nutzung von Linked Geodata in QGIS Software können zum Beispiel durch die Integration des SPARQLing Unicorn QGIS Plugin gelöst werden.

Zuletzt sei dem Mainzer Zentrum für Digitalität in den Geistes- und Kulturwissenschaften und den Research Squirrel Engineers, insbesondere Sophie Charlotte Schmidt und Timo Homburg, für die Unterstützung in diesem kleinen COVID-19 Data Projekt gedankt.

## Literatur

Abhayaratna, J., van den Brink, L., Car, N., Atkinson, R., Homburg, T., Knibbe, F., McGlenn, K., Wagner, A., Bonduel, M., Rasmussen, M. H., Thiery, F. (2020): [https://github.com/opengeospatial/geosemantics-dwg/blob/master/white\\_paper/wp.pdf](https://github.com/opengeospatial/geosemantics-dwg/blob/master/white_paper/wp.pdf), letzter Zugriff 04.06.2020.

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., Garry, R. F. (2020): The proximal origin of SARS-CoV-2. In: *Nature Medicine*, Volume 26, 450–452. DOI: 10.1038/s41591-020-0820-9.

Berners-Lee, T. (2006): [www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html), letzter Zugriff 04.06.2020.

Berners-Lee, T., Hendler, J., Lassila, O. (2001): The Semantic Web. In: *Scientific American*, Volume 284, Nr. 5, 34–43. DOI: 10.1038/scientificamerican0501-34.

BKG (2019): GDI-DE. Geodatendienste im Internet. Ein Leitfadens. 4. Aufl., Frankfurt a. M.

Brack, G. (2020): <https://t1p.de/im4y>, letzter Zugriff 04.06.2020.

Colavizza, G. (2020): COVID-19 research in Wikipedia. In: *bioRxiv*. DOI: 10.1101/2020.05.10.087643.

Dutta, B. (2020): <https://github.com/biswanathdutta/CODO>, letzter Zugriff 04.06.2020.

Homburg, T., Prudhomme, C., Boochs, F. (2018): <https://t1p.de/88kj>, letzter Zugriff 04.06.2020.

Homburg, T., Steppan, S., Würriehausen, F. (2019): Semantic Data Integration and Quality Assurance of Thematic Maps in the German Federal Agency for Cartography and Geodesy. In: Abramowicz, W., Corchuelo, R. (Hrsg.): *Business Information Systems Workshops. BIS 2019. Lecture Notes in Business Information Processing*. Volume 373, Springer, Cham. DOI: 10.1007/978-3-030-36691-9\_46.

JHU (2020): <https://t1p.de/ftd11>, letzter Zugriff 04.06.2020.

QGIS (2020): <https://plugins.qgis.org/plugins/sparqlunicorn>, letzter Zugriff 04.06.2020.

RKI (2020a): [https://www.rki.de/DE/Home/homepage\\_node.html](https://www.rki.de/DE/Home/homepage_node.html), letzter Zugriff 04.06.2020.

RKI (2020b): <https://t1p.de/eumxt>, letzter Zugriff 04/06/2020.

Schmid, C. (2020): <https://github.com/nevrome/covid19germany>, letzter Zugriff 04.06.2020.

Thiery, F. (2019a): Linked Data Cloud: Bindeglied zwischen Geodäsie und Gesellschaft. DOI: 10.5281/zenodo.3403029.

Thiery, F. (2019b): Sphere 7 Data: LOUD and FAIR Data for the Research Community. In: Florian Thiery et al. (Hrsg.): *Research Squirrel Engineers – Squirrel Papers*. DOI: 10.5281/zenodo.2643469.

Thiery, F. (2020a): Linked COVID-19 Data: Johns Hopkins University (JHU) and European Centre for Disease Prevention and Control (ECDC). DOI: 10.5281/zenodo.3757279.

Thiery, F. (2020b): Linked COVID-19 Data: Robert Koch-Institut (RKI). DOI: 10.5281/zenodo.3757283.

Thiery, F., Homburg, T. (2020): QGIS – A SPARQLing Unicorn? Eine Einführung in Linked Open Geodata zur Integration von RDF in QGIS Plugins. In: FOSSGIS e.V., Thomsen, J. (Hrsg.): *FOSSGIS 2020: Anwenderkonferenz für Freie und Open Source Software für Geoinformationssysteme*, FOSSGIS e.V., 68–72. DOI: 10.5281/zenodo.3719127.

Thiery, F., Homburg, T., Schmidt, S.-C., Trognitz, M., Przybilla, M. (2020a): SPARQLing Geodesy for Cultural Heritage – New Opportunities for Publishing and Analysing Volunteered Linked (Geo-)Data. In: *Fédération Internationale des Géomètres (Hrsg.): FIG Working Week 2020 Proceedings*. ISBN 978-87-92853-93-6. DOI: 10.5281/zenodo.3751769.

Thiery, F., Schmidt, S.-C., Homburg, T., Trognitz, M. (2020b): The SPARQL Unicorn: An introduction. In: Florian Thiery et al. (Hrsg.): *Research Squirrel Engineers – Squirrel Papers*. DOI: 10.5281/zenodo.3742185.

Thiery, F., Trognitz, M., Gruber, E., Wigg-Wolf, D. (2019): Hic sunt dracones! the modern unknown Data Dragons. In: Florian Thiery et al. (Hrsg.): *Research Squirrel Engineers – Squirrel Papers*. DOI: 10.5281/zenodo.3345711.

Trognitz, M., Thiery, F. (2019): Wikidata – A SPARQL(ing) Unicorn?. DOI: 10.5281/zenodo.3471404.

Vrandečić, D., Krötzsch, M. (2014): Wikidata: a free collaborative knowledgebase. In: *Scientific American*, Volume 57, Nr.10, 78–85. DOI: 10.1145/2629489.

Wikidata (2020): [www.wikidata.org/wiki/Wikidata:WikiProject\\_COVID-19](http://www.wikidata.org/wiki/Wikidata:WikiProject_COVID-19), letzter Zugriff 04.06.2020.

Wikipedia (2020): <https://de.wikipedia.org/wiki/COVID-19-Pandemie>, letzter Zugriff 04.06.2020.

Wöhrbach, O. (2020): <https://t1p.de/eivc>, letzter Zugriff 04.06.2020.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L. (2020): A pneumonia outbreak associated with a new coronavirus of probable bat origin. In: *Nature*, Volume 579, 270–273. DOI: 10.1038/s41586-020-2012-7.

## Kontakt

Florian Thiery M.Sc.  
 Research Squirrel Engineers  
 Josef-Traxel-Weg 4, 55128 Mainz  
 rse@fthiery.de

Dieser Beitrag ist auch digital verfügbar unter [www.geodaesie.info](http://www.geodaesie.info).