

Hochschule Merseburg
University of Applied Sciences



Fachbereich Wirtschaftswissenschaften und Informationswissenschaften

Fachgebiet Wirtschaftsinformatik

Masterarbeit:

**Evaluierung der Nutzung von Geowebsservices im digitalen
Baugenehmigungsverfahren**

vorgelegt bei

Prof. Dr. Ronny Weinkauf

Sommersemester 2023

Zweitprüfer: Martin Becker

Eingereicht von:

Tim Bernutz

Studiengang: Wirtschaftsinformatik (Master)

Datum der Einreichung: 16.11.2023

Abstract

Das Ziel dieser Masterarbeit ist es zu bestimmen, ob sich frei zugängliche Geowebsservices für eine automatisierte Nutzung im digitalen Baugenehmigungsverfahren eignen. Die Dienste werden zum Zeitpunkt dieser Arbeit manuell und dementsprechend zeitaufwendig abgefragt. Zur Erreichung dieses Ziels muss überprüft werden, ob der Service sich für die Automatisierung eignet und ob relevante Daten zur Weiterverarbeitung geliefert werden. Außerdem muss ein Messwerkzeug für die Datensammlung implementiert werden. Die Abfragen dürfen nicht länger dauern als eine manuelle Abfrage. Um dieses Ziel zu erreichen, wird ein Experiment durchgeführt. Dabei werden die jeweiligen Datensätze in Dateien abgespeichert. Die Ergebnisse werden mithilfe der deskriptiven Statistik und Hypothesentests analysiert und ausgewertet. Die Auswertung hat gezeigt, dass eine Automatisierung dieses Prozesses möglich ist und die Abfragen im Durchschnitt 0,2 Sekunden benötigt haben. Dementsprechend bietet das Messwerkzeug einen deutlichen Mehrwert gegenüber händischen Abfragen. Außerdem liefern die Dienste relevante Daten für die weitere Verarbeitung. Jedoch gab es bei der Erreichbarkeit des Dienstes ein paar Einbrüche, die weiter untersucht werden müssen.

Schlagwörter: Automatisierung, Baugenehmigungsverfahren, Experiment, Geography Markup Language, Geowebsservice, Messwerkzeug, Web Feature Service

The aim of this master's thesis is to determine whether freely accessible geowebsservices are suitable for automated use in the digital planning permission process. At the time of this thesis, the services are queried manually and are therefore time-consuming. To achieve this goal, it is necessary to check whether the service is suitable for automation and whether relevant data is supplied for further processing. In addition, a measurement tool must be implemented for data collection. The queries must not take longer than a manual query. An experiment is carried out to achieve this goal. The respective data records are saved in files. The results are analysed and evaluated using descriptive statistics and hypothesis tests. The evaluation has shown that it is possible to automate this process and that the queries took an average of 0.2 seconds. Accordingly, the measurement tool offers significant added value compared to manual queries. The services also provide relevant data for further processing. However, there were a few drops in the accessibility of the service that need to be investigated further.

Inhaltsverzeichnis

Abstract.....	II
Abbildungsverzeichnis.....	V
Tabellenverzeichnis.....	VI
Abkürzungsverzeichnis	VII
1 Einleitung.....	1
2 Theoretischer Rahmen.....	5
2.1 Extensible Markup Language	5
2.1.1 Geography Markup Language	7
2.2 Deskriptive Statistik.....	9
2.2.1 Univariate Analyse	9
2.2.2 Multivariate Analyse	11
2.3 Web Application Architecture.....	13
2.4 Geowebsservices.....	16
2.4.1 Web Feature Service.....	16
2.4.2 Infrastructure for Spatial Information in Europe.....	18
2.5 Geographic Information Systems.....	19
2.6 Experimentieren nach Wohlin	21
3 Zieldefinitionsphase	23
4 Planungsphase.....	26
4.1 Formulierung der Hypothesen	26
4.2 Auswahl der Variablen und Subjekte	28
4.3 Versuchsplanung	29
4.4 Instrumentierung	30
4.5 Konzeption des Messwerkzeugs	30
4.5.1 Anwendungsfalldiagramm.....	31
4.5.2 Klassendiagramm.....	33
4.5.3 Sequenzdiagramm	35

4.6	Eingabedatei für das Messwerkzeug	38
4.7	Eingabedatei als Webseite	41
4.8	Messwerkzeug	42
4.9	Ausgabedateien des Messwerkzeugs	46
4.9.1	Textdatei	46
4.9.2	CSV-Datei	46
4.9.3	XML-Datei	50
4.9.4	XML-Datei aus der POST-Abfrage	50
5	Durchführungsphase	52
5.1	Vorbereitungen	52
5.2	Ausführung durch Anwendungsfälle	53
5.2.1	POST	53
5.2.2	Hochschule Merseburg	56
5.2.3	Aktivierung des Testlaufs	58
5.2.4	Abfrage eines ausländischen Web Feature Service	59
5.2.5	Experiment	60
5.3	Überprüfung der Daten	63
6	Analyse und Interpretation	64
6.1	Deskriptive statistische Auswertung	64
6.2	Reduzierung des Datensatzes	73
6.3	Hypothesentests	74
6.3.1	Erste Hypothese	74
6.3.2	Zweite Hypothese	77
6.3.3	Dritte Hypothese	78
6.4	Diskussion der Ergebnisse	81
7	Zusammenfassung und Ausblick	83
	Literaturverzeichnis	86
	Eidesstattliche Erklärung	88

Abbildungsverzeichnis

Abbildung 1: Anwendungsfalldiagramm des Messwerkzeugs	32
Abbildung 2: Klassendiagramm einer Messung	34
Abbildung 3: Sequenzdiagramm GET	36
Abbildung 4: Sequenzdiagramm POST	37
Abbildung 5: Ausschnitt aus der Abfrage der GetCapabilities des BfN	38
Abbildung 6: Eingabemaske einer Webseite für das Messwerkzeug	41
Abbildung 7: Bounding Boxen für eine Abfrage	45
Abbildung 8: automatisch generierte XML-Datei vom Messwerkzeug	54
Abbildung 9: Bounding Box mit dem Hauptgebäude der Hochschule Merseburg	57
Abbildung 10: Eingabedatei Teil 1	61
Abbildung 11: Eingabedatei Teil 2	62
Abbildung 12: Histogramm für die gemessene Systemzeit	67
Abbildung 13: Histogramm für die Anzahl der Antwortobjekte	68
Abbildung 14: Histogramm für die Startzeit aus Excel	69
Abbildung 15: Boxplot für die gemessene Systemzeit	70
Abbildung 16: Boxplot für die gemessene Systemzeit exklusive der Extremwerte	71
Abbildung 17: Streudiagramm mit der gemessenen Systemzeit und dem Statuscode	76
Abbildung 18: Streudiagramm mit der gemessenen Systemzeit und der Startzeit	79
Abbildung 19: Streudiagramm mit der gemessenen Systemzeit und der Arbeitszeit	80

Tabellenverzeichnis

Tabelle 1: Vorlagen für die Zieldefinition.....	24
Tabelle 2: gekürztes Abfrageergebnis einer POST-Abfrage	56
Tabelle 3: gekürztes Abfrageergebnis der Hochschule Merseburg	58
Tabelle 4: gekürztes Abfrageergebnis des Testlaufs	59
Tabelle 5: gekürztes Abfrageergebnis des ausländischen WFS	60
Tabelle 6: Deskriptive Statistik Teil 1	65
Tabelle 7: Deskriptive Statistik Teil 2	66
Tabelle 8: Korrelationskoeffizienten der Variablen aus den Hypothesen	72
Tabelle 9: Regressionsanalyse mit Excel	78

Abkürzungsverzeichnis

BBox	Bounding Box
CSV	Comma-separated values
EPSG	European Petroleum Survey Group Geodesy
GIS	Geographische Informationssysteme
GML	Geography Markup Language
H_0	Nullhypothese
H_a	Alternativhypothese
HTTP	Hypertext Transfer Protocol
OGC	Open Geospatial Consortium
UML	Unified Modeling Language
URL	Uniform Resource Locator
WFS	Web Feature Service
XML	Extensible Markup Language

1 Einleitung

Die Digitalisierung und digitale Transformation sind mittlerweile unaufhaltsame Prozesse, die nicht mehr wegzudenken sind. Täglich werden digitale Prozesse eingesetzt, die das Leben erleichtern. Dabei ist die Unterscheidung in Digitalisierung und digitale Transformation wichtig. Digitalisierung bedeutet, dass das Analoge in Digitales übersetzt wird. Das Einscannen von Texten oder das Fotografieren einer Landkarte ist also Digitalisierung. Hierbei wird ein Blatt Papier in ein anderes Datenformat wie PDF oder JPEG übertragen. Die Form der Darreichung ändert sich also bei der Digitalisierung. Dadurch werden die Daten einfacher zugänglich und sie werden brauchbarer. Beispielsweise muss bei der Digitalisierung einer Landkarte der Beifahrer nicht mehr mit der Faltechnik von analogen Landkarten kämpfen.¹

Bei Transformationsprozessen wird der Gegenstand selbst geändert. Dabei geht es um die Schaffung eines neuen Produkts oder dass ein vorhandenes Produkt auf ein neues Level gehoben wird. Bei der Landkarte bedeutet das, dass neue Funktionen geschaffen werden. Das ist beispielsweise das Erkennen und Umfahren eines Staus oder einer Baustelle in Echtzeit oder die Suche nach dem günstigsten Benzin in der näheren Umgebung. Die Fahrtzeit wird ebenfalls live berechnet und ändert sich bei Pausen oder Geschwindigkeitsänderungen. Diese Funktionen können jedoch erst bei Vorhandensein einer digitalen Karte entwickelt werden. Bei analogen Landkarten musste eine Alternativroute manuell bestimmt werden und Staus zu umfahren war kaum möglich, da sie auf der Karte nicht markiert sind. Außerdem mussten früher jährlich neue Karten gekauft werden, um die Aktualität der Infrastruktur aufrechtzuerhalten. Dies kann je nach genutzten Dienst entfallen, da beispielsweise Google Maps kostenlos ist.²

Nach Müller und Scheffler gilt das deutsche Bau- und Planungswesen als wenig digitalisiert. Obwohl die Mehrheit der internen Akteure bereits IT-gestützt arbeitet, basiert die Kommunikation von Informationen und Daten oft noch auf Papier. Wenn IT-Systeme verwendet werden, kommt es zu Medienbrüchen und Insellösungen, die den Datenaustausch behindern. Aufgrund der niedrigen Finanzierungskosten für Bauvorhaben nimmt die Bautätigkeit seit Jahren kontinuierlich zu und die Nachfrage nach Wohnraum steigt. Zur Abarbeitung sind vernetzte IT-gestützte Prozesse und Verfahren unerlässlich. Als Grundlage dienen interoperable IT-Systeme und offizielle Datenaustauschstandards. Das sorgt dafür, dass die übermittelten Daten von allen Systemen entgegengenommen und gelesen werden

¹ Vgl. Friesike und Sprondel 2021, S. 10–11.

² Vgl. Friesike und Sprondel 2021, S. 12–13.

können. Da oft nicht die Potenziale, sondern nur die Kosten betrachtet werden, werden diese Standards noch nicht flächendeckend angewendet.³

Bei Bauvorhaben spielen Geodaten eine große Rolle. Das digitale Baugenehmigungsverfahren hat in den letzten Jahren zunehmend an Bedeutung gewonnen und bietet mehrere Vorteile gegenüber papierbasierten Verfahren. Die Geodaten stellen die Grundlage vieler Verwaltungsanwendungen dar. Durch die Standardisierungen und Regelungen wird eine einfachere Nutzung von digitalen Geodaten aus verschiedenen Quellen ermöglicht. Mittlerweile sind in der Form von webbasierten Geoinformationsdiensten viele Geodaten verfügbar. Diese Daten werden von den Informationssystemen der Behörden und von den Planungsverfahren für die elektronische Regierung und Verwaltung genutzt.⁴

Bei der Prüfung von Baugenehmigungsverfahren sind unter anderem Berechnungen notwendig, die Abstandsflächen und Maßzahlen der Baunutzungsverordnung darstellen. Es werden auch andere Angaben in Betracht gezogen, wie beispielsweise, ob die Baufläche ein Schutzgebiet schneidet. Die Prüfung dieser Angaben musste bisher per Hand durchgeführt werden.⁵

Das bedeutet, dass die Informationen manuell über eine Software mit grafischer Benutzeroberfläche von einem Mitarbeiter visualisiert wurden. Dadurch ist es möglich das Bauvorhaben mit den beispielsweise Schutzgebieten zu vergleichen und Schnittmengen zu sehen. Dabei müssen die Koordinaten vorher gesucht werden. Mithilfe dieser Arbeit soll dieser Prozess automatisiert werden. Außerdem sollen durch Messungen die webbasierten Geoinformationsdienste überprüft werden, ob sie zur Automatisierung geeignet sind. Durch die Automatisierung soll die Fehleranfälligkeit bei der Überprüfung der Koordinaten, beispielsweise durch Zeitdruck, gesenkt werden. Außerdem soll dadurch der Prozess deutlich beschleunigt werden. Durch die elektronische Übermittlung und Auswertung der Daten wird eine Verringerung der Durchführungszeit erwartet.⁶

Obwohl die Digitalisierung amtlicher Geodaten in den letzten zehn Jahren enorme Fortschritte gemacht hat und ihre Verfügbarkeit deutlich erhöht werden konnte, gibt es noch viele Herausforderungen für zukünftige Entwicklungen. Bisher gibt es nur wenige (durchgängige) E-Government-Anwendungen, die Geodaten oder Geoinformationsdienste direkt in Verwaltungsprozesse integrieren können. Es ist wichtig, die Fähigkeiten zu erweitern und

³ Vgl. Müller und Scheffler 2017, S. 99.

⁴ Vgl. Klenk et al. 2020, S. 101–102.

⁵ Vgl. Stember et al. 2021, S. 143.

⁶ Vgl. Streicher 2020, S. 43.

Elemente wie gemeinsame Lizenzierungsregeln, Sicherheitsstandards für sensible Daten und bedarfsgerechte, aber nicht zu komplexe Geodatenangebote zu entwickeln. Dieser Bedarf wurde an verschiedenen Stellen erkannt. Eine Studie der EU zur Umsetzung der INSPIRE-Richtlinie hat gezeigt, dass uneinheitliche Zugangsregeln und ein Mangel an Anwendungen, die INSPIRE nutzen können, die Hauptprobleme sind. Durch die Prozessautomatisierung dieser Arbeit sollen Geodaten für weitere Arbeit zur Verfügung stehen. Die Integration von Geowebsservices eröffnet neue Möglichkeiten zur Nutzung und Verarbeitung geografischer Informationen. Damit kann die Effizienz bürokratischer Prozesse gesteigert werden. Diese Arbeit liefert somit einen wertvollen Beitrag zur Forschung im Bereich der digitalen Transformation im Bauwesen.⁷

Die Umsetzung findet als Experiment statt. Dabei wird anhand von Schutzgebieten ein Web Feature Service (WFS) beispielhaft überprüft und die Ergebnisse ausgewertet. Neben der statistischen Sammlung von Daten steht auch die Prozessautomatisierung im Vordergrund. Die Anwendung beinhaltet beide Anwendungsfälle. Durch Konfiguration des Skripts soll die Nutzung angepasst werden. Im Laufe dieser Arbeit sollen die folgenden drei Forschungsfragen beantwortet werden.

Die Geodaten aus dem Dienst müssen für die Automatisierung und für die Mitarbeiter, die sie benötigen von Nutzen sein. Aus dieser Anforderung wird die erste Forschungsfrage abgeleitet: *Sind die geprüften WFS grundsätzlich für das Verfahren geeignet?* Damit soll statistisch überprüft werden, ob die Geowebsservices für die Automatisierung relevante Daten liefern. Diese Daten müssen für die weitere Verwendung geeignet sein. Es reicht nicht, dass die reine Erreichbarkeit des Dienstes geprüft wird, sondern das Messwerkzeug muss die erhaltenen Daten zum Teil und für weitere Anwendungen auswerten.

Für die Automatisierung ist es nötig, dass die Datenquelle zu jedem Zeitpunkt einer Abfrage Daten für die weitere Verwendung liefern kann. Daraus ergibt sich die zweite Forschungsfrage: *Sind die derzeit verfügbaren WFS technisch in der Lage, bei synchronen Aufrufen die für die RPA benötigten Inputdaten am Beispiel von Schutzgebietsdaten zu liefern?*

Die letzte Forschungsfrage beschäftigt sich mit der Einteilung in Arbeitszeit und Freizeit. Hierbei wird überprüft, ob die Anfragen an den Geowebsservice von den Anfragen der Behörden oder Ämtern beeinflusst wird. Dabei wird die benötigte Zeit der Abfrage ausgewertet und überprüft, ob die Zeit steigt oder sogar sinkt während der Arbeitszeit. Die dritte Forschungsfrage lautet demnach:

⁷ Vgl. Klenk et al. 2020, S. 109.

Gibt es einen signifikanten Unterschied in der Antwortzeit des WFS während der Kernarbeitszeit von 9-15 Uhr im Vergleich zur Freizeit (außerhalb der Kernarbeitszeit)?

Die Arbeit unterteilt sich in sieben Kapitel. Nach der Einleitung werden in Kapitel 2 die theoretischen Grundlagen genannt und erklärt, um den aktuellen Stand aufzuzeigen und die Fachwörter zu erläutern. Die nächsten vier Kapitel orientieren sich an der Planung eines Experiments nach Wohlin. Im dritten Kapitel findet die Scopingphase statt. Danach erfolgt in Kapitel 4 die Planningphase bei der das Messwerkzeug konzipiert wird. Im fünften Kapitel werden die Messungen und das Experiment durchgeführt. Außerdem werden alle Funktionen des Skripts gezeigt und erklärt. Nachdem die Daten gesammelt wurden, findet im sechsten Kapitel die Auswertung dieser über statistische Methoden statt. Zum Schluss werden die Erkenntnisse aus dem Experiment zusammengefasst und im Ausblick werden weitere Ideen für die Erweiterung aufgelistet.

2 Theoretischer Rahmen

Dieses Kapitel dient dazu eine grundlegende Basis zu schaffen wie die Systeme definiert werden. Es werden sechs Teilgebiete näher betrachtet, die im Laufe der Arbeit eine hohe Wichtigkeit aufweisen. Es ist wichtig ein einheitliches Verständnis aufzubauen, um gewisse Teile des Experiments nachvollziehen zu können. Es geht um Grundlagen aus der Informatik, Statistik und Planung von Experimenten, die im Laufe der Arbeit verwendet werden. Außerdem werden für das Experiment maßgebende Systeme vorgestellt, weil sie beispielsweise für die Datenerhebung wichtige Schnittstellen darstellen. Mit den wissenschaftlichen Grundlagen und der einheitlichen Definition von Fachtermini können weitere Kapitel darauf aufbauen.

2.1 Extensible Markup Language

Die eXtensible Markup Language (XML) ist ein flexibles Textformat. Ursprünglich für die Anforderungen des elektronischen Publizierens in umfangreichem Maßstab entwickelt, spielt XML eine zunehmend wichtige Rolle beim Austausch verschiedener Daten im Internet und anderen Bereichen. XML dient als Standard zur Modellierung von Daten durch eine Baumstruktur. Es besteht aus Regeln, die die Struktur von Dokumenten definieren und es ermöglichen, Daten gemäß dieser Struktur zu speichern. XML bietet einen Rahmen zur Definition von Datenstrukturen, ohne jedoch festzulegen, wie ein tatsächliches Dokument verarbeitet werden soll. Programme, die XML verarbeiten können, erfordern eine detaillierte Beschreibung der Elemente in den Dokumenten. Dies gilt insbesondere für die Definition von Strukturelementen und deren Anordnung innerhalb des Dokuments. Zu diesem Zweck stehen zwei Optionen zur Verfügung: Die ältere Methode ist die Document Type Definition (DTD), während XML Schema als modernere Möglichkeit betrachtet wird. XML ist also ein Standard zur Definition beliebiger Auszeichnungssprachen, die strukturell ähnlich sind.⁸

Die Extensible Markup Language besitzt nach Margit Becher zwei Hauptaufgabengebiete. Zum einen ermöglicht XML die layoutunabhängige Beschreibung von Dokumenten und zum anderen ist diese Sprache ein universelles Datenaustauschformat. Durch die Beschreibung von Dokumenten kann die Extensible Markup Language für Single-Source-Publishing genutzt werden. Hierbei werden Dokumente nicht einzeln per Hand erstellt, sondern aus einer Quelle automatisiert und auf die jeweiligen Bedürfnisse angepasst, generiert.⁹

⁸ Vgl. Shekhar 2008, S. 307.

⁹ Vgl. Becher 2021, S. 2.

Für die Zielsetzung dieser Arbeit ist jedoch die zweite Hauptfunktion von XML von entscheidender Bedeutung. XML dient als universelles Format für den Austausch von Daten. Dies bietet den Vorteil, dass der Datentransfer zwischen zwei oder mehr Anwendungen standardisiert ist. Dadurch wird die Kommunikation von Schnittstellen vereinfacht und sichergestellt, dass die übermittelten Daten korrekt interpretiert werden können. Ein weiterer Vorteil der Extensible Markup Language besteht darin, dass sie sowohl von Menschen als auch von Maschinen lesbar ist. Darüber hinaus ermöglicht XML die Abbildung komplexer und hierarchischer Strukturen. Aufgrund dieser Vorteile wird es von vielen Anwendungen und Schnittstellen unterstützt und hat sich daher als bevorzugtes Format für den Datenaustausch im Internet etabliert.¹⁰

Jedes XML-Dokument folgt einem bestimmten Aufbau. Am Anfang sollte der Prolog stehen. In diesem Teil befindet sich die XML-Deklaration, welche enthalten sein muss. Die Verarbeitungsanweisung und der Verweis auf DTD oder XML-Schemata sind optional im Prolog zu finden. Danach folgen die XML-Daten. Sie enthalten die grundlegenden Bestandteile in XML. Dies sind Attribute, Elemente und den Textinhalt. Der Textinhalt sind die Daten, die in einem Element gespeichert werden. Er befindet sich zwischen dem Start- und Endtag eines Elements und kann sowohl einfacher Text als auch strukturierte Daten sein. Der Textinhalt ermöglicht es, konkrete Informationen innerhalb eines Elements zu speichern und zu übertragen. Elemente repräsentieren die einzelnen Komponenten eines XML-Dokuments und können sowohl den Inhalt als auch die Hierarchie der Daten darstellen. Sie werden durch Start- und Endtags gekennzeichnet. Außerdem können sie einen Textinhalt aufweisen. Attribute hingegen dienen dazu, zusätzliche Informationen zu einem Element bereitzustellen, indem sie Eigenschaften oder Merkmale des Elements definieren. Dabei kann ein Element beliebig viele Attribute beinhalten. Attribute werden innerhalb des Starttags eines Elements angegeben und bestehen aus einem Namen-Wert-Paar. Damit wird die Darstellung und der Austausch von Daten in einer hierarchischen Baumstruktur ermöglicht.¹¹

XML Schema ist eine sprachliche Spezifikation, die zur Definition und Validierung der Struktur und des Inhalts von XML-Dokumenten verwendet wird. Es handelt sich um eine XML-basierte Sprache, die es ermöglicht, Regeln und Einschränkungen für die Elemente, Attribute und den Textinhalt eines XML-Dokuments festzulegen. XML Schema definiert komplexe Datentypen, ermöglicht die Festlegung von Beziehungen zwischen Elementen und erlaubt die Definition von Einschränkungen wie Mindest- und Höchstwerten, Mustern oder Wertebereichen. Jedes der Elemente und Attribute in XML-Schema ist mit einem spezifischen Typ verbunden. XML-

¹⁰ Vgl. Becher 2021, S. 6.

¹¹ Vgl. Becher 2021, S. 8–14.

Schema ermöglicht eine klare Trennung der Konzepte von Elementen und Attributen von ihren zugehörigen Typen. Diese Trennung erlaubt es, unterschiedliche Namen für Daten zu verwenden, die strukturell identisch sind. Neben der Verwendung unterschiedlicher Namen haben Entwickler auch die Freiheit, die entsprechenden Elemente an verschiedenen Stellen im XML-Dokument zu platzieren. Darüber hinaus ist es möglich, zwei Elementdeklarationen mit demselben Namen, aber unterschiedlichen Typen in verschiedenen Kontexten zu haben. Diese Flexibilität erlaubt es den Entwicklern, die Struktur und den Inhalt von XML-Dokumenten präzise anzupassen und verschiedene Kontexte zu berücksichtigen.¹²

2.1.1 Geography Markup Language

Im Bereich der Geodaten hat XML eine weite Verbreitung gefunden, wobei insbesondere die Geography Markup Language (GML) als XML-Sprache für Geodaten von großer Bedeutung ist. Ein Open Geospatial Consortium (OGC) konformer Dienst zur Übertragung von georeferenzierten Vektordaten sendet diese Informationen als XML-Datei im GML-Format zurück. GML ist eine Open-Source-Kodierung, die auf der Extensible Markup Language basiert und für die Darstellung geografischer Objekte geeignet ist. GML-Objekte sind unter anderem als eine Hierarchie von Merkmalen, Sammlungen und Geometrien organisiert und werden nach realen Entitäten modelliert, die durch Eigenschaften und Zustände gekennzeichnet sind. Darüber hinaus wurde GML als Informationsaustausch- und Speicherformat definiert, mit dem unterschiedliche Systeme gemeinsame geografische Daten nutzen können. GML-Schemata schaffen einen Standardentwurf dafür, wie geografische Objekte von einem System definiert und von anderen Systemen herstellerunabhängig verstanden werden können.¹³

GML dient als führende Modellierungssprache für die Darstellung von Daten mit geografischen Informationen über Standorte. Geodaten sind die äußerst bedeutsamen Informationen über die Koordinatenwerte der physischen Einheiten des Planeten, die in Form von Zahlenwerten in einem Geodatenystem dargestellt werden können. Die Geography Markup Language ist eine Standardsprache für die Speicherung und den Transport von Geodaten, um gemeinsame Richtlinien, Grundsätze, Rahmenrichtlinien und Standards für die Interoperabilität und Veränderbarkeit von Geodaten und -diensten zu fördern. GML arbeitet für Geographiesysteme und ist verantwortlich für das Austauschformat für geografische Datentransaktionen im Internet.¹⁴

¹² Vgl. Walmsley 2013, S. 18–22.

¹³ Vgl. Shekhar 2008, S. 364.

¹⁴ Vgl. Bajwa et al. 2020, S. 343–348.

Das XML-Dokument kann auch ohne umfangreiche Erläuterungen und tiefgreifende Kenntnisse von GML leicht verstanden werden. Jede Fläche ist eindeutig durch eine Flächennummer identifiziert und enthält Informationen über ihre Zugehörigkeit zu einer Verwaltungseinheit, ihren Besitzer und ihren Grenzen. Die geometrischen Eigenschaften eines Features werden durch das Element ‚geometryProperty‘ gekennzeichnet. Eine FeatureCollection ist eine Sammlung von einzelnen Feature-Elementen. Um die Geometriedaten korrekt darzustellen, muss das räumliche Bezugssystem spezifiziert werden, wobei die Abkürzung ‚srs‘ für spatialreferencesystem steht. Hierbei werden sogenannte EPSG-Codes verwendet.¹⁵

Zur Identifizierung der verschiedenen Koordinatensysteme werden EPSG-Codes eingesetzt. Diese wurden von der ursprünglichen European Petroleum Survey Group Geodesy (EPSG) zusammengestellt. Diese Gruppe wurde im Jahr 1986 gegründet. Im Jahr 2005 wurde die Gruppe durch das Surveying and Positioning Committee der International Association of Oil and Gas Producers ersetzt. Die Codes bestehen aus vier- bis fünfstelligen eindeutigen Schlüsselnummern für Koordinatenreferenzsysteme sowie Beschreibungen von Koordinatentransformationen. Das Schlüsselsystem ist äußerst detailliert und bietet verschiedene Codes für verschiedene Koordinatenreihenfolgen bei gleichem Gebiet.¹⁶

¹⁵ Vgl. Lange 2020, S. 249.

¹⁶ Vgl. Lange 2020, S. 177.

2.2 Deskriptive Statistik

Die deskriptive Statistik ist ein grundlegendes Instrument zur Analyse und Darstellung von Daten und eignet sich somit für diese Arbeit. Sie ermöglicht es, komplexe Datensätze zu strukturieren, zu beschreiben und zu interpretieren. Durch die Anwendung deskriptiver Statistik können wichtige Informationen über eine Stichprobe gewonnen werden. Nach Starbuck besteht dieses Teilgebiet der Statistik aus rudimentären Analysetechniken zur Unterstützung der Beschreibung einer Variable. Außerdem können mit der deskriptiven Statistik größere Datenmengen eines Parameters sinnvoll zusammengefasst werden. Dadurch ist die deskriptive Statistik hilfreich bei der Interpretation der verfügbaren Daten. Die deskriptive Statistik lässt sich in die univariate oder auch eindimensionale Analyse und die bivariate Analyse oder multivariate Analyse einteilen. Die bivariate Analyse behandelt zwei Variablen und deren Abhängigkeiten und die multivariate Analyse mehrere Variablen.¹⁷

2.2.1 Univariate Analyse

Mit der univariaten Deskription und Exploration lassen sich eindimensionale Daten in geeigneter Weise darstellen, beschreiben und analysieren. Eindimensionale Daten sind Daten aus Beobachtungen eines einzelnen Merkmals beziehungsweise einer einzelnen Variable. Dadurch wird jede Variable unabhängig untersucht. Starbuck unterscheidet innerhalb der univariaten Analyse zwischen den Maßen der zentralen Tendenz und der Maße für die Streuung. Bei der zentralen Tendenz wird die Lage der Häufigkeitsverteilung beziehungsweise die zentrale Position in einer Datenmenge einer gemessenen Variable beschrieben. Die gängigsten Maße der zentralen Tendenz werden im Folgenden aufgelistet und im Fall der späteren Verwendung beschrieben.¹⁸

Das erste und wohl bekannteste Maß ist der Durchschnitt oder das arithmetische Mittel. Es wird berechnet, indem die Summe aller Werte in der Datenreihe durch die Anzahl der Werte geteilt werden. Das arithmetische Mittel gibt an, wie die Werte im Durchschnitt verteilt sind und ermöglicht somit einen Vergleich zwischen verschiedenen Datenreihen. Ein weiteres Maß ist der Median. Dieser Wert wird verwendet, um den zentralen Wert einer Datenreihe zu bestimmen. Bei der Ermittlung des Medians werden die Daten in aufsteigender Reihenfolge sortiert und dann wird der mittlere Wert ausgewählt. Bei einer geraden Anzahl der Werte ist der Median der Durchschnitt der beiden mittleren Werte. Dadurch ist der Median weniger anfällig für Ausreißer im Gegensatz zum arithmetischen Mittel. Das dritte Maß ist der

¹⁷ Vgl. Starbuck 2023, S. 97.

¹⁸ Vgl. Fahrmeir et al. 2016, S. 29.

Modalwert oder auch Modus. Das Ergebnis ist die am häufigsten vorkommende Zahl in der zu untersuchenden Menge von Werten. Im Gegensatz zu den beiden vorangegangenen Maßen, fokussiert sich der Modalwert auf die Häufigkeit der Werte innerhalb einer Datenreihe. Der Modalwert kann bei diskreten und kontinuierlichen Daten eingesetzt werden und eignet sich besonders, um den dominierenden Wert oder die dominierende Kategorie zu identifizieren. Diskrete Daten können nur abzählbare Werte annehmen. Sie sind in der Regel ganzzahlig wie beispielsweise beim HTTP-Statuscode. Kontinuierliche Daten können in unendliche Zwischenwerte unterteilt sein. Die gemessenen Zeiten sind dementsprechend kontinuierliche Daten. Das letzte Maß ist die Spannweite. Sie gibt den Unterschied zwischen dem größten und kleinsten Wert in einer Zahlenreihe an. Dadurch wird Auskunft über die Streuung der Daten gegeben. Es wird eine Möglichkeit geschaffen die Bandbreite der beobachteten Werte schnell zu erfassen. Die Spannweite berücksichtigt jedoch nicht die Verteilung der Werte und ist daher stark beeinflussbar von Ausreißern.¹⁹

Bei dem Maß der Streuung wird der Fokus auf die Verteilung und Variabilität von Daten gelegt. Die verschiedenen Maße ermöglichen die Abweichung der einzelnen Datenpunkte vom Mittelwert zu bestimmen. Ein größeres Maß der Streuung deutet auf eine größere Variabilität der Daten hin, während ein kleineres Maß der Streuung auf eine geringere Variabilität hinweist. Bei einer hohen Variabilität sind die Datenpunkte weit voneinander entfernt. Dadurch werden Vergleiche zwischen verschiedenen Gruppen möglich. Ein Maß für die Variabilität der Daten ist die Varianz. Sie wird anhand des Durchschnitts der quadrierten Differenzen vom Mittelwert berechnet. Die Varianz beschreibt die Streuung oder die Abweichung der einzelnen Datenpunkte einer Datenreihe um den Durchschnittswert. Eine hohe Varianz deutet auf eine große Streuung der Daten hin. Die Standardabweichung ist ein weiteres Maß, welches ebenfalls die Streuung oder Abweichung einzelner Datenpunkte einer Datenreihe um den Durchschnittswert beschreibt. Im Gegensatz zur Varianz wird die Standardabweichung in derselben Einheit wie die Daten gemessen, was sie leichter interpretierbar macht. Eine hohe Standardabweichung deutet auf eine große Streuung der Daten hin. Die Standardabweichung wird berechnet, indem die Wurzel aus der Varianz gezogen wird.²⁰

Zum Schluss kommen die grafischen Darstellungen wie Histogramme und Box-Plots. Histogramme gruppieren und visualisieren die Daten. Dadurch können Muster, Trends und Ausreißer leichter erkannt werden. Außerdem ist damit eine schnelle Ermittlung der Form der

¹⁹ Vgl. Starbuck 2023, S. 97–101.

²⁰ Vgl. Fahrmeir et al. 2016, S. 64–66.

Verteilung (beispielsweise die Symmetrie oder Schiefe) möglich. Ebenso können die verschiedenen Klassen im Diagramm schnell miteinander verglichen werden.²¹

Box-Plots bieten eine kompakte Zusammenfassung der Verteilung von Daten. Diese Grafik ermöglicht eine schnelle Einschätzung wichtiger Maße wie Streuung und Symmetrie der Daten. Außerdem können damit schnell verschiedene Kategorien verglichen werden, indem mehrere Box-Plots nebeneinander platziert werden. Dadurch werden auf einem Blick Unterschiede bei Ausreißern, Mustern und Verteilungen ersichtlich. Durch den Vergleich univariater Ergebnisse können erste Hypothesen über mögliche Zusammenhänge gewonnen werden und somit den Weg für weiterführende Untersuchungen mit Methoden ebnen, die eine genauere Analyse von zwei- oder mehrdimensionalen Abhängigkeiten ermöglichen.²²

2.2.2 Multivariate Analyse

Bei vielen Auswertungen ist das Interesse nicht nur an einem einzigen, sondern auch an vielen erhobenen Merkmalen gegeben. Findet eine Beschränkung auf ein Merkmal statt, so lassen sich die Variablen wie Messzeiten, Arbeitszeiten oder Statuscodes mit den univariaten Maßen untersuchen. Jedoch können wesentliche Fragestellungen wie bei der dritten Hypothese, nämlich die Abhängigkeit zweier Variablen, nicht beantwortet werden. Dafür ist es notwendig die Daten gemeinsam zu erheben. Das bedeutet, dass eine Abfrage mehrere Werte enthält beziehungsweise die Werte der Abfrage zugeordnet werden können. Diese Daten werden mehrdimensionale Daten genannt.²³

Die Korrelation ist ein weiteres standardisiertes Maß, welches die Stärke und Richtung des linearen Zusammenhangs zwischen zwei Werten misst. Mit der Korrelation kann sowohl die Richtung als auch die Stärke bivariater Beziehungen quantifiziert werden. Die Korrelationskoeffizienten liegen im Wertebereich von minus eins und eins. Dieser Wert gibt an, wie eng die untersuchten Variablen miteinander verbunden sind. Wenn beide Variablen tendenziell zusammen variieren, ist der Korrelationskoeffizient positiv. Ein Wert von null deutet darauf hin, dass die beiden Variablen keine lineare Beziehung aufweisen. Die Korrelation ist deswegen ein wichtiger Bestandteil der Statistik und findet häufig bei der Analyse zwischen zwei oder mehreren Variablen Verwendung. Für Schlussfolgerungen aus den Ergebnissen ist jedoch zu beachten, dass die Korrelation nur den linearen Zusammenhang zwischen den untersuchten Variablen misst. Es wird dadurch keine Kausalität impliziert. Das bedeutet, dass ein hoher Korrelationskoeffizient nicht unbedingt aussagt, dass eine Variable die andere

²¹ Vgl. Fahrmeir et al. 2016, S. 38.

²² Vgl. Fahrmeir et al. 2016, S. 59–62.

²³ Vgl. Fahrmeir et al. 2016, S. 105.

verursacht. Es sind weitere Untersuchungen erforderlich, um eine kausale Beziehung zu bestätigen.²⁴ Die Kausalität ist jedoch kein Bestandteil dieser Arbeit.

Streudiagramme bilden den letzten Punkt für dieses Kapitel. Sie sind ein wichtiger Bestandteil für die Visualisierung von Beziehungen zwischen numerischen Variablen. Hierbei werden die Werte in einem Koordinatensystem als beispielsweise Punkte oder Kreuze dargestellt. Jeder Datenpunkt repräsentiert hierbei eine Beobachtung und zeigt damit den Wert der einen Variable auf der Abszisse und den Wert der anderen Variable auf der Ordinate an. Das Streudiagramm besteht aus einer horizontalen x-Achse und einer vertikalen y-Achse. Durch die Analyse des Musters der Datenpunkte im Streudiagramm können Informationen über den möglichen Zusammenhang zwischen den Variablen gewonnen werden. Es kann abgelesen werden, ob sie positiv, negativ oder unabhängig voneinander variieren.²⁵

²⁴ Vgl. Starbuck 2023, S. 113–115.

²⁵ Vgl. Fahrmeir et al. 2016, S. 121–122.

2.3 Web Application Architecture

Das moderne World Wide Web wurde nicht isoliert erfunden, sondern es basiert auf den Kernprotokollen des Internets, die bereits seit vielen Jahren existierten. Für die Entwicklung robuster Webanwendungen sind diese Protokolle ein wichtiger Faktor. Ein solches Protokoll ist beispielsweise Transmission Control Protocol/Internet Protocol (TCP/IP).²⁶

TCP/IP-Anwendungen folgen in der Regel dem Client-Server-Paradigma, das bedeutet, dass in diesen Anwendungen Server, auch als Dienste oder Daemons bezeichnet, auf Anfragen von Client-Programmen warten und diese bearbeiten. Die Client-Programme können entweder Anwendungen sein, die von Menschen genutzt werden, oder Server, die ihre eigenen spezifischen Anforderungen haben, die nur von anderen Servern erfüllt werden können. In den meisten Fällen befinden sich Client und Server auf separaten Rechnern und kommunizieren über eine Netzwerkverbindung. Die Kommunikation zwischen Client- und Serverprogrammen erfolgt unter Verwendung von Anfrage-Antwort-Interaktionen. Diese Interaktionen sind durch die Beschränkungen der Internet-Kommunikationsprotokolle definiert, was bedeutet, dass selbst einfache Befehlszeilenschnittstellen TCP/IP-Dienste nutzen können. Nachdem der Server die erfolgreiche Verbindung bestätigt hat, sendet der Client seine Befehle. Der Server antwortet dann auf jeden Befehl, normalerweise mit einer Zeile, die einen Antwortcode enthält. Ein zustandsbehaftetes Protokoll ermöglicht es einem Client, eine Sequenz von Befehlen zu unterstützen. Der Server muss den Zustand der Verbindung während der Übertragung aufeinanderfolgender Befehle aufrechterhalten, bis die Verbindung beendet wird. Diese Sequenz von übertragenen und ausgeführten Befehlen wird oft als Sitzung bezeichnet. Die meisten Internet-Dienste sind sitzungsorientiert.²⁷

Tim Berners-Lee hat drei grundlegende Komponenten entwickelt, die den Kern der Web-Technologie ausmachen. Zum einen eine Auszeichnungssprache für die Formatierung von Hypertext-Dokumenten mit dem Namen HyperText Markup Language (HTML). Außerdem entwickelte er ein einheitliches Notationsschema für die Adressierung von über das Netz zugänglichen Ressourcen. Das Schema wird als Uniform Resource Identifier (URI) oder auch als Uniform Resource Locator (URL) bezeichnet. Und die dritte Komponente ist das Protokoll für den Transport von Nachrichten über das Web. Dieses Protokoll trägt den Namen HyperText Transfer Protocol (HTTP).²⁸

²⁶ Vgl. Shklar und Rosen 2012, S. 28–29.

²⁷ Vgl. Shklar und Rosen 2012, S. 35–36.

²⁸ Vgl. Shklar und Rosen 2012, S. 59.

HTTP verfügt über mehrere Abfragemethoden. Es gibt GET, HEAD, POST, PUT, DELETE, TRACE, OPTIONS und CONNECT. Im Folgenden werden GET und POST erläutert, da diese beiden von besonderer Bedeutung für diese Arbeit sind. GET ist die einfachste der Anforderungsmethoden. Jeder Seitenaufruf verwendet diese Methode. Eine GET-Anfrage hat keinen Textkörper, muss jedoch eine Kopfzeile enthalten. Die Werte werden über die URL an den Server gegeben. Der Query-String wird durch ein Fragezeichen vom Rest der URL getrennt. POST-Anfragen haben dagegen einen Body. Dieser besteht aus dem Inhalt, der nach den Kopfzeilen folgt. POST-Anfragen schicken also Dateien an den Server.²⁹

HTTP schickt nach einer Anfrage in der ersten Zeile einen dreistelligen Statuscode zurück. Der Statuscode teilt einem HTTP-Client (Browser oder Proxy) entweder mit, dass die Antwort wie erwartet generiert wurde, oder dass der Client eine bestimmte Aktion durchführen muss. Die Codes sind in fünf Kategorien aufgeteilt. Die erste Ziffer bestimmt die Kategorie. Beginnt der Statuscode mit einer eins sind sie rein informativ. Bei einer zwei war die Anfrage erfolgreich. Eine drei weist den Client an zusätzliche Aktionen durchzuführen. Das kann beispielsweise eine Umleitung zu einer anderen Ressource sein. Bei Codes, die mit der Ziffer vier beginnen, trat ein Fehler bei der Client-Anfrage auf. Eine fünf deutet hingegen auf einen Serverfehler hin.³⁰

Bei Webanwendungen handelt es sich um Client-Server-Anwendungen, die einen Browser als Client-Programm verwenden. Webanwendungen sind jedoch nicht auf die Interaktionen zwischen Browsern und Servern beschränkt. Serverseitige Anwendungskomponenten können als Clients agieren und mit Backend-Datenquellen wie Datenbanken und Webdiensten interagieren. In mehrschichtigen Webanwendungen stellt jeder Satz benachbarter Schichten ein Paar aus einem Client und einem Server dar, und jede Zwischenschicht kann entweder als Client oder als Server fungieren, je nachdem, mit welchem ihrer Nachbarn sie gerade interagiert. So wie sich beispielsweise der Browser mit dem Webserver verbindet, um eine Anfrage zu stellen, kann sich die auf dem Webserver ausgeführte Anwendungskomponente mit einer Geschäftslogikschicht verbinden und als Client für die Dienste dieser Schicht fungieren.³¹

Die Web Application Architecture (WAA) weist drei verschiedene Schichten auf. Die Schichten sollten als konzeptionelle Grenzen betrachtet werden, müssen aber nicht physisch voneinander isoliert sein, beispielsweise durch eine virtuelle Maschine. Bei einer

²⁹ Vgl. Shklar und Rosen 2012, S. 71.

³⁰ Vgl. Shklar und Rosen 2012, S. 74.

³¹ Vgl. Shklar und Rosen 2012, S. 394.

Webanwendung werden die Schichten in der Regel in derselben virtuellen Maschine ausgeführt. Die erste der drei Schichten ist die Präsentationsschicht (Presentation Layer). Sie ist für die Darstellung der Benutzeroberfläche verantwortlich. Diese Schicht ist meistens webbasiert. Diese Schicht umfasst HTML, CSS und JavaScript und soll so dünn wie möglich sein. Der Benutzer interagiert mit dieser Schicht. Die Präsentationsschicht soll auf einer gut konzipierten Anwendungsschicht aufbauen. Die Anwendungsschicht (Application Layer) ist das System, das die Geschäftslogik enthält. In dieser Schicht werden die Anfragen verarbeitet. Sie bildet die Transaktionsgrenze des Systems und soll so wenig wie möglich über Persistenz, also den Zustand von Objekten längere Zeit zu speichern, wissen. Die Anwendungsschicht kommuniziert mit der Präsentationsschicht. Die dritte Schicht ist die Datenbankschicht (Data Layer). Sie speichert und verwaltet die Daten. Dazu werden Datenbanken verwendet. Diese Schicht darf keine Geschäftslogik enthalten. Diese drei Schichten kommunizieren von oben (Präsentation) nach unten (Datenbank). Durch die Zusammenarbeit wird eine vollständige Webanwendung erstellt. Die Aufteilung in Schichten ist die Webanwendung modular. Dadurch werden eine einfachere Wartung und Erweiterung sichergestellt.³²

³² Vgl. Deinum und Cosmina 2021, S. 55–59.

2.4 Geowebsservices

Webdienste sind Softwareanwendungen, die über das Internet verfügbar sind und das standardisierte XML-Nachrichtenprotokoll als Kommunikationssystem verwenden. Ein solches Protokoll ist beispielsweise Simple Object Access Protocol (SOAP) oder Representational State Transfer (REST). Durch die Nutzung nicht-proprietärer Standards fördern Webdienste die Interoperabilität, weil Maschinen unabhängig von ihrem Betriebssystem, ihrer Programmiersprache, Entwicklungsumgebungen oder ihrem Standort die übermittelten Informationen verstehen können. Die grundlegende Idee eines Webdienstes besteht darin, dass Maschinen Informationen über die Art der von ihnen angebotenen Dienste und ihnen übergebene Parameter veröffentlichen können. Ein Webdienst basiert auf einer dienstorientierten Architektur, die den Prinzipien der Serviceorientierten Architektur (SOA) folgt. Dabei gibt es drei grundlegende Operationen. Diese sind veröffentlichen, verknüpfen und verbinden. Die Dienstanbieter veröffentlichen ihre Dienste bei einem Dienstvermittler, während Dienstanforderer den Dienst finden und sich mit diesem verbinden. Dieses Konzept lässt sich ebenfalls auf geographische Informationen anwenden.³³

Geodatendienste repräsentieren eine spezifischere Ausprägung von Webdiensten oder Webservices, die geographische Informationen und Funktionen bereitstellen. Durch den Einsatz von Geowebsservices wird der Zugriff auf geographische Daten wie Karten, Geodatenbanken oder Kodierungen über das Internet ermöglicht. Diese Dienste nutzen häufig standardisierte Protokolle wie den Web Map Service (WMS), den Web Coverage Service (WCS) oder insbesondere für diese Arbeit relevant, den Web Feature Service (WFS).³⁴

2.4.1 Web Feature Service

Der Web Feature Service ist eine Schnittstelle, die vom Open Geospatial Consortium entwickelt wurde, um den Austausch von geografischen Daten über das Internet zu ermöglichen. Diese Schnittstelle definiert die Regeln und Protokolle für die Abfrage und den Abruf von geografischen Informationen über das Hyper Text Transmission Protocol. Der Web Feature Service ermöglicht es, dass geografische Merkmale abgefragt und manipuliert werden können. Zur Übertragung der Informationen wird GML verwendet. WFS unterstützt das Vektordatenmodell. Das OGC spielte eine entscheidende Rolle bei der Entwicklung des Web Feature Service. Es ist ein internationales Industriekonsortium, bestehend aus Unternehmen, Behörden und Universitäten, das durch einen Konsensprozess öffentlich verfügbare

³³ Vgl. Shekhar 2008, S. 1256.

³⁴ Vgl. Lange 2020, S. 250.

Schnittstellenspezifikationen entwickelt. Im Bereich der Webdienste hat das OGC spezifische Webschnittstellen definiert, die auf einem Modell basieren, welches Anfrage- und Antwortregeln unter Verwendung des HTTP und XML unterstützt. Die Spezifikation für den Web Feature Service wurde erstmals im Mai 2002 eingeführt, als das OGC die Version 1.0 des WFS veröffentlichte.³⁵

Der Web Feature Service ist eine Spezifikation, die es einer Maschine (z. B. Maschine A) ermöglicht, eine andere Maschine (z. B. Maschine B) nach geographischen Informationen zu fragen. Der Grund für die Informationsanfrage besteht darin, dass Maschine A selbst nicht über diese Informationen verfügt und daher auf Maschine B angewiesen ist, um sie zu erhalten. Sobald Maschine A die benötigten Informationen erhalten hat, kann sie diese entsprechend ihren Anforderungen nutzen. Dies wird durch die Verwendung von XML-basiertem GML ermöglicht. Durch diese Kodierung sind die Informationen maschinenunabhängig. Geographische Daten werden normalerweise zur Erstellung einer Karte oder zum Hinzufügen zusätzlicher Ebenen zu einer bereits vorhandenen Karte verwendet. Dies ist die gängigste Art und Weise, wie die Dienste eines Web Feature Servers genutzt werden können. Es gibt jedoch auch andere Möglichkeiten, den WFS zu nutzen. Eine weitere Verwendungsmöglichkeit ist die spezifische Abfrage des Dienstes auf der Grundlage von räumlichen und nicht-räumlichen Beschränkungen. Nicht-räumliche Beschränkungen sind beispielsweise Jahreszahlen, Kategorien oder Namen.³⁶

Ein WFS, der OGC-konform ist, besitzt drei Funktionen. Diese Operationen können beispielsweise über die URL per HTTP-Anfrage an den Service geschickt werden.

1. GetCapabilities: Durch diese Abfrage werden die Fähigkeiten des WFS beschrieben. Standardmäßig wird eine XML-Datei zurückgesendet in der Metainformationen enthalten sind. Darin befinden sich zum Beispiel die Feature Types, Angaben zum Datenanbieter, Typenames und die möglichen Operationen.
2. DescribeFeatureType: Mit dieser Operation wird die Struktur jedes einzelnen Feature Types der Geoobjekte beschrieben.
3. GetFeature: Diese Funktion liefert Vektordaten, die mit den Merkmalen aus der Abfrage verbunden sind, im GML-Format zurück. Mit dieser Operation können Geoobjekte zur Bearbeitung gesperrt, geändert, erzeugt oder gelöscht werden. Jedoch werden in dieser Arbeit ausschließlich Objekte abgefragt und nicht manipuliert.³⁷

³⁵ Vgl. Shekhar et al. 2017, S. 2481.

³⁶ Vgl. Shekhar 2008, S. 1256–1257.

³⁷ Vgl. Lange 2020, S. 255.

Für eine GetFeature-Abfrage ist es erforderlich, einen Typnamen anzugeben. Dies ist der Name eines bestimmten Feature-Typs, den der spezifische WFS anbietet. Durch den Feature-Typ werden die Struktur und Attribute eines geographischen Features definiert. Dadurch kann dieser spezielle Feature-Typ bei einer GetFeature-Abfrage abgerufen oder bearbeitet werden. Durch den Typenname kann ein spezifisches Feature identifiziert und angesprochen werden. Ein Feature-Typ kann noch weitere Elemente außer den Namen aufweisen. Das beinhaltet zum Beispiel einen Titel, der für Menschen besser lesbar ist, Schlüsselwörter für eine einfachere Suche oder die Bounding Box, die angibt wo Instanzen eines bestimmten Merkmalstyps vorhanden sind.³⁸

2.4.2 Infrastructure for Spatial Information in Europe

Die Entwicklung von Geoportalen führte zu einer erleichterten Zugänglichkeit räumlicher Informationen. Davor wurden Karten per Hand von Kartographen erstellt, doch mittlerweile sind sie digitalisiert worden. Diese Geoinformationen finden in einer Vielzahl von Lebensbereichen Anwendung. Innerhalb der EU existieren zahlreiche Datenbanken, die je nach Fachgebiet auf verschiedene Institutionen verteilt sind. Die umfassende Sammlung räumlicher Informationen über eine spezifische Region ist sehr zeitaufwändig und der Vergleich von Daten zwischen verschiedenen EU-Ländern gestaltet sich als kompliziert, da bisher keine Harmonisierung stattgefunden hat. Um diese Situation zu verbessern, verabschiedete die Europäische Kommission im Jahr 2007 eine Richtlinie mit dem Ziel, eine einheitliche Geodateninfrastruktur in der Europäischen Gemeinschaft zu etablieren. Diese Richtlinie trägt den Namen Infrastructure for Spatial Information in Europe (INSPIRE). Das Hauptziel von INSPIRE ist es, den Zugang zu Daten zu erleichtern und zu beschleunigen sowie die Möglichkeiten des Datenaustauschs innerhalb der EU-Länder zu verbessern.³⁹

Die Implementierung der INSPIRE-Richtlinie erfordert die Überprüfung der Konformität einer großen Anzahl von Netzwerkdiensten mit den Durchführungsbestimmungen von INSPIRE. Die Bewertung, ob ein Dienst vollständig mit INSPIRE konform ist, ist eine komplexe Aufgabe und erfordert den Einsatz spezialisierter Prüfwerkzeuge. Es gibt mit dem INSPIRE-Register eine zentrale Datenbank⁴⁰, die Informationen über INSPIRE-konforme Dienste enthält.⁴¹

³⁸ Vgl. Vretanos 2010, S. 46–48.

³⁹ Vgl. Ogryzek et al. 2020, S. 1–3.

⁴⁰ <https://inspire.ec.europa.eu/registry/>

⁴¹ Vgl. Huerta et al. 2014, S. 325–327.

2.5 Geographic Information Systems

Geographische Informationen stellen ein Wissensgebiet dar, welches durch die Verarbeitung geographisch referenzierter Daten gewonnen wird. Die geographischen Informationssysteme (GIS) sind Funktionen, die von einer Softwareeinheit über ihre Schnittstellen zur Verfügung gestellt werden. Außerdem bieten sie die Bereitstellung von Informationen, die aus Geodaten generiert werden. Die Entwicklung von Geoinformationsdiensten ist stark mit der Technologie verteilter Systeme und dem Internet verknüpft. Sie stellt einen bedeutenden Fortschritt dar, um einen einfachen Zugang zu geografischen Informationen und zur Geoverarbeitungstechnologie zu ermöglichen. Aus der Perspektive der Softwareentwicklung repräsentieren Geoinformationsdienste eine vertikale Domäne von Informationsdiensten, in der verschiedene IT-Dienstkomponenten zu Informationsanwendungen kombiniert werden. GI-Service-Komponenten sind eigenständige, selbstbeschreibende und wiederverwendbare Softwareobjekte, die in mehreren Adressräumen und verteilten Umgebungen veröffentlicht, lokalisiert und aufgerufen werden können. Diese Komponenten können unterschiedliche Größen haben und in verschiedenen Frameworks wie RPC, .NET und Webservices arbeiten.⁴²

Nach Lü et al. gibt es viele Definitionen von geographischen Informationen und den Systemen, die verwendet werden, um Daten zu speichern, abzurufen, zu analysieren und darzustellen, die räumlich oder geographisch dargestellt werden. Sie definieren GIS als Informationssystem, das geographische, räumliche oder georäumliche Daten für den raumzeitlichen Gebrauch und die geographische Forschung verarbeitet. GIS unterstützen von der computergestützten Kartierung über die räumliche Analyse bis hin zur geographischen Problemlösung. Die geographischen Informationssysteme haben drei Hauptziele. Sie umfassen die Erfassung geographischer Informationen, die Untersuchung und Visualisierung von Geobjekten und deren Beziehungen sowie die Erforschung geographischer Regeln.⁴³

In der Welt der geographischen Informationssysteme werden Daten üblicherweise in einer von zwei grundlegenden Datenstrukturen betrachtet: Raster und Vektoren. Rasterdaten sind Daten, die das Fernerkundungsvokabular von Pixeln, räumlicher Auflösung, Bildern und Bildsammlungen verwenden. Durch diese Darstellung des Geländes können Analysen von räumlichen Mustern durchgeführt werden. Jedoch ist die Arbeit mit Vektordaten ebenfalls ein wichtiger Bereich. Diese Daten repräsentieren geographische Informationen als Punkte, Linien oder Polygone. Dadurch wird eine präzise Darstellung von geographischen Objekten ermöglicht. Vektordaten eignen sich dementsprechend für die Visualisierung von Standorten

⁴² Vgl. Shekhar 2008, S. 578.

⁴³ Vgl. Lü et al. 2019, S. 346–348.

beziehungsweise Grenzen und sind somit auch für diese Arbeit wichtig. Sowohl Vektor- als auch Rasterdaten sind wertvolle Werkzeuge in der GIS-Analyse und bieten unterschiedliche Stärken und Anwendungsbereiche. Ein Punkt kann ein spezifischer Standort sein. Mithilfe von zwei Punkten kann eine Bounding Box generiert werden, die die Form eines Rechtecks oder Quadrats annimmt. Linien werden können ebenfalls aus zwei oder mehr Punkten generiert werden. Dadurch können beispielsweise Bahnstrecken oder Straßen abgebildet werden. Polygone bestehen aus vielen Punkten und haben eine Fläche. Damit eignen sie sich zur Darstellung von Grenzen oder Gebieten. Jeder Punkt hat eine eindeutige Position auf der Erdoberfläche.⁴⁴

GIS ist ein übergeordneter Begriff für die spezifischen Softwarelösungen. Eine dieser Lösungen ist Quantum GIS (QGIS). Dies ist ein in der Programmiersprache C++ geschriebenes Open-Source GIS. Das bedeutet, dass der Quellcode für die Öffentlichkeit frei zugänglich ist. Es läuft auf den meisten existierenden Betriebssystemen, einschließlich Linux, Unix, Mac OS X und Windows. Das Hauptaugenmerk von QGIS liegt auf der interaktiven zweidimensionalen Darstellung von räumlichen Daten. Es gibt jedoch auch Funktionen zur Bearbeitung von Vektordaten und ein GRASS-Plugin, mit dem die analytischen Funktionen des GRASS-Programms innerhalb der QGIS-GUI genutzt werden können. QGIS unterstützt eine große Anzahl von Vektor- und Rasterformaten, einschließlich PostGIS, GRASS, Shapefile, GML, WFS, GPX, WMS, GeoTiff, PNG, JPG und viele andere.⁴⁵ Dadurch eignet es sich für die Aufgaben, die bei dieser Arbeit bearbeitet werden.

⁴⁴ Vgl. Cardille et al. 2023, S. 424.

⁴⁵ Vgl. Shekhar 2008, S. 935.

2.6 Experimentieren nach Wohlin

Für die wissenschaftliche Durchführung eines Experiments ist es notwendig die Schritte und Ergebnisse zu dokumentieren. Nach Wohlin gibt es für wissenschaftliche Versuche fünf wichtige Phasen, die im Folgenden genannt und erklärt werden. Der erste wichtige Punkt ist das Scoping. Hierbei wird der Umfang des Experiments definiert. Diese Phase ist wichtig, da sie das Fundament des Experiments darstellt und bei fehlerhaften oder schlechten Vorbereitungen im späteren Experimentverlauf Probleme auftreten. Das kann zu Verzögerungen führen. Im schlimmsten Fall sind die aus dem Experiment gewonnenen Ergebnisse beziehungsweise Datensätze nicht weiter verwendungsfähig. Durch die Scopingphase sollen die Ziele des Experiments anhand eines definierten Rahmens festgelegt werden. Wohlin bietet hierfür eine Vorlage an mit der es einfacher sein soll den Kontext zu bestimmen. Dieses Template enthält fünf Punkte. Der erste ist der Untersuchungsgegenstand. Dies kann ein Produkt, Prozess, Modell oder eine Theorie sein. Als nächstes folgt der Zweck, der aufzeigt warum das Experiment durchgeführt wird. Der Qualitätsschwerpunkt ist der primäre Effekt des Experiments in der Untersuchung. Die Perspektive gibt die Sichtweise der Interpretation an. Der Kontext definiert Subjekte und Objekte, die an dem Experiment beteiligt sind. Daraus ergibt sich dann der Zieldefinitionsrahmen.⁴⁶

Nach der abgeschlossenen Scopingphase findet die Planung statt. Beim Scoping wird die Grundlage des Experiments bestimmt. Die Planung bereitet die Durchführung vor. Die Pläne sind wichtig für die Kontrolle des Experiments. Nach Wohlin gibt es bei der Planung sieben Schritte. Diese werden im Folgenden aufgezählt und kurz erläutert. Ausgehend von der Zieldefinition aus der Scopingphase wird in der Kontextauswahl die Umgebung ausgewählt, in der das Experiment durchgeführt werden soll. Anschließend erfolgt die Hypothesenformulierung und die Variablenauswahl der unabhängigen und abhängigen Variablen. Die Auswahl der Versuchspersonen wird ebenfalls durchgeführt. Die Art der Versuchsplanung wird auf der Grundlage der ausgewählten Hypothese und Variablen gewählt. Anschließend wird die Instrumentierung für die praktische Durchführung des Experiments vorbereitet. Schließlich zielt die Auswertung der Validität darauf ab, die Gültigkeit des Experiments zu überprüfen. Der Planungsprozess wird so lange wiederholt, bis ein vollständiger Versuchsplan fertig ist. Da für dieses Experiment nicht alle Schritte benötigt werden, finden sich in Kapitel 4 nur die verwendeten Schritte wieder.⁴⁷

⁴⁶ Vgl. Wohlin et al. 2012, S. 85–88.

⁴⁷ Vgl. Wohlin et al. 2012, S. 89.

Nach Abschluss des Entwurfs und der Planung kann das Experiment zur Datenerhebung durchgeführt werden. Aus diesem Grund ist das die Phase der Durchführung des Experiments. In dieser operativen Phase findet die Interaktion mit Versuchspersonen und die Sammlung der Daten statt. Die Durchführungsphase wird in drei Schritte eingeteilt. Der erste Schritt ist die Vorbereitung bei der Versuchspersonen ausgewählt und Formulare vorbereitet werden. Als nächstes folgt die tatsächliche Durchführung des Experiments mit der Sammlung der Daten. Und zum Schluss erfolgt die Datenvalidierung, bei der die gesammelten Daten überprüft werden. Da das Experiment dieser Arbeit ausschließlich technologieorientiert ist, wird dieses Kapitel sich nicht mit Versuchspersonen beschäftigen, sondern sich rein auf das Messwerkzeug und die Daten fokussieren.⁴⁸

Nachdem die Daten erhoben wurden, kann die Phase der Analyse und Interpretation starten. Auf der Grundlage der gewonnenen Daten können dann Schlussfolgerungen gezogen werden. Um dies zu erreichen, müssen die Versuchsdaten richtig interpretiert werden. Die quantitative Anforderung kann in drei Schritten erfolgen. Im ersten Schritt werden die gewonnenen Daten mit Hilfe der deskriptiven Statistik charakterisiert und visualisiert. Im zweiten Schritt werden anormale oder falsche Datenpunkte ausgeschlossen. Im letzten Schritt werden die Daten durch Hypothesentests analysiert.⁴⁹

Nach Abschluss eines Experiments erfolgt die letzte Phase. Sie wird als Präsentation oder Bericht über das Experiment bezeichnet. Dabei ist es wichtig die Präsentation der Ergebnisse zielgruppengerecht zu gestalten. Dies kann beispielsweise in Form eines Artikels, eines Berichts oder Pakets für die Nachbildung des Experiments geschehen. Wird das Experiment von oder mit einem Unternehmen durchgeführt, ist es angebracht die verschiedenen Erfahrungen in einer Erfahrungsdatenbank für Mitarbeiter zu speichern.⁵⁰

⁴⁸ Vgl. Wohlin et al. 2012, S. 117.

⁴⁹ Vgl. Wohlin et al. 2012, S. 123.

⁵⁰ Vgl. Wohlin et al. 2012, S. 153.

3 Zieldefinitionsphase

In dieser Phase des Experiments wird der Umfang durch die Festlegung der Ziele bestimmt. Dies wird durch eine Zieldefinitionsvorlage erreicht, die vor dem Start der nächsten Phase festgelegt werden muss. Der Zweck einer Zieldefinitionsvorlage ist es, sicherzustellen, dass wichtige Aspekte eines Experiments definiert sind bevor die Planung und Durchführung erfolgt. Dadurch werden Fehlerquellen reduziert oder ausgeschlossen. Durch die Definition des Ziels des Experiments nach dieser Vorlage wird die Grundlage dafür gelegt.⁵¹

Die Vorlage enthält fünf wichtige Komponenten, die in Tabelle 1 als Spalten aufgeführt werden. Der erste Punkt ist der Untersuchungsgegenstand des Experiments. Der Forschungsgegenstand kann aus Produkten, Prozessen, Ressourcen, Modellen, Messgrößen oder Theorien bestehen. Beispiele sind das Endprodukt, der Entwicklungs- oder Prüfprozess oder ein Modell zur Steigerung der Zuverlässigkeit. Der Zweck legt fest, was mit dem Experiment beabsichtigt ist. Es kann darum gehen, die Auswirkungen von zwei verschiedenen Techniken zu bewerten oder die Lernkurve einer Organisation zu beschreiben. Der Qualitätsschwerpunkt ist der primäre Effekt, der in dem Experiment untersucht wird. Der Qualitätsschwerpunkt kann Wirksamkeit, Kosten, Zuverlässigkeit oder etwas anderes sein. Die Perspektive gibt an, aus welchem Blickwinkel die Versuchsergebnisse interpretiert werden. Beispiele für Perspektiven sind Entwickler, Projektmanager, Kunden und Forscher. Der Kontext ist die Umgebung, in der das Experiment durchgeführt wird. Der Kontext definiert kurz, welches Personal an dem Experiment beteiligt ist (Subjekte) und welche Software-Artefakte (Objekte) in dem Experiment verwendet werden. Probanden können durch Erfahrung, Teamgröße, Arbeitsbelastung usw. charakterisiert werden. Objekte können durch Größe, Komplexität, Priorität, Anwendungsbereich usw. charakterisiert werden. In Tabelle 1 sind weitere Beispiele von Wohlin et al. aufgeführt, die bei der Zielsetzung helfen können. Mithilfe dieser Auswahl kann eine Zieldefinition ohne viel Aufwand erstellt werden.⁵²

⁵¹ Vgl. Wohlin et al. 2012, S. 85.

⁵² Vgl. Wohlin et al. 2012, S. 85–87.

Tabelle 1: Vorlagen für die Zieldefinition

Gegenstand der Studie	Zielsetzung/Zweck	Qualitätsfokus	Perspektive	Kontext
Produkt	Charakterisierung	Wirksamkeit	Entwickler	Subjekte
Prozess	Überwachung	Kosten	Betreuer	Objekte
Modell	Auswertung	Verlässlichkeit	Projektleiter	
Metrik	Vorhersagung	Instandhaltbarkeit	Unternehmensleiter	
Theorie	Steuerung	Übertragbarkeit	Kunde	
	Änderung		Benutzer	
			Forscher	

Quelle: In Anlehnung an Wohlin et al., 2012, S. 87

Es gibt die Unterscheidung in Experiment und Quasi-Experiment. Bei einem Quasi-Experiment fehlt die Randomisierung der Probanden oder Objekte. Ein-Objekt-Studien werden mit einem einzigen Subjekt und einem einzigen Objekt durchgeführt. Die Studie mit nur einem Objekt und Subjekt ist ein Quasi-Experiment, wenn das einzelne Subjekt und das Objekt nicht nach dem Zufallsprinzip ausgewählt werden, aber sie ist ein Experiment, wenn das Subjekt und das Objekt nach dem Zufallsprinzip ausgewählt werden. Die Randomisierung ist deswegen ein wichtiger Faktor.⁵³

Mithilfe der allgemeinen Zieldefinitionsvorlage kann eine Zieldefinition für das durchzuführende Experiment erstellt werden. Aufgabenstellung ist es einen Web Feature Service mit einem vorher an Geowebdienste angepassten und implementierten Messwerkzeug für den Zeitraum von sieben Tagen, also einer Woche abzufragen. Die zurückerhaltenen Daten müssen nach der Abfrage für anschließende Evaluierungen zur Verfügung stehen. Der qualitative Schwerpunkt liegt auf der Verlässlichkeit des Services. Mit den gespeicherten Datensätzen kann dann festgestellt werden, ob die Services dauerhaft für den Gebrauch nutzbare Ergebnisse liefern oder nicht. Das bedeutet, dass nicht nur die Verbindung zum Server getestet wird, sondern reelle Abfragen. Die Perspektive ist aus der Sicht des Experimentators. Eine Sicht des Softwareentwicklers ist beim Start und Ablauf der Messung jedoch auch von Vorteil. Das Experiment wird mit einem ausgewählten Web Feature Service und Server durchgeführt. In Deutschland verfügt jedes Bundesland über einen eigenen Web Feature Service für sein jeweiliges Gebiet. Dies impliziert, dass bei Verwendung des WFS für Sachsen-Anhalt ausschließlich geographische Daten bereitgestellt werden können, die sich auf Sachsen-Anhalt beziehen. Da die Koordinaten deutschlandweit generiert werden, wird ein

⁵³ Vgl. Wohlin et al. 2012, S. 86.

deutschlandweiter WFS benötigt. Das Bundesamt für Naturschutz (BfN) bietet einen Web Feature Service an, der Naturschutzgebiete über ganz Deutschland enthält. Neben dem WFS⁵⁴ wird auch ein Dienst im Browser zur Visualisierung von den im WFS enthaltenen Gebieten angeboten. Mit der interaktiven Karte des BfN⁵⁵ lassen sich die jeweiligen Grenzen und Naturschutzgebiete visualisieren. Es ist ebenfalls möglich in der Karte zu zeichnen und den EPSG-Wert zu ändern. Es werden außerdem eine Legende und weitere Informationen für die jeweiligen Schutzgebiete angeboten.

Nachdem alles definiert wurde, ergibt sich folgender Zieldefinitionsrahmen für das Experiment dieser Arbeit:

Analysieren des Prozesses der Abfrage eines Web Feature Service im Hinblick der Evaluation in Bezug auf ihre Verlässlichkeit aus dem Blickwinkel eines Experimentators im Kontext von einem Geowebsevice. Damit ist das Ziel des Experiments definiert und die nächste Phase kann bearbeitet werden.

⁵⁴ <https://geodienste.bfn.de/ogc/wfs/schutzgebiet>

⁵⁵ <https://geodienste.bfn.de/schutzgebiete>

4 Planungsphase

Nachdem das Ziel definiert und der Kontext abgegrenzt wurde, kann die Planungsphase stattfinden. In dieser Phase wird geklärt, wieso das Experiment durchgeführt wird. Außerdem werden bei der Planung erste Diagramme erstellt, die bei der Implementierung als Grundlage und Hilfe dienen. Eine häufige Situation in einem Experiment ist, dass ein bestehender Prozess mit etwas Neuem verglichen wird. Dadurch kann ein Vergleich gezogen werden, ob die neue oder alte Methode besser ist. Ein Problem ist, dass die alte Methode selten ausreichend dokumentiert ist und nicht einheitlich angewandt wird.⁵⁶

Experimente können menschenorientiert oder technologieorientiert sein. Für diese Arbeit wird ein technologieorientiertes Experiment durchgeführt. Der Fokus liegt auf dem Antwortverhalten des WFS. Bei technologieorientierten Experimenten werden in der Regel technische Werkzeuge auf verschiedene Objekte angewandt. Das menschenorientierte Experiment hat weniger Kontrolle als das technologieorientierte, da Menschen sich bei verschiedenen Gelegenheiten unterschiedlich verhalten, während Werkzeuge in den meisten Fällen deterministisch sind.⁵⁷

4.1 Formulierung der Hypothesen

Die Grundlage für die statistische Analyse eines Experiments ist die Hypothesenprüfung. In der Planungsphase wird die Definition des Experiments in Hypothesen formalisiert. Eine Hypothese wird formell aufgestellt, und die im Laufe des Experiments gesammelten Daten werden verwendet, um die Hypothese nach Möglichkeit zu verwerfen. Wenn die Hypothese verworfen werden kann, können auf der Grundlage der Hypothesenprüfung unter Berücksichtigung der Risiken Schlussfolgerungen gezogen werden. In der Planungsphase wird die Versuchsdefinition in Hypothesen formalisiert. Es müssen je zwei Hypothesen formuliert werden. Eine Nullhypothese (H_0), die besagt, dass es keine zugrunde liegenden Trends oder Muster in der Versuchsanordnung gibt. Diese Hypothese soll mit möglichst hoher Signifikanz zurückgewiesen werden. Im Gegensatz dazu ist die Alternativhypothese (H_a) die Hypothese, für die die Nullhypothese verworfen wird. Die Nullhypothese ist so formuliert, dass es keinen signifikanten Unterschied zwischen den Gruppen und Bedingungen gibt. Die

⁵⁶ Vgl. Wohlin et al. 2012, S. 89–90.

⁵⁷ Vgl. Wohlin et al. 2012, S. 16.

Alternativhypothese wird gegenteilig formuliert. Das heißt, dass es einen signifikanten Unterschied gibt.⁵⁸

Für die Arbeit sind drei Hypothesen von Bedeutung. Die erste Nullhypothese bezieht sich auf die erste Forschungsfrage. Die erste Forschungsfrage FF₁ lautet:

Sind die geprüften WFS grundsätzlich für das Verfahren geeignet?

Daraus ergibt sich folgendes Hypothesenpaar:

H₀ = Bei der Abfrage der Geowebsservices werden keine für das Verfahren relevanten Daten geliefert.

H_{a1} = Bei der Abfrage der Geowebsservices werden für das Verfahren relevante Daten geliefert.

Die relevanten Daten sind so definiert, dass der HTTP-Statuscode einer Anfrage 200 sein muss. Andere HTTP-Statuscodes im 200er Bereich kamen während des Experiments nicht vor. Das bedeutet, dass die Anfrage diesbezüglich erfolgreich war. Die Daten des WFS müssen jedoch auch plausibel sein. Aus diesem Grund muss die Anzahl der zurückgelieferten Objekte größer oder gleich null sein. Wenn diese beiden Bedingungen erfüllt sind, gelten die Daten der Anfrage als relevant.

Für das zweite Hypothesenpaar wird die zweite Forschungsfrage FF₂ als Grundlage verwendet:

Sind die derzeit verfügbaren WFS technisch in der Lage, bei synchronen Aufrufen die für die RPA benötigten Inputdaten am Beispiel von Schutzgebietsdaten zu liefern?

Die Hypothesen, die sich daraus ableiten, lauten:

H₀ = Der überprüfte WFS ist derzeit technisch nicht in der Lage die benötigten Daten am Beispiel von Schutzgebietsdaten zu liefern.

H_{a2} = Der überprüfte WFS ist derzeit technisch in der Lage die benötigten Daten am Beispiel von Schutzgebietsdaten zu liefern.

Der Teil ‚technisch in der Lage‘ wird über die gemessene Systemzeit definiert. Wenn die Bearbeitung und Ausführung der Anfrage und das zurücksenden der Antwort mehr als zehn Sekunden in Anspruch nimmt, gilt sie als nicht in der Lage die Daten zu liefern.

⁵⁸ Vgl. Wohlin et al. 2012, S. 91–92.

Die letzten Hypothesen ergeben sich aus der dritten Forschungsfrage FF₃:

Gibt es einen signifikanten Unterschied in der Antwortzeit des WFS während der Kernarbeitszeit von 9-15 Uhr im Vergleich zur Freizeit (außerhalb der Kernarbeitszeit)?

Die Hypothesen dazu lauten:

H₀ = Zwischen der (Kern-)Arbeitszeit und außerhalb der Kernarbeitszeit besteht kein Zusammenhang zur Antwortzeit des WFS.

H_{a3} = Zwischen der (Kern-)Arbeitszeit und außerhalb der Kernarbeitszeit besteht ein Zusammenhang zur Antwortzeit des WFS.

Die Kernarbeitszeit wird von 9-15 Uhr definiert. Das bedeutet, dass der Zeitraum von 15-9 Uhr in diesem Fall nicht mehr als Arbeitszeit, sondern als Freizeit angesehen wird. Dadurch sind die Berechnungen für die Auswertung weniger kompliziert.

Um diese Hypothesen zu überprüfen, werden statistische Tests verwendet. Sie dienen zur Bewertung der Ergebnisse eines Experiments. Das Testen von Hypothesen birgt verschiedene Risiken. Entweder weist der Test eine wahre Hypothese zurück oder der Test weist eine falsche Hypothese nicht zurück. Diese Risiken werden als Typ-I-Fehler und Typ-II-Fehler bezeichnet. Ein Typ-I-Fehler liegt vor, wenn ein statistischer Test ein Muster oder eine Beziehung anzeigt, auch wenn es in Wirklichkeit kein Muster gibt. Ein Typ-II-Fehler liegt vor, wenn ein statistischer Test kein Muster oder keinen Zusammenhang anzeigt, obwohl es tatsächlich ein Muster gibt.⁵⁹

4.2 Auswahl der Variablen und Subjekte

Bevor das Experiment beginnen kann, müssen die abhängigen und unabhängigen Variablen ausgewählt werden. Die unabhängigen Variablen sind die Variablen, die kontrolliert werden können. Die Variablen sollten eine gewisse Auswirkung auf die abhängige Variable haben. Die Auswahl der unabhängigen und abhängigen Variablen erfolgt oft gleichzeitig oder in umgekehrter Reihenfolge. Die Auswahl der unabhängigen Variablen umfasst auch die Wahl der Messskalen, des Bereichs für die Variablen und der spezifischen Niveaus, auf denen die Tests durchgeführt werden sollen. Die Wirkung der Behandlungen wird anhand der abhängigen Variable(n) gemessen. Häufig gibt es nur eine abhängige Variable, die daher direkt aus der Hypothese abgeleitet werden sollte. Die drei abhängigen Variablen sind also der Statuscode, die Anzahl der zurückgelieferten Objekte und die gemessene Systemzeit. Die Wahl der abhängigen Variable bedeutet auch, dass die Messskala und der Messbereich der

⁵⁹ Vgl. Wohlin et al. 2012, S. 91–92.

Variablen festgelegt werden. Ein Grund dafür, nur eine abhängige Variable zu wählen, ist, dass bei mehreren die Gefahr besteht, dass die Schwankungs- und Fehlerquote für die Gültigkeit der Schlussfolgerungen zu groß wird.⁶⁰

Bei der Stichprobe aus der Grundgesamtheit kann es sich entweder um eine Wahrscheinlichkeitsstichprobe oder um eine Nichtwahrscheinlichkeitsstichprobe handeln. Der Unterschied zwischen beiden besteht darin, dass bei der Wahrscheinlichkeitsstichprobe die Wahrscheinlichkeit der Auswahl der einzelnen Subjekte bekannt ist, während sie bei der Nichtwahrscheinlichkeitsstichprobe unbekannt ist. Bei einer einfachen Zufallsstichprobe sind die Subjekte in einer Liste der Grundgesamtheit. Aus dieser Liste wird das Subjekt dann per Zufallsprinzip ausgewählt. Das bedeutet jedes Element hat die gleiche Wahrscheinlichkeit ausgewählt zu werden. Dies ist ein Beispiel für Wahrscheinlichkeitsstichprobenverfahren. Bei Nichtwahrscheinlichkeitsstichprobenverfahren gibt es zum Beispiel Zufallsstichproben. Hierbei wird das nächstgelegene und am besten geeignete Subjekt gewählt. Die Zufallsstichprobe wird normalerweise für jedes Element verwendet. In dem Experiment werden die Koordinaten in einem bestimmten Bereich zufällig erzeugt. Dadurch sind die generierten Bounding Boxen auch zufällig gewählt. Der Umfang der Stichprobe wirkt sich auch auf die Ergebnisse bei der Verallgemeinerung aus. Je größer die Stichprobe ist, desto geringer ist der Fehler bei der Verallgemeinerung der Ergebnisse. Der Stichprobenumfang steht auch in engem Zusammenhang mit der Aussagekraft des statistischen Tests.⁶¹

4.3 Versuchsplanung

Die Planung und Interpretation sind stark miteinander verbunden. Um aussagekräftige Schlussfolgerungen aus einem Experiment ziehen zu können, müssen Analysemethoden auf die gesammelten Daten angewendet werden. Danach können die Ergebnisse interpretiert werden. Um valide Daten aus dem Experiment zu erhalten, muss es gut geplant und konzipiert sein. Die gemessenen Variablen und Messskalen bestimmen welche statistischen Analysen angewendet werden können. Ein Versuchsplan beschreibt die Organisation und Durchführung der Versuche. Dabei spielen die gebildeten Hypothesen aus Kapitel 4.1 eine große Rolle. Sie haben einen maßgeblichen Einfluss welche Analysen und Hypothesentests durchgeführt werden sollten. Außerdem ist der Versuchsplan die Grundlage für eine Replikation. Bei der Planung eines Versuchs müssen viele Aspekte berücksichtigt werden. Die allgemeinen Planungsprinzipien nach Wohlin et al. sind Randomisierung, Blocking und Balancierung. Der Fokus für diese Arbeit liegt auf der Randomisierung. Hierbei werden die Objekte und Subjekte

⁶⁰ Vgl. Wohlin et al. 2012, S. 92.

⁶¹ Vgl. Wohlin et al. 2012, S. 92–93.

zufällig ausgewählt. Dadurch können Auswirkungen eines möglicherweise nicht beachteten und vorhandenen Faktors ausgeglichen werden.⁶²

4.4 Instrumentierung

Es gibt drei Arten von Instrumenten für ein Experiment, nämlich Objekte, Leitlinien und Messinstrumente. Für dieses Experiment steht das Messwerkzeug im Vordergrund. Bei der Planung eines Experiments werden die Instrumente ausgewählt. Vor der Durchführung wird das Messwerkzeug für das spezifische Experiment entwickelt. Bei den Versuchsobjekten kann es sich z. B. um Spezifikations- oder Codedokumente handeln. Bei der Planung eines Experiments ist es wichtig, geeignete Objekte zu wählen. Die Messungen in einem Experiment erfolgen über die Datenerhebung. Das übergeordnete Ziel der Instrumentierung besteht darin, Mittel zur Durchführung des Experiments und zu dessen Überwachung bereitzustellen, ohne die Kontrolle des Experiments zu beeinträchtigen.⁶³

Wie in Kapitel 1 beschrieben, dient das Messwerkzeug nicht nur zum Sammeln großer Datenmengen für die statistische Auswertung. Es bietet auch die schnelle Durchführung von Anwendungsfällen an. Durch diese Teilung kann es auch produktiv einen Mehrwert als Automatisierungstool darstellen.

4.5 Konzeption des Messwerkzeugs

Die Unified Modeling Language (UML) dient der Bestimmung, Dokumentation und Erstellung eines Softwareprodukts. Alle Darstellungen in diesem Kapitel wurden mit der Software Microsoft Visio⁶⁴ erstellt. Durch UML-Diagramme können Entwicklungsprozesse beschleunigt werden.⁶⁵

Die dargestellten Diagramme befinden sich in einer speziellen Reihenfolge. Zuerst wird ein Anwendungsfalldiagramm dargestellt, um die an dem Prozess beteiligten Systeme zu ermitteln. Danach wird auf diesen Erkenntnissen aufgebaut und ein Klassendiagramm mit den Systemen erstellt. Darauf folgen zwei Sequenzdiagramme mit zwei ausgewählten Anwendungsfällen, um das abgebildete Szenario besser verdeutlichen zu können.

⁶² Vgl. Wohlin et al. 2012, S. 94.

⁶³ Vgl. Wohlin et al. 2012, S. 101.

⁶⁴ <https://www.microsoft.com/de-de/microsoft-365/visio/flowchart-software>

⁶⁵ Vgl. Fauzan et al. 2019, S. 3.

4.5.1 Anwendungsfalldiagramm

Bei einem Anwendungsfalldiagramm wird aufgezeigt, wie ein System mit externen Entitäten interagiert. Details sind bei diesem Diagrammtyp nicht im Blickpunkt, sondern wie sich das System intern verhält und wie die externe Umgebung aufgebaut ist. Bei Use Case Diagrammen gibt es vier Hauptsymbole. Das erste ist der Akteur. Dies ist eine Person, eine Gruppe, eine Organisation oder ein externes System, das eine Rolle in einer Interaktion mit dem System spielt. Dies können auch Netzwerke, Computer oder Programme sein. Ein Akteur wird durch ein Strichmännchen dargestellt und besitzt einen Namen. Das zweite Hauptsymbol ist der Anwendungsfall. Er beschreibt eine Abfolge von Aktionen. Diese Aktionen müssen einem Akteur einen messbaren Wert zurückgeben. Ein Anwendungsfall wird durch eine horizontale Ellipse dargestellt. Das dritte Symbol ist die Assoziation. Dies ist eine Interaktion, die durch einen Anwendungsfall beschrieben wird. Eine Assoziation wird durch Linien zwischen Anwendungsfällen und Akteuren dargestellt. Außerdem können die Verbindungen mit einer Pfeilspitze ausgestattet sein. Damit kann die Richtung des Aufrufs der Beziehung nachvollzogen werden. Das letzte Hauptsymbol ist die Systemgrenze. Sie wird als großes Rechteck um die Anwendungsfälle dargestellt. Alles innerhalb dieses Rahmens ist die Funktionalität im Geltungsbereich des Systems.⁶⁶

Die Akteure werden in ihren einzunehmenden Rollen nach Anwendungsfall unterschieden. Beispielsweise hat ein Kunde andere funktionale Anforderungen als ein Lieferant. Durch den Anwendungsfall wird beschrieben was das System leisten muss, jedoch nicht wie. Im Use Case Diagramm gibt es noch zwei Besonderheiten. Bei Fehlern oder Sonderfällen wird meistens der Anwendungsfall ausgelagert und dann mit einer mit ‚extends‘, zu Deutsch erweitern, beschrifteten Assoziation mit dem Normalfall verbunden. Bei einer Assoziation mit ‚include‘, zu Deutsch einschließen, wird ein Anwendungsfall geschaffen, der den gleichen Teil mehrerer Anwendungsfälle beschreibt. Ein Use Case Diagramm ist einfach gehalten. Ein Verhaltensdiagramm beschreibt den Ablauf eines Anwendungsfalls aus dem Anwendungsfall genauer. Dies geschieht im nächsten Unterkapitel per Sequenzdiagramm.⁶⁷

⁶⁶ Vgl. Seidl et al. 2015, S. 23–27.

⁶⁷ Vgl. Seidl et al. 2015, S. 30–32.

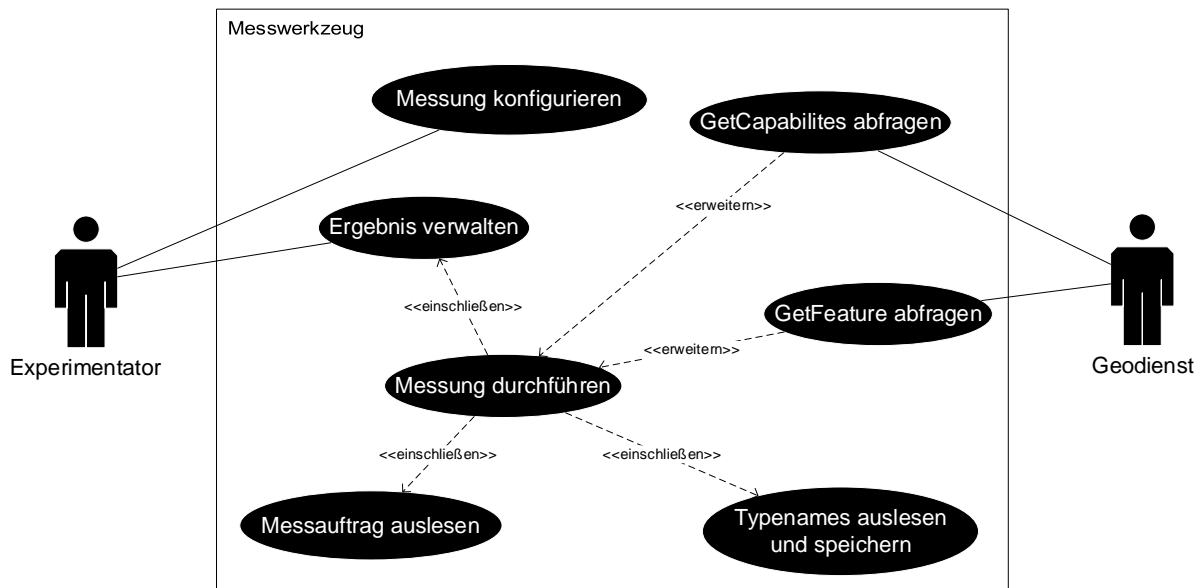


Abbildung 1: Anwendungsfalldiagramm des Messwerkzeugs ⁶⁸

In Abbildung 1 sind zwei Akteure dargestellt. Zum einen gibt es den Experimentator als Konfigurator der Eingabedatei. Da das Messwerkzeug und dementsprechend die Messungen abhängig von der Eingabedatei sind, ist diese Funktion sehr wichtig. Das Messwerkzeug bildet die Systemgrenze, mit der die Akteure interagieren. Dies wird durch den Anwendungsfall Messung konfigurieren dargestellt. Die Anwendungsfälle sind durch Ellipsen abgebildet. Auf der rechten Seite ist der abzufragende Geodienst dargestellt, mit welchem die Messung durchgeführt wird. Bei der Messung wird zwischen GetFeature und GetCapabilities unterschieden. Diese beiden Anwendungsfälle sind mit dem Geodienst über Assoziationen verbunden. Beide Fälle führen dazu, dass eine Messung vom Messwerkzeug durchgeführt wird. Vor der Ausführung wird die Eingabedatei als Messauftrag ausgelesen und das Messwerkzeug passt die Anfragen dann an diese Bedürfnisse an. Wenn eine Messung durchgeführt wird, werden die Typenames temporär im Messwerkzeug abgespeichert. Nachdem die Abfrage des WFS durchgeführt wurde, wird das Ergebnis für spätere Auswertungen menschenlesbar abgespeichert. Auf diese Ergebnisse kann der Experimentator im letzten Anwendungsfall zugreifen und diese dann verwalten.

⁶⁸ Eigene Darstellung

4.5.2 Klassendiagramm

Klassendiagramme sind wichtig für die meisten Systemmodellierungen. Diese Diagramme beschreiben die Architektur eines Systems. Besonders wichtig ist das Konzept Klasse, denn es wird durchgängig in der Modellierung und Programmierung eingesetzt. Klassendiagramme bilden die Grundlage für weitere Diagrammart, da diese sich auf den Inhalt des Klassendiagramms beziehen. Eine Klasse hat mehrere Aufgaben wie beispielsweise die Kapselung von Attributen und Methoden, die Typisierung von Objekten und sie dienen als Charakterisierung der möglichen Struktur für das spätere System. Attribute und Methoden können je nach Grad der Detaillierung weggelassen werden.⁶⁹

Eine Klasse besteht aus Attributen und Methoden. Diese legen den Zustand und das Verhalten der Objekte fest. Klassen werden durch Assoziationen miteinander verbunden. Durch den Klassennamen kann die Klasse identifiziert werden. Ein Attribut ist eine Zustandskomponente einer Klasse. Sie beinhalten einen Namen und Typ. Sie dienen zur Beschreibung von Eigenschaften. Ein Typ ist ein Grunddatentyp aus der Programmierung wie beispielsweise Integer (int). Eine Methode beinhaltet die Funktionalität einer Klasse. Sie beschreibt die Implementierung. Eine Assoziation ist eine binäre Beziehung von Klassen. Sie dient zur Realisierung struktureller Informationen. Eine Assoziation besitzt einen Namen und Kardinalitäten. Optional sind Rollennamen am Ende zu jeder Klasse. Die Kardinalität wird am Ende jeder Assoziation angegeben. Eine beispielhafte Form ist ‚0...1‘, ‚1‘ oder ‚0...*‘. Sie beschreibt, ob eine Assoziation in dieser Richtung eindeutig ist oder durch ein Sternchen, mehrfache Bindung erlaubt.⁷⁰

In Abbildung 2 sind insgesamt acht Klassen zu sehen. Jede einzelne ist gekennzeichnet durch ein Rechteck. Attribute und Methoden werden in diesem Diagramm nicht dargestellt. Insgesamt gibt es sieben Assoziationen. Außerdem sind vier Kardinalitäten abgebildet.

⁶⁹ Vgl. Rumpe 2016, S. 13–14.

⁷⁰ Vgl. Rumpe 2012, S. 34–35.

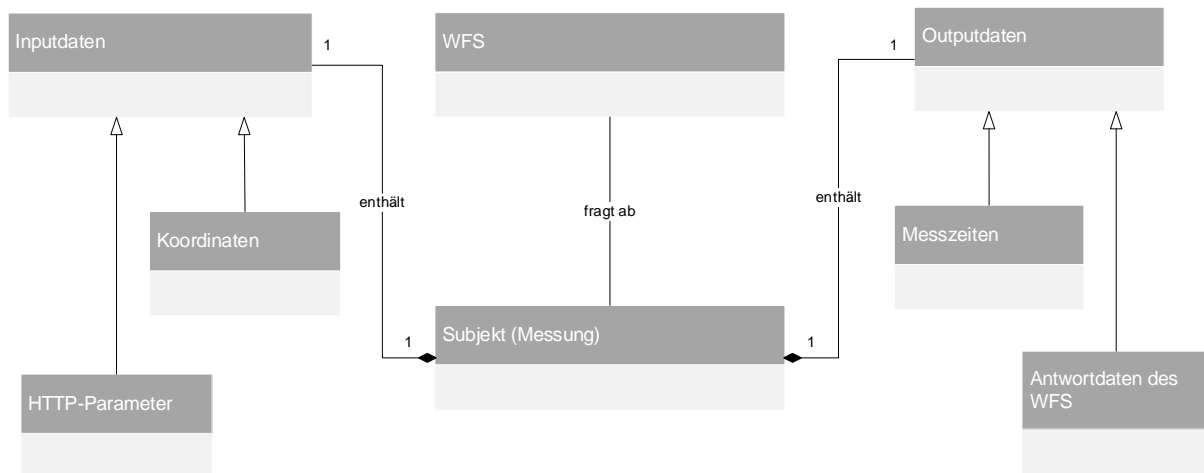


Abbildung 2: Klassendiagramm einer Messung⁷¹

Das Klassendiagramm wird genutzt, um Konzepte der realen Welt zu strukturieren. Das Diagramm ist ohne Methoden, Attribute und Typen, um die Lesbarkeit zu erhöhen. Im Mittelpunkt des Klassendiagramms steht die Messung. Diese Klasse besitzt eine Verbindung zum WFS. Der Web Feature Service wird bei einer Messung abgefragt und liefert Datensätze als Antwort zurück. Bevor eine Messung gestartet werden kann, werden Inputdaten benötigt. Eine Vielzahl dieser Daten wird bei jeder Messung mit abgespeichert. Dadurch weist die Klasse Messung die Komposition ‚enthält‘ mit den Inputdaten auf. Die Komposition ist eine Sonderform der Aggregation und ist an einer gefüllten Raute zu erkennen. Dadurch wird ausgedrückt, dass die Teile von der Existenz des Ganzen abhängig sind. Wird bei einer Komposition die übergeordnete Klasse gelöscht, so wird die Unterklasse ebenso zerstört. Das bedeutet, dass bei einer Löschung einer Messung die abgespeicherten Inputdaten mitgelöscht werden. Die Inputdaten können weiterhin unterteilt werden. Es gibt beispielsweise die HTTP-Parameter aus der Eingabedatei wie beispielsweise die Methode oder die URL. Außerdem gibt es noch die Klasse Koordinaten, die zum Beispiel die Grenzen der Bounding Box enthalten. Neben den Inputdaten besteht die Messung auch aus Outputdaten. Darin sind alle Messzeiten enthalten. Diese könnten weiter eingeteilt werden, da sie nicht vom WFS geliefert, sondern vom Messwerkzeug gemessen werden. Eine weitere Unterklasse der Outputdaten sind die Antwortdaten des WFS. Diese beinhaltet beispielsweise die Anzahl der zurückgegebenen Objekte oder die Typenamen. Für diese Klasse ist die Verbindung ebenfalls eine weitere Komposition mit dem Namen ‚enthält‘. Bei der Löschung einer Messung, werden die Outputdaten ebenfalls gelöscht.

⁷¹ Eigene Darstellung

4.5.3 Sequenzdiagramm

Sequenzdiagramme modellieren Interaktionen zwischen Objekten. Es stellt einen beispielhaften Ausschnitt aus dem Ablauf des zu modellierenden Softwaresystems dar. Das Diagramm zeigt die auftretenden Interaktionen und Aktivitäten. Das Sequenzdiagramm beschreibt die Reihenfolge, in der Methodenaufrufe absolviert und beendet werden. Die Objekte werden in einer Reihe nebeneinander platziert und mit einer vertikalen Zeitlinie versehen, die verdickt dargestellt wird, wenn das Objekt aktiv ist.⁷²

Ein Sequenzdiagramm enthält fünf wichtige Teile. Als erstes das Objekt, welches einen Namen und einen Typ besitzt. Die zeitlich aufeinanderfolgenden Ereignisse werden von oben nach unten gelesen. Die Objekte in einem Sequenzdiagramm besitzen eine Zeitlinie oder Lebenslinie, die der zweite Teil eines Sequenzdiagramms ist und das Voranschreiten der Zeit des zu betrachteten Objekts zeigt. Die Aufgabe der Zeitlinie ist es, die zeitliche Reihenfolge von Interaktionen einfacher nachvollziehen zu können. Der dritte Teil ist die Interaktion zwischen zwei Objekten. Sie kann beispielsweise durch Methodenaufrufe, Returns, Fehlermeldungen oder asynchrone Nachrichten ausgelöst werden. Der vierte Teil ist der Aktivitätsbalken. Durch ihn wird die aktive Zeit eines Objekts bei einem Methodenaufruf dargestellt. Bei Rekursion ist der Aktivitätsbalken geschachtelt. Der letzte Teil ist die Bedingung. Sie dient zur Beschreibung von Eigenschaften während eines Ablaufs.⁷³

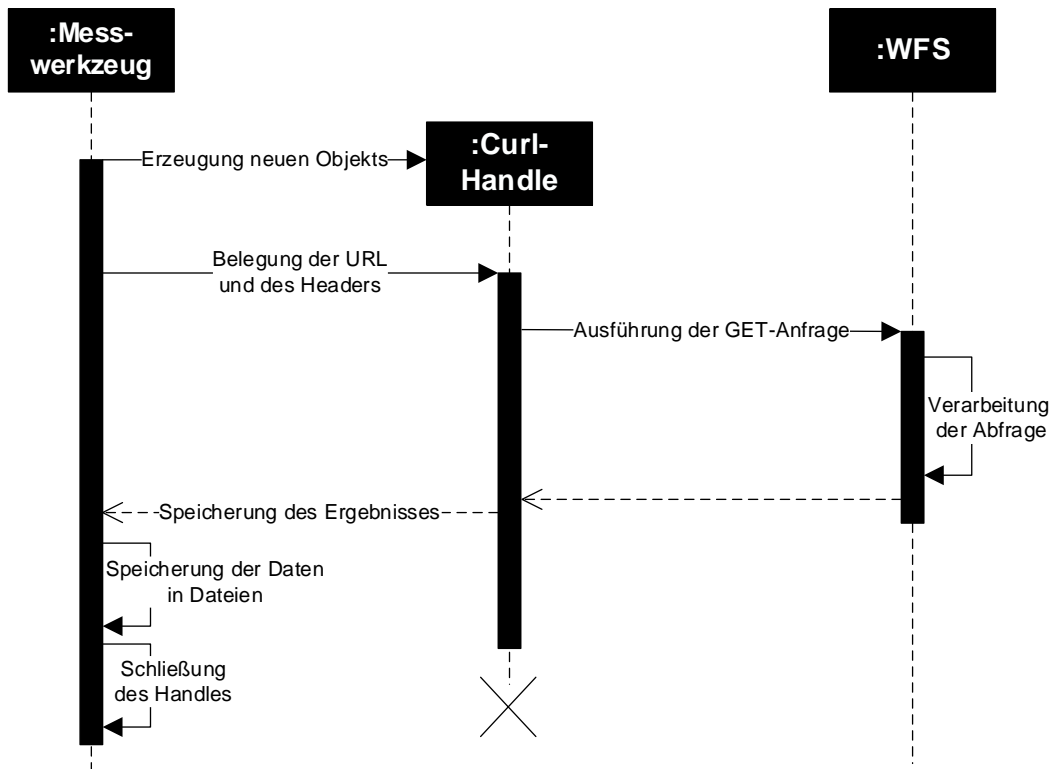
Jede Interaktion an das Nachbarobjekt wird durch einen horizontalen Pfeil dargestellt. Das zeigt, dass die hierbei verbrauchte Zeit vernachlässigt wird. An einem solchen Pfeil befinden sich beispielsweise Methodenaufrufe oder Fehlermeldungen. Wenn ein Objekt im Laufe der Interaktion neu erzeugt wird, beginnt die Zeitlinie des Objekts erst später. Ein gestrichelter Pfeil in Rückrichtung wird verwendet, um das Ergebnis des Methodenaufrufs zurückzuschicken, also den Returnwert. Ein Pfeil von einem Objekt auf sich selbst stellt eine interne Operation dar. Eine Nachricht von einem Objekt kann nur versendet werden, wenn die beiden durch eine Assoziation verbunden sind oder das Objekt das zweite erzeugt hat.⁷⁴

Ein Sequenzdiagramm stellt eine Abstraktion eines Ablaufs dar und somit kann es weder alle Objekte des Systems noch alle Interaktionen beinhalten. Deswegen sind die folgenden Diagramme in GET und POST unterteilt und vereinfacht dargestellt.

⁷² Vgl. Rumpe 2016, S. 191–192.

⁷³ Vgl. Rumpe 2012, S. 69.

⁷⁴ Vgl. Seidl et al. 2015, S. 108–110.

Abbildung 3: Sequenzdiagramm GET ⁷⁵

Im Sequenzdiagramm aus Abbildung 3, welches die Abfrage mittels GET veranschaulicht, wird der Anwendungsfall Messung durchführen dargestellt. Das Diagramm beinhaltet drei Objekte. Die zwei Objekte, die von Anfang an aktiv beziehungsweise bereit sind, sind das Messwerkzeug und der WFS. Das dritte Objekt wird erst im Laufe der Abfrage vom Skript erzeugt. Das Startobjekt für dieses Diagramm ist das Messwerkzeug. Dies erzeugt ein neues Objekt namens Curl-Handle. Zu diesem Zeitpunkt beginnt die Lebenslinie des Handles. Das Objekt wird mit der abzufragenden URL und Daten für den Header belegt. Das wird durch die Aktivierung auf dem Aktivitätsbalken angezeigt. Nach etwas Zeit führt das Objekt die Abfrage an die URL, also dem WFS, aus und dies aktiviert das dritte Objekt im Diagramm. Der WFS nimmt die Anfrage an und verarbeitet sie. Danach wird das Ergebnis an das Handle gereicht und die Aktivierung des WFS wird für diesen Anwendungsfall beendet. Das Curl-Handle enthält jetzt das Ergebnis der Abfrage. Danach wird dies an das Messwerkzeug gereicht und als Variable abgespeichert. Die Daten werden dann in Dateien geschrieben und stehen dem Skript für weitere Verarbeitungen als Variable zur Verfügung. Nach dieser Aktion wird das Handle geschlossen, um die Systemressourcen wieder freizugeben. Die Lebenslinie des

⁷⁵ Eigene Darstellung

Handles wird durch ein großes Kreuz beendet und zeigt so, dass das Objekt nicht mehr verfügbar ist.

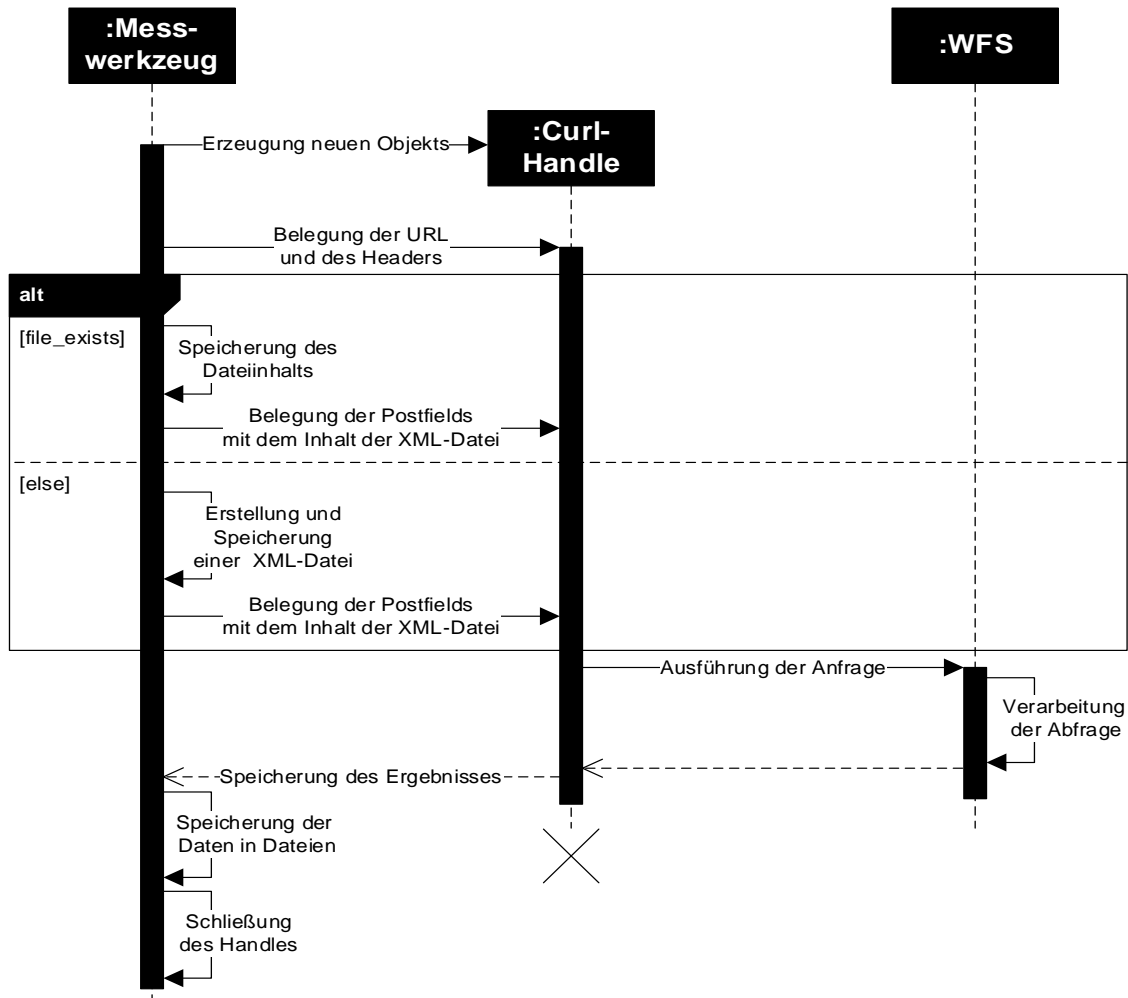


Abbildung 4: Sequenzdiagramm POST ⁷⁶

In Abbildung 4 ist der Fall der POST-Abfrage zu sehen. Bis zur Belegung der URL und des Headers ist der Ablauf wie in Abbildung 3. Danach folgt eine bedingte Anweisung. Wenn die XML-Datei existiert, wird sie ausgelesen und das Ergebnis wird zur Laufzeit im Messwerkzeug gespeichert. Danach wird der Inhalt der Datei an das Handle gegeben. Falls die Datei nicht existiert oder der Dateipfad falsch ist, generiert das Messwerkzeug automatisiert eine valide XML-Datei für die Abfrage, die im gleichen Verzeichnis wie das Skript abgespeichert wird. Mit dem Inhalt der Datei wird die Abfrage durchgeführt. Durch die Speicherung der Datei kann diese später weiter angepasst oder erweitert werden, wenn Bedarf besteht. Nach der Belegung der Postfields im Handle erfolgt die gleiche Prozedur wie in Abbildung 3.

⁷⁶ Eigene Darstellung

4.6 Eingabedatei für das Messwerkzeug

Das Messwerkzeug erfordert zusätzliche Parameter, die über eine Eingabedatei an das Skript geliefert werden. Über diese Datei ist es dem Nutzer möglich die durchzuführende Messung anzupassen. In diesem Kapitel wird der Stand der Eingabedatei für die experimentelle Messung beschrieben, im Ausblick werden dann weitere Ideen erläutert.

Die Eingabedatei erwartet, dass 13 Parameter angegeben werden. Diese sind zusätzlich eingeteilt in Pflichtfelder und optionale Felder. Pflichtfelder sind in der Datei durch einen Stern markiert und sollten nicht leer bleiben. Das kann zum Absturz des Messwerkzeugs oder zu unerwarteten Ergebnissen führen. Außerdem gibt es zwei Parameter, die mit Fragezeichen gekennzeichnet sind. Sie können nur die Werte true und false annehmen. Dies sind logische Werte und können mit nur einem Bit dargestellt werden. Dadurch können logische Operationen einfacher durchgeführt werden.⁷⁷ Im Folgenden werden die Parameter aufgelistet, übersetzt und erklärt:

1. **URL** ist der Link zum WFS beziehungsweise die Adresse der abzufragenden Webseite. Zu beachten bei diesem Parameter ist die korrekte Eingabe. Sowohl das Protokoll (http/https) inklusive der Doppelpunkte und Schrägstriche (Slashes) als auch das Fragezeichen am Schluss der URL muss mit angegeben werden. Wenn die URL des GetCapabilities-Aufrufs bekannt ist, kann diese bis zum Fragezeichen eingekürzt werden. Die WFS geben bei der GetCapabilities-Abfrage die zu benutzenden URLs für GET und POST an. Für gewöhnlich unterscheiden sich diese aber nicht. Dieser Parameter ist ein Pflichtfeld.

```

-<ows:OperationsMetadata>
  -<ows:Operation name="GetCapabilities">
    -<ows:DCP>
      -<ows:HTTP>
        <ows:Get xlink:href="https://geodienste.bfn.de/ogc/wfs/schutzgebiet?"/>
        <ows:Post xlink:href="https://geodienste.bfn.de/ogc/wfs/schutzgebiet"/>
      </ows:HTTP>
    </ows:DCP>
  
```

Abbildung 5: Ausschnitt aus der Abfrage der GetCapabilities des BfN ⁷⁸

In Abbildung 5 aus der Abfrage des WFS vom Bundesamt für Naturschutz lassen sich die zu verwendenden URLs für GET und POST für diesen Web Feature Service ablesen.

⁷⁷ Vgl. Lange 2020, S. 89.

⁷⁸ <https://geodienste.bfn.de/ogc/wfs/schutzgebiet>

2. **Time interval between measurements** steht für den Zeitabstand, der zwischen den Messungen bestehen soll. Da dies im Programmcode als Gleitkommazahl (float) behandelt wird, sollten ausschließlich natürliche Zahlen oder Dezimalzahlen angegeben werden. Außerdem sollten die Werte im positiven Wertebereich liegen, da das Messwerkzeug keine negativen Zeiten umsetzen kann. Bei der Eingabe einer Dezimalzahl ist auf das Trennungszeichen zu achten. Es sind ausschließlich Punkte oder Kommata als Trennzeichen erlaubt. Die eingegebene Zahl stellt Minuten dar. Bei der Eingabe einer eins liegt eine Minute zwischen den Durchläufen. Für eine Stunde wird dementsprechend der Wert 60 eingetragen und für beispielsweise sechs Sekunden der Wert 0.1 oder 0,1. Dieser Parameter ist ein Pflichtfeld.
3. **Number of runs** ist die vorgegebene Anzahl an Durchläufen, bis das Messwerkzeug stoppt. Bei der Eingabe ist auf eine natürliche Zahl zu achten, also keine Buchstaben oder Gleitkommazahlen wie zuvor. Die Zahl muss außerdem im positiven Wertebereich liegen. Dieser Parameter ist ein Pflichtfeld.
4. **Coordinates separated with spaces** erwartet Koordinaten, die mit Leerzeichen getrennt sind. Diese Koordinaten müssen sich an das später eingestellte EPSG richten. Es ist möglich ein oder mehrere Koordinaten anzugeben, sofern das Trennzeichen der Koordinaten ein Leerzeichen ist. Dies ist wichtig, um die verschiedenen Geometrietypen zu testen. In diesem Experiment liegt der Fokus jedoch auf der Bounding Box und deswegen werden meistens zwei Koordinaten benutzt. Koordinaten mit Nachkommastellen sollten einen Punkt enthalten, da ansonsten die Gültigkeit der Abfrage verloren geht oder das Ergebnis der Abfrage nicht den Erwartungen entspricht. Dieser Parameter ist optional.
5. **Version** ist der Parameter, der angibt in welcher Version der WFS antwortet. Nach heutigem Stand gibt es drei mögliche Zustände dafür: 1.0.0 oder 1.1.0 oder 2.0.0. Dieser Parameter ist optional.
6. **Http method** steht für die HTTP-Anfragemethode, die an den WFS gestellt wird. HTTP unterstützt neun solcher Methoden. Für diese Arbeit sind jedoch nur zwei, GET und POST, in Verwendung. Ein weiterer Zustand, den dieser Parameter annehmen kann, ist both/beides/POST&GET. Damit werden beide Abfragen hintereinander durchgeführt. Dieser Parameter ist ein Pflichtfeld.
7. **EPSG** gibt das zu nutzende Koordinatensystem an. Wird dieser Parameter leer gelassen, wird der dazugehörige Standardwert des WFS benutzt. Dieser Wert ist optional.
8. **Path to the xmlfile** verlangt einen validen Pfad zu einer existierenden XML-Datei. Dieser Parameter ist optional.

9. **Geometry type** bezeichnet den Geometrietypen, der abgefragt werden soll. Zum Zustand dieser Arbeit kann das Messwerkzeug Punkte, Linien, Bounding Boxen, Multipunkte und Multilinien verarbeiten und abfragen. Es sollte jedoch vorher sichergestellt werden, dass der WFS den abzufragenden Typ auch verarbeiten kann. Dieser Parameter ist ein Pflichtfeld.
10. **Trialrun** bezeichnet einen Testlauf. Die Funktionalität wird im Kapitel 5.2.3 erläutert. Dies ist ein boolean-Wert innerhalb des Programmcodes und kann daher nur zwei Werte annehmen. Wird das Feld leer gelassen bedeutet es, dass kein Testlauf durchgeführt wird. Wenn das Feld befüllt ist, egal ob eine Zahl oder Buchstaben, wird der Testlauf aktiviert. Dieser Parameter ist optional.
11. **Name of the files** wird als Dateiname für die verschiedenen Ausgabedateien des Messwerkzeugs verwendet. Hierbei sind die gängigen Regeln zur Benennung von Dateien zu beachten. Das bedeutet keine Anführungszeichen, Slashes, Doppelpunkte, Fragezeichen etc. verwenden. Dieser Parameter ist ein Pflichtfeld.
12. **Boundaries of the Bounding Box** bezeichnet die Grenzen der zu verwendenden Bounding Box. Auf die Funktionalität wird im Kapitel 4.8 näher eingegangen. Dieser Parameter ist auszufüllen wie beim vierten Parameter. Es werden zwei Koordinaten mit einem Leerzeichen getrennt erwartet. Die Aktivierung ist nur sinnvoll, wenn der folgende, der 13. Parameter, aktiviert ist. Dieser Parameter ist optional.
13. **Randomly generate centrepoint** ist wie der Testlauf ein boolean. Hierbei wird in der vorher angegebenen BBox aus 12. ein zufälliger Punkt in diesen Grenzen generiert. Wird das Feld leer gelassen, bedeutet es, dass keine Zufallspunkte generiert werden. Dieser Parameter ist optional.

Durch diese 13 Parameter, von denen sechs Pflichtfelder sind, lässt sich das Messwerkzeug an verschiedene Experimente und Anwendungsfälle anpassen. In Kapitel 5.2 werden verschiedene Anwendungsszenarien vorgestellt und Potentiale des Skripts erläutert. Die Parameter werden mit Hilfe eines Zeilenumbruchs voneinander getrennt. Im Entwicklungsprozess gab es einen 14. Parameter namens Typename. Damit wurde ein spezifischer Typename vom WFS abgefragt. Diese Eingabe wurde jedoch aus zwei Gründen entfernt. Zum einen ist das eine weitere Nutzereingabe, da dieser Parameter ein Pflichtfeld war. Diese Eingabe muss ebenfalls überprüft werden. Da Typenames konkret vom WFS vorgegeben sind und dementsprechend individuell sind, muss der Benutzer wissen welcher Typename zu seinen Bedürfnissen passt und diesen aus den GetCapabilities auslesen. Zum anderen wurde dieser Parameter aus den genannten Gründen komplett ersetzt. Das Messwerkzeug fragt automatisch alle Typenames ab. Dies ist für die Überprüfung und Sammlung von Datensätzen für diese Arbeit ebenfalls sinnvoller als spezifische Typenames vorzugeben.

4.7 Eingabedatei als Webseite

Die Eingabedatei ist für einen produktiven Gebrauch für Endnutzer ungeeignet, da die Eingaben nicht vollständig und vor dem Start überprüft werden und die Eingabe in die Datei ebenfalls fehleranfällig ist. Ein ungewollter Zeilenumbruch führt beispielsweise dazu, dass die Parameter nicht mehr richtig eingelesen werden können. Die Eingabedatei ist nicht intuitiv für unerfahrene Benutzer und bedarf einer Anleitung. Außerdem kann das Messwerkzeug so nur über eine Kommandozeile gestartet werden. Tippfehler, die nicht erkannt werden, sind ebenfalls ein Problem, da sie zu unerwünschtem Verhalten des Messwerkzeugs oder WFS führen können. Aus diesen Gründen wurde die Kompatibilität des Messwerkzeugs mit Daten aus Formularen einer Webseite getestet. Dadurch ist die Eingabe der Parameter und der Start des Messvorgangs nutzungsfreundlicher und weniger fehleranfällig. Außerdem ist der Weg über den Browser ein schon bekannter Prozess, der ohne eine Anleitung oder Einführung auskommt und daher eher Zuspruch findet. Im Folgenden wird eine beispielhafte Eingabemaske als Webseite veranschaulicht.

URL des WFS *

Zeitabstand in min (0.1 \triangleq 6sek) *

Anzahl der Durchläufe *

Koordinate(n) *

Version

Methode

EPSG

Abbildung 6: Eingabemaske einer Webseite für das Messwerkzeug ⁷⁹

In Abbildung 6 sind einige Parameter aus Kapitel 4.6 als Formulare dargestellt. Das Feld der URL ist wie in der Eingabedatei ein Freitextfeld. Beim Zeitabstand sind im Eingabefeld rechts zwei Pfeile zu erkennen. Diese zeigen an, dass es sich um ein Eingabefeld für Zahlen handelt.

⁷⁹ Eigene Darstellung

Die Pfeile können den eingegebenen Zahlenwert inkrementieren, also erhöhen oder dekrementieren, also verringern. In HTML-Formularen ist es möglich die Eingabe bei Bestätigung eines Buttons zu überprüfen. Dadurch kann die Nutzereingabe sofort und vor Start der Messung auf Validität geprüft werden. Wenn die Eingabe nicht passt oder falsch ist, wird das betroffene Feld markiert und der Nutzer kann seine Eingabe verbessern. Wie in Kapitel 4.6 sind Pflichtfelder mit einem Sternchen markiert. Wenn sie nicht befüllt sind, gibt es ebenfalls eine Fehlermeldung. Es ist möglich wie bei dem Feld der URL einen Platzhalter einzufügen, sodass der Nutzer weiß, was vom Formular erwartet wird. Eine weitere Möglichkeit sind die Dropdownfelder wie bei der Version oder der Methode. Damit sind Standardwerte vorgegeben und der Nutzer muss diese nur noch auswählen. Die Vorteile der Webseite sind die einfache Zuordnung der Felder zu ihrem Inhalt für den Nutzer und auch für das Skript und die Überprüfung der Eingabewerte vor Start der Messung. Dies wird durch den Aufbau der Webseite erreicht. Bei HTML kann in den Formularen über die Attribute Name und ID deren Inhalte dem Skript zur Verfügung gestellt werden.

4.8 Messwerkzeug

Die Inbetriebnahme des Messwerkzeugs erfolgt über die Kommandozeile. Es ist keine Parameterübergabe erforderlich, da dies nicht wie bei herkömmlichen Linux-Befehlen implementiert ist. Die zuvor beschriebene Eingabedatei mit den eingegebenen Daten wird dann mithilfe einer Funktion eingelesen. Die internen Funktionen können jedoch nur Eingaben mit einem Dezimalpunkt verarbeiten. Bei Verwendung eines Dezimalkommas können Probleme auftreten, die im schlimmsten Fall das Skript beenden. Normalerweise werden bei der Typumwandlung die Nachkommastellen abgeschnitten. Dies ist jedoch unerwünscht, da dadurch die Genauigkeit verloren geht. Aus diesem Grund wird der Wert vorher umgewandelt, indem alle Kommata durch Punkte ersetzt werden. Für das korrekte Einlesen der Eingabedatei ist es wichtig, dass sich diese im gleichen Verzeichnis wie das Messwerkzeug befindet. Andernfalls wird das Skript mit einer Fehlermeldung beendet. Dies liegt daran, dass wichtige Variablen fehlen, wenn die Eingabedatei nicht vorhanden ist und diese im Messwerkzeug nicht als Standard definiert sind. Die Datei mit den Daten wird zeilenweise ausgelesen und die Nutzereingaben als Variablen gespeichert. Anschließend werden die verfügbaren Variablen mithilfe von Funktionen an die Anforderungen des Programmcodes angepasst. Dies beinhaltet beispielsweise das Entfernen von Leerzeichen vor oder nach einem Wert. Eine weitere Funktion besteht in der Durchführung von Variablentypumwandlungen, auch als ‚type casting‘ bezeichnet. Dabei erfolgt eine Umwandlung des Datentyps einer Variable, beispielsweise von einem Integer zu einem String. Durch diese Umwandlung werden potenzielle Typverletzungen vermieden, die zu Fehlverhalten oder Abstürzen des Messwerkzeugs führen könnten. Nach Abschluss dieses Prozesses wird überprüft, ob bei der

Eingabe die Methode GET, POST oder beide Methoden ausgewählt wurden. Im Standardfall wird GET gewählt, falls beispielsweise ein Tippfehler vorliegt. Anschließend wird im gleichen Verzeichnis wie das Messwerkzeug eine Textdatei mit dem Namen aus der Eingabedatei erstellt. Die genaue Beschreibung dieser Textdatei erfolgt im Kapitel 4.9.1. Danach wird die Kopfzeile der CSV-Datei geschrieben und das Skript beginnt mit der Vorbereitung und Ausführung der anstehenden Anfragen. Die Funktionalitäten der CSV-Datei werden im Kapitel 4.9.2 näher erläutert. Die vom Benutzer eingegebene URL wird um statische Parameter erweitert, um den ersten Aufruf des gewünschten WFS vorzubereiten und durchzuführen. Diese erste Abfrage, welche wichtige Informationen ausliest, ist die GetCapabilities-Anfrage. Dabei wird der WFS mit der zuvor definierten URL aufgerufen und das Ergebnis wird gespeichert. Darüber hinaus werden interne Messungen durchgeführt, um sie mit den Ergebnissen der Curl-Funktion zu vergleichen. Das Ergebnis des Aufrufs wird ausgewertet und gespeichert. Dies erfolgt beispielsweise durch das Durchsuchen der XML-Datei nach der Kennzahl ‚numberReturned‘, die angibt, wie viele Objekte vom WFS zurückgeliefert wurden. Am Ende der Abfrage werden zwei Dateien beschrieben. Die erste ist die Textdatei, die wichtige Informationen für den Experimentator enthält. Die zweite ist die CSV-Datei, in der die Kennzahlen abgespeichert werden, um sie anschließend einfach mit Excel analysieren und weiterverarbeiten zu können. Nachdem das Ergebnis gespeichert wurde, wird es auf bestimmte Zeichenketten, sogenannte Strings, überprüft. Dieser Vorgang wird als ‚Parsing‘ bezeichnet. Das Ergebnis der GetCapabilities-Anfrage wird nach den verschiedenen Typenames geparkt, die vom zuvor abgefragten WFS angeboten werden. Jedes Mal, wenn ein Typename erreicht wird, wird er in einem Array gespeichert. Dieser Prozess muss in mehrere Schritte unterteilt werden, da zwei verschiedene Elemente zur Bezeichnung der Typenames verwendet werden. Einige WFS verwenden den Bezeichner ‚<Name>‘, während andere ‚<wfs:Name>‘ verwenden. Dies muss beim Parsing berücksichtigt werden, um die gewünschten Ergebnisse zu erzielen. Nachdem alle verfügbaren Typenames ausgelesen und gespeichert wurden, stehen vier verschiedene Pfade zur Auswahl, je nach Nutzereingabe. Die Abfragen werden so oft durchgeführt wie vom Nutzer in der Eingabedatei definiert.

Wenn ein Testlauf mit Zufallszahlen gewünscht ist, werden nacheinander fünf Abfragen getätigt. Zuerst wird ein Punkt mit zufälligen Koordinaten erstellt und dann eine Funktion aufgerufen, die die URL für die Abfrage anfertigt. Im Messwerkzeug gibt es die Funktionen ‚prepare_URL_short‘ und ‚prepare_URL‘. Zum Stand dieser Arbeit ist die erste Funktion im Produktiveinsatz. In diesen Funktionen werden die Koordinaten angepasst. Dabei werden die Leerzeichen durch Kommata ersetzt. Je nachdem welcher Typ mitgegeben wird, gibt die Funktion eine dafür passende URL zurück. Die zweite Funktion besitzt den gleichen Aufbau und versucht die Abfrage mit Intersects zu gestalten. Danach werden die Koordinaten nach jedem Leerzeichen getrennt und einzeln in einem Array gespeichert. Im nächsten Schritt wird

ausgewertet, ob die Nutzereingabe GET, POST oder beide Methoden gewählt wurde. Da für das Experiment nur GET von Verwendung war und POST im Kapitel 5.2.1 beschrieben wird, liegt der Fokus auf der Funktion ‚receive_data‘. In dieser Funktion wird das Handle mit der neuen URL abgesetzt. Diesmal nicht als GetCapabilities sondern als GetFeature. Das bedeutet die Anfrage wird mit konkreten Koordinaten durchgeführt. Nach diesem Prozess wird dann statt eines Punktes eine Linie mit zwei Zufallskordinaten im gleichen Schema abgefragt. Danach folgen noch eine Bounding Box, ein Multipunkt und eine Multilinie.

Wenn ein Testlauf ohne Zufallspunkte durchgeführt werden soll, wird der gleiche Prozess ohne Zufallskordinaten gestartet. Es startet wieder mit einem Punkt, jedoch kommen die Koordinaten aus der Eingabedatei, das bedeutet der Nutzer gibt diese vor. Die Abfolge des Sendens der Abfrage folgt dem gleichen Prinzip wie im vorherigen Prozess. Die fünf Geometrietypen werden als Standard von den meisten WFS deutschlandweit unterstützt und eignen sich aus diesem Grund für solch einen Testlauf. Da der Testlauf für dieses Experiment nicht genutzt wurde, wird in Kapitel 5.2.3 ein Anwendungsszenario erläutert.

Wenn kein Testlauf durchgeführt werden soll, aber Zufallspunkte generiert werden, dann erfolgt der Ablauf des Experiments. Hierbei wird die Funktion ‚receive_data‘ dreimal hintereinander ausgeführt. Bei jedem Durchlauf wird die Größe der Bounding Box gesteigert. Beim ersten Aufruf wird eine Bounding Box erstellt, die eine Seitenlänge von 100 Metern aufweist. Dies entspricht in diesem Experimentaufbau einer Baufläche. In der Funktion ‚make_bbox‘ werden Berechnungen durchgeführt, um Koordinaten zu erzeugen, die eine BBox erstellen. Diese Box liegt in den angebenen Grenzen der großen Bounding Box und die Seitenlänge beträgt 100 Meter. Dieser Prozess ist nur möglich, wenn das angegebene Koordinatensystem mit Metern rechnen kann. Bei Koordinatensystemen, die auf Gradwerten aufbauen, werden die Berechnungen zu falschen Ergebnissen führen. Mit diesen berechneten Koordinaten wird dann die URL erstellt und die Abfrage an den WFS durchgeführt. Nachdem die Antwort des Web Feature Service eingetroffen ist, wird die Seitenlänge auf 1000 Meter gesetzt. Dies soll ein Stadtgebiet darstellen. Dabei wird ebenfalls eine Bounding Box neu berechnet. Zum Schluss wird die Seitenlänge auf 10000m gesetzt. Dies entspricht einer Region von 100km². Diese Abfrage liefert die meisten Objekte zurück. Erst nachdem alle drei Abfragen durchgeführt wurden, werden neue Zufallspunkte generiert. Die Boxen weisen alle den gleichen Mittelpunkt auf.

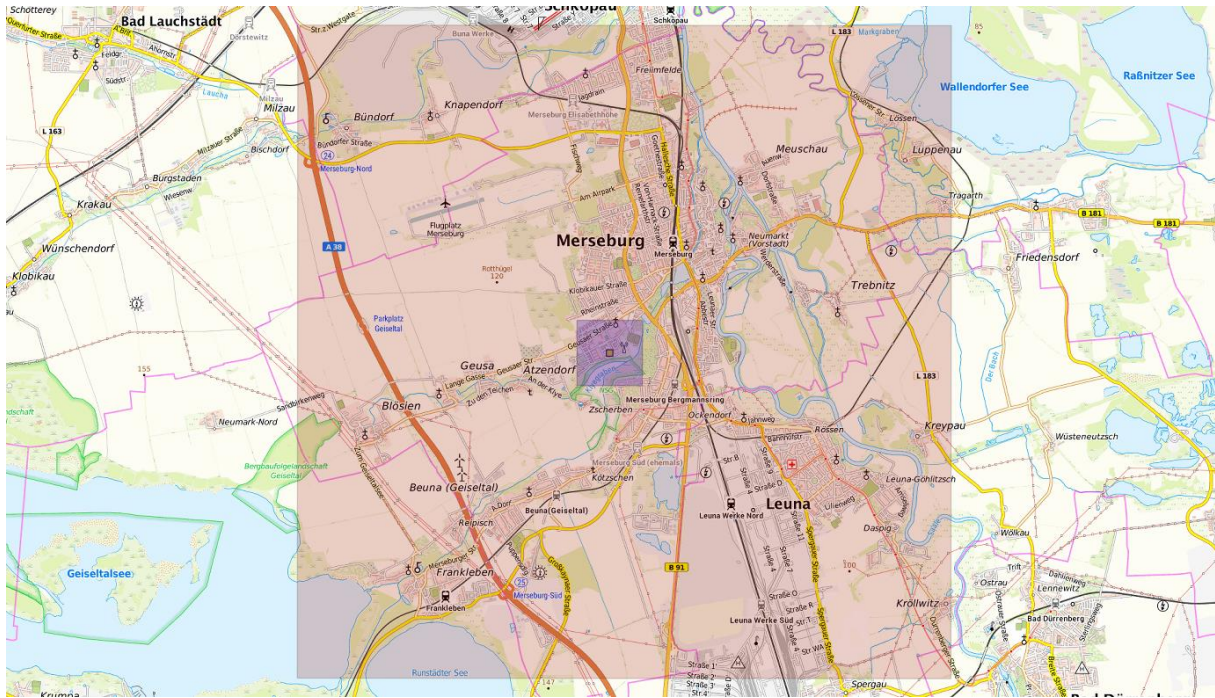


Abbildung 7: Bounding Boxen für eine Abfrage⁸⁰

In Abbildung 7 ist die Stadt Merseburg und die nähere Umgebung abgebildet. Diese Abfrage wurde mittels QGIS visualisiert. In der Mitte der Abbildung befindet sich ein Punkt, der aufgrund der Skalierung der Karte nicht sichtbar ist. Um diesen Punkt ist die erste Bounding Box, die orangefarben in der Mitte der Abbildung eingezeichnet ist. Sie hat eine Seitenlänge von 100 Metern. Dies ist die erste Abfrage. Danach wird eine weitere Box gebildet, die eine größere Umgebung abdeckt. Sie ist violett eingefärbt und enthält die kleinere Box. Zum Schluss wird die größte BBox mit einer Seitenlänge von 10000 Metern gebildet, sie ist in Abbildung 7 transparent rot dargestellt. Für die Abfragen werden Quadrate als Umkreis benutzt, weil Kreise in der Berechnung komplizierter wären und nicht so gut von den Diensten unterstützt werden. Wenn weder Testlauf noch Zufallsgenerierung gewünscht sind, wird der letzte Zweig gewählt. Das Messwerkzeug bildet wie bei den anderen Szenarien auch eine passende URL. Diesmal werden die Koordinaten des Nutzers verwendet. Die drei Unterscheidungen nach der HTTP-Methode werden ebenfalls durchgeführt, sodass die Abfrage an GET oder POST angepasst und ausgeführt wird. Nachdem dieser Prozess durchgeführt wurde, wird eine XML-Datei mit Ergebnissen der Abfrage beschrieben. In Kapitel 4.9.3 wird diese Datei genauer erklärt. Wenn das Beschreiben abgeschlossen ist, wird der gesamte Prozess nach dem Parsing für jeden in den GetCapabilities gefundenen Typename wiederholt.

⁸⁰ Eigene Darstellung

4.9 Ausgabedateien des Messwerkzeugs

Das Messwerkzeug generiert im Normalfall drei und in speziellen Fällen vier Dateien. In diesem Abschnitt werden alle Dateien näher erläutert.

4.9.1 Textdatei

Als Softwareentwickler ist es wichtig spezifische Zustände von Variablen auszulesen. Da das Messwerkzeug eine lange Laufzeit haben kann, ist es von Vorteil diese Variablen und deren Zustände in einer Datei für die spätere Auswertung zu speichern. Die Datei enthält eine Kopfzeile, die schreibend erstellt wird. Das bedeutet, dass die Textdatei mit jedem neuen Start des Messwerkzeugs überschrieben wird. Damit ist sichergestellt, dass nur der neueste Durchlauf gespeichert wird. Für die Auswertung spart dies Zeit. Die Kopfzeile enthält die URL, die Koordinaten, die Version, die HTTP-Methode und den EPSG-Wert. Nachdem die gewünschte Abfrage durchgeführt wurde, wird das Ergebnis menschenlesbar in die Datei geschrieben. Der Inhalt ist variabel und wird vom Softwareentwickler bestimmt. Als Standard sind die Zeiten des Aufrufs aufgelistet und darunter weitere Daten wie der Statuscode, die Uploadgröße und die Anzahl der zurückgegebenen Objekte. Diese Ausgabe ist jedoch variabel und kann sehr einfach erweitert werden. Dadurch kann eine spezifische Variable über den kompletten Zeitraum der Messungen abgespeichert und im Nachhinein ausgewertet werden. Somit muss die Person, die das Verhalten der Variable auswerten möchte, den Durchlauf des Skripts nicht beobachten. Die Funktion ‚write_txt‘ kann dafür beliebig an die jeweiligen Anforderungen angepasst werden. Soll keine Textdatei erstellt werden, weil dies für den Produktivbetrieb nicht notwendig ist, muss diese Funktion auskommentiert oder gelöscht werden. Die Informationen werden aus zwei Gründen nicht in die CSV-Datei geschrieben. Zum einen wegen der Übersichtlichkeit. Eine CSV-Datei muss einem einheitlichen Aufbau folgen. Wenn die Datei jedoch nur in einem einfachen Editor geöffnet wird, ist sie schnell sehr unübersichtlich und es nimmt Zeit in Anspruch die Variablen zuzuordnen. Zum anderen kann es leichter übersehen werden, dass in die CSV-Datei noch Variablen geschrieben werden, die gar nicht gespeichert werden sollen. Das kann für Verwirrung und Ärger bei Nutzern führen. Deswegen ist es besser die produktive Datei für die richtigen Messergebnisse und die Datei mit den zu überprüfenden Werten zu trennen. Die Zahlenwerte werden außerdem durch eine weitere Funktion an den europäischen Standard angepasst. Das bedeutet statt einem Punkt wird ein Komma als Dezimaltrennzeichen verwendet.

4.9.2 CSV-Datei

Das Messwerkzeug liefert wichtige Informationen aus den Messungen zurück. Diese müssen abgespeichert werden. Das geschieht über die CSV-Datei (Comma-separated values), die vom Skript erstellt und befüllt wird. Die Kopfzeile dieser Datei wird schreibend erstellt, das

bedeutet, dass bei einer bestehenden Datei die Daten überschrieben werden. Dadurch ist sichergestellt, dass verschiedene Anfragen getrennt bleiben. Es kann jedoch auch sinnvoll sein die Datei fortlaufend zu beschreiben. Während des Testens, ob alles sauber abgespeichert wird, traten Probleme mit Excel auf. In der CSV-Datei sind viele Werte, die zeitliche Messungen darstellen. Die Spaltennamen enthalten dann den Wert time. Bei diesen Messwerten war das Trennzeichen ein Punkt. Bei CSV-Dateien ist der Separator standardmäßig ein Komma. Dementsprechend wurden die ersten Versuche mit Punkten als Trennzeichen bei Zahlen und Kommata als Trennzeichen für Excel durchgeführt. Dies war jedoch fehlerbehaftet, da Excel manche Zeiten nicht richtig interpretierte und dadurch falsche Werte anzeigte. Aus diesem Grund gibt es eine Funktion die Punkte durch Kommata ersetzt. Dadurch ist das Komma als Trennzeichen für die CSV-Datei nicht mehr valide. Es würden dann falsche Trennungen entstehen. Deswegen ist das neue Trennzeichen ein Semikolon, was auch von Excel als gültiges Trennzeichen unterstützt wird. Die Kopfzeile enthält 24 Spaltennamen und dementsprechend Variablen im Messwerkzeug, die im Folgenden aufgelistet, übersetzt und erklärt werden:

1. **Number** steht für den momentanen Wert der Durchläufe. Der Wert reicht von eins bis zur vom Nutzer eingegebenen Zahl aus der Eingabedatei. Dieser Parameter wird dann erhöht, wenn die Messungen abgeschlossen sind und die Wartezeit vorüber ist. Es können mehrere Messungen zum gleichen Wert von number enthalten sein.
2. **Date** ist das Datum im Format DD.MM.YYYY und dient der schnellen Orientierung bei Messungen über mehrere Tage, Monate oder Jahre. Außerdem lassen sich damit Auswertungen über das Verhalten des WFS bei verschiedenen Tagen, Wochen, Monaten, Jahreszeiten oder Jahren erstellen.
3. **Start time** speichert die Uhrzeit zu der die Messung gestartet und durchgeführt wird. Diese Zeit wird im Format HH:MM:SS für leichte Vergleichbarkeit abgelegt. Dadurch können Untersuchungen über das Antwortverhalten des WFS bei verschiedenen Uhrzeiten durchgeführt werden. An der Uhrzeit kann die Wartezeit des Messwerkzeugs ausgerechnet werden. Die Laufzeit der Abfrage findet sich auch in diesen Werten wieder.
4. **Returned objects** bezeichnet die Anzahl der zurückgelieferten Objekte nach einer Abfrage. Dieser Wert kommt aus der Antwort des WFS und muss ausgelesen werden. Bei dem Wert null wurden keine Objekte zurückgeliefert. Bei einem Wert von fünf wurden fünf unterschiedliche Objekte zurückgegeben. Bei unerwarteten Fehlern nimmt der Wert auch null an. Dies kann der Fall sein, wenn der WFS nicht antwortet.
5. **Uploadsize** steht für die Gesamtmenge an Bytes, die über die Anfrage gesendet werden. Es ist Bestandteil der Curl-Funktion und wird daher nur ausgelesen. Dieser

Wert entspricht der Größe des Bodys beziehungsweise der Daten, die übertragen wurden. Der Header ist darin nicht enthalten.⁸¹

6. **Nameslookup_time** ist wie die folgenden sechs Parameter ein Ergebnis aus einer Funktion von Curl. Der Wert gibt die Zeit in Sekunden an, die vom Start bis zum Abschluss der Namensauflösung vergeht.⁸²
7. **Connect_time** steht für die Zeit in Sekunden, die vom Start bis zum Abschluss der TCP-Verbindung zum entfernten Host (oder Proxy) vergeht.⁸³
8. **Appconnect_time** ist die Zeit in Sekunden, die vom Start bis zum Abschluss der SSL/SSH-Verbindung/des Handshakes mit dem entfernten Server vergeht.⁸⁴
9. **Pretransfer_time** beschreibt die Zeit in Sekunden, die vom Start bis zum Beginn der Dateiübertragung vergeht. Dies schließt alle Befehle und Verhandlungen vor der Übertragung ein, die für das betreffende Protokoll spezifisch sind.⁸⁵
10. **Redirect_time** ist die Zeit in Sekunden, die für alle Umleitungsschritte einschließlich Namenssuche, Verbindung, Vorübertragung und Übertragung benötigt wird, bevor die endgültige Transaktion gestartet wird. Der Wert zeigt die gesamte Ausführungszeit für mehrere Umleitungen. Bei GET bleibt das Ergebnis null Sekunden.⁸⁶ Aus diesem Grund ist dieser Parameter ausschließlich bei POST relevant.
11. **Starttransfer_time** entspricht der Zeit in Sekunden, die vom Start bis zum Empfang des ersten Bytes verging. Dies beinhaltet `pretransfer_time` und auch die Zeit, die der Server zur Berechnung des Ergebnisses benötigt hat.⁸⁷
12. **Total_time** ist die Gesamtzeit, in Sekunden, die der gesamte Vorgang gedauert hat.
13. **Measured_time** ist eine von der Curl-Funktion unabhängige Messung der Dauer. Diese Variable misst vom Start der Abfrage bis zu dem Zeitpunkt, an dem das Ergebnis im Programmcode zur Verfügung steht. Der gemessene Wert muss aus diesem Grund dem `total_time` aus der Curl-Funktion ähneln.
14. **Process_time** bezeichnet eine intern berechnete Variable. Sie wird als Differenz zwischen der Startzeit des Transfers und der Zeit für Vorübertragungen abgespeichert. Dieser Wert gibt an, wie viel Zeit tatsächlich für den Transfer der eigentlichen Daten benötigt wurde, nachdem alle Vorübertragungen abgeschlossen

⁸¹ Vgl. Stenberg o.D., Absatz 11.

⁸² Vgl. Stenberg o.D., Absatz 11.

⁸³ Vgl. Stenberg o.D., Absatz 11.

⁸⁴ Vgl. Stenberg o.D., Absatz 11.

⁸⁵ Vgl. Stenberg o.D., Absatz 11.

⁸⁶ Vgl. Stenberg o.D., Absatz 11.

⁸⁷ Vgl. Stenberg o.D., Absatz 11.

sind. Ein niedriger Wert deutet darauf hin, dass nur eine geringe oder keine Vorabübertragung erforderlich war und der Datenübertragungsprozess schnell begonnen hat. Ein höherer Wert kann darauf hindeuten, dass mehr Zeit für Vorabübertragungen benötigt wurde, bevor der eigentliche Datenfluss begann.

15. **Method** steht für die HTTP-Methode. Das Messwerkzeug gibt hier ausschließlich GET oder POST zurück. Wenn bei der Eingabedatei beides gewählt wird, ist das in der CSV-Datei nur durch einen POST direkt nach einer GET-Abfrage zu erkennen.
16. **Status code** entspricht dem zurückgelieferten, dreistelligen HTTP-Statuscode aus der Curl-Abfrage. Dieser Code wird für jede einzelne Abfrage neu erstellt und an das Messwerkzeug gesendet.⁸⁸
17. **Call** bezeichnet die Operationen des WFS. Die drei Operationen sind GetCapabilities, DescribeFeatureType und GetFeature. Da die Abfrage des DescribeFeatureTypes im Messwerkzeug nur beispielhaft angedeutet und auskommentiert ist, wird dieser Wert in der CSV-Datei zum Stand der Arbeit nicht auftauchen. Dementsprechend finden sich die Werte GetCapabilities und GetFeature in dieser Spalte und zeigen an welcher Aufruf durchgeführt wird. Dieser Wert ist unabhängig von der Eingabedatei und wird vom Messwerkzeug selbst bestimmt.
18. **Type** ist der Geometrietyt aus der Eingabedatei, der abgefragt wird.
19. **Coordinates** sind die eingegebenen Koordinaten aus der Eingabedatei und daher unabhängig von den zufallsgenerierten Punkten. Werden keine Koordinaten angegeben, weil das nicht für die Abfrage benötigt wird oder ein Prozess fehlgeschlagen ist, wird ein Bindestrich generiert. Dies hat den Grund, dass ein leeres Feld Nutzer verunsichern kann. Außerdem erleichtert der Bindestrich die Auswertung der Felder mit Software wie beispielsweise Excel.
20. **Typename** ist ein Identifikator für den Dienst. Dadurch wird es Nutzern ermöglicht Featuretypes zu identifizieren und darauf zuzugreifen.
21. **Version** enthält die eingegebene Version des WFS aus der Eingabedatei.
22. **EPSG** bezeichnet das gewünschte Koordinatensystem. Es ist darauf zu achten, dass es, wenn Zufallspunkte gewünscht und der Testlauf unerwünscht ist, ein Koordinatensystem sein muss, welches auf Metern basiert.
23. **Short URL** ist die URL aus der Eingabedatei, sodass der abgefragte WFS weiterhin bekannt bleibt und verschiedene WFS in der gleichen Datei unterschieden werden können.

⁸⁸ Vgl. Bressoud und White 2020, S. 635.

24. **Random_point** entspricht dem generierten Zufallspunkt. Wird kein Zufallspunkt generiert, weil das aus der Eingabedatei so hervorgeht, ein Prozess fehlgeschlagen ist oder keine Zufallskordinaten benötigt werden, wird ein Bindestrich generiert. Dies hat den Grund, dass ansonsten der Eindruck entstehen kann, dass das Skript keinen Zufallspunkt generieren konnte.

4.9.3 XML-Datei

Das Messwerkzeug erstellt eine XML-Datei, die das Ergebnis des WFS abspeichert. Sie wird so benannt wie vom Nutzer bestimmt. Diese Datei hat im Gegensatz zur CSV-Datei keine Kopfzeile und wird aus diesem Grund anhängend beschrieben. Das bedeutet, dass zusätzliche Informationen, also Messungen, an das Dateiende angehängt werden. Eine Kopfzeile für diese Datei ist nicht von Nöten, da Metainformationen wie das Datum der Antwort und die zurückgelieferten Objekte in der Antwort bereits enthalten sind. Dadurch kann die Datei nahezu endlos mit weiteren Datensätzen erweitert werden. Zu dem Zeitpunkt des Experiments befinden sich Trennlinien und Leerzeilen zwischen den Antworten. Dies wird mit Hilfe der Kommentarfunktion aus XML bewerkstelligt. Das soll das Lesen und Überfliegen der Datei erleichtern. Wenn die Datei mit einer Integrated Development Environment (IDE) oder einem Editor geöffnet wird, können durch Syntax-Highlighting Kommentare durch ihre spezielle Färbung leicht erkannt werden. Für eine weitere Verwendung der Datei, bei der der Inhalt maschinell geparkt wird, können weitere Anpassungen dahingehend durchgeführt werden. Damit können entweder eigene Trennungszeichen erstellt oder diese komplett entfernt werden. Mit den Ergebnissen innerhalb der Datei können weitere Informationen ausgelesen werden. Ein wichtiges Beispiel ist der Name des betroffenen Gebiets. Da diese Datei anhängend beschrieben wird, muss auf die Anzahl der Messungen und die abgefragten Koordinaten beziehungsweise Bounding Boxen geachtet werden. Der Grund ist, dass diese beiden Parameter einen großen Einfluss auf die Größe der Antworten des WFS haben. Je größer die Antwort wird desto mehr wird abgespeichert. Das führt dazu, dass die Datei je nach Messung und Konfiguration groß werden kann.

4.9.4 XML-Datei aus der POST-Abfrage

Bei einem POST werden Daten an den Dienst gesendet. Eine Datei, in der diese Daten enthalten sind, kann an den WFS geschickt werden. Wenn der Benutzer eine eigene XML-Datei hat mit der der WFS abgefragt werden soll, muss diese angegeben werden. Sie kann zur Einfachheit im gleichen Verzeichnis wie das Messwerkzeug liegen. Dann muss nur der Name der Datei angegeben werden. Es ist aber auch möglich den kompletten Dateipfad, statt nur den Namen der Datei in die Eingabedatei einzutragen. Ansonsten kann die Datei nicht gefunden und zur Messung verwendet werden. Für eine valide XML-Datei, die vom WFS

akzeptiert und verarbeitet werden kann, sind einige Parameter wichtig. Die Art des Requests, also beispielsweise GetFeature, der Service und die Version müssen angegeben werden. Je nach der Abfrage sollten noch der Typename, die Koordinaten, der Geometrietyp und das EPSG angegeben werden. Wird POST in der Eingabedatei verlangt und keine XML-Datei mitgeliefert, dann versucht das Messwerkzeug aus den anderen Parametern der Eingabedatei eine sinnvolle Abfrage zu bilden und diese als XML-Datei abzuspeichern. Dadurch können auch ohne Kenntnisse über XML/GML oder WFS POST-Abfragen durchgeführt werden.

5 Durchführungsphase

Wenn ein Experiment entworfen und geplant wurde, muss es durchgeführt werden, um die Daten zu sammeln, die analysiert werden sollen. Dies ist die Phase der Durchführung eines Experiments. In der operativen Phase eines Experiments werden die Behandlungen an den Versuchspersonen oder Messungen durchgeführt. Experimente in der Softwaretechnik haben in den meisten Fällen Schnittstellen mit Menschen, obwohl es möglich ist, rein technologieorientierte Experimente durchzuführen. Die operative Phase eines Experiments besteht aus drei Schritten: Vorbereitung, bei der die Versuchspersonen ausgewählt und Formulare usw. vorbereitet werden, Durchführung, bei der Daten gesammelt werden, und Datenvalidierung, bei der die gesammelten Daten validiert werden.⁸⁹

Das Experiment dieser Arbeit ist technologieorientiert und hat zum Zeitpunkt der Durchführung nur eine Schnittstelle mit dem Experimentator/Softwareentwickler. Dadurch entfällt die Erstellung von Formularen und die Befragung von Versuchspersonen. Es ist auch möglich die Vorgaben für die Eingabedatei automatisiert zu generieren. Die Ergebnisse der Messungen können auch mit Testpersonen auf Verständlichkeit und Nutzerfreundlichkeit über Fragebögen in einem neuen Experiment ausgewertet werden.

5.1 Vorbereitungen

Bevor das Experiment tatsächlich durchgeführt wird, müssen einige Vorbereitungen getroffen werden. Je besser diese Vorbereitungen getroffen werden, desto einfacher wird die Durchführung des Experiments. Bei der Vorbereitung gibt es zwei wichtige Aspekte. Der erste besteht darin, die Teilnehmer auszuwählen und zu informieren, und der zweite darin, Material wie Formulare und Werkzeuge vorzubereiten. Da für dieses Experiment keine Versuchsteilnehmer benötigt werden, weil das Messwerkzeug automatisch laufen soll, entfällt der erste Aspekt der Vorbereitungen in der Durchführungsphase. Bevor der Versuch durchgeführt werden kann, müssen alle Versuchsinstrumente bereitstehen. Dazu gehören z. B. die Versuchsobjekte, Richtlinien für den Versuch sowie Messformulare und -werkzeuge. Welche Instrumente benötigt werden, hängt von der Konzeption des Experiments und der Methode der Datenerhebung ab. Wenn alle Vorbereitungen abgeschlossen sind, kann die Durchführung des Experiments beginnen.⁹⁰

⁸⁹ Vgl. Wohlin et al. 2012, S. 117.

⁹⁰ Vgl. Wohlin et al. 2012, S. 117–119.

5.2 Ausführung durch Anwendungsfälle

Das Experiment kann auf verschiedene Weise durchgeführt werden. Einige Experimente werden jedoch über einen viel längeren Zeitraum durchgeführt, und es ist unmöglich, dass der Versuchsleiter an jedem Detail des Versuchs und der Datenerhebung teilnehmen kann. Die Daten können entweder manuell von den Teilnehmern durch Ausfüllen von Formularen, manuell unterstützt durch Instrumente, in Interviews oder automatisch durch Instrumente erhoben werden. Ein Nachteil einer langen Messung ist, dass der Experimentator keine Möglichkeit hat, Ungereimtheiten, Unsicherheiten und Fehler in den Dateien usw. direkt aufzudecken. Diese Art von Fehlern kann erst nach der Datenerhebung aufgezeigt werden.⁹¹

Die weiteren Abschnitte sind Anwendungsfälle des Messwerkzeugs. Sie sollen die Zweiteilung für die Nutzung von Messungen und für die Automatisierung eines Prozesses aufzeigen. Der erste Anwendungsfall ist, dass das Skript auch die HTTP-Methode POST verwenden kann. Dabei kann überprüft werden, ob der abgefragte Dienst POST-Anfragen verarbeiten kann. Der nächste Abschnitt befasst sich mit einem konkreten Beispiel bei dem genaue Koordinaten abgefragt werden sollen. Dies zeigt das Potential als Automatisierungstool. Da es nur eine Reihe aufeinanderfolgender Abfragen gibt und keine Wartezeit wie bei den Messungen, ist der Prozess viel schneller als ein manueller Vergleich mit grafischer Software oder eine manuelle Abfrage des WFS. Im nächsten Kapitel wird der Testlauf aktiviert und erklärt. Damit können mehrere Geometrietypen in einem kurzen Zeitraum überprüft werden. Der darauffolgende Abschnitt zeigt, dass das entwickelte Werkzeug auch international verwendet werden kann. Dabei wird ein WFS außerhalb von Deutschland und Europa abgefragt. Als letzter Anwendungsfall wird das Experiment zur Datenerhebung für die statistische Auswertung zur Beantwortung der Hypothesen aufgeführt. Dabei werden viele Datensätze mit einem Start des Messwerkzeugs abgefragt und in Dateien gespeichert.

5.2.1 POST

Neben einer einfachen GET-Abfrage gibt es bei einem Web Feature Service auch die Möglichkeit einen POST durchzuführen. Dies hat den Vorteil, dass die URL dafür nicht bearbeitet werden muss. Für einen Benutzer, der die Abfrage händisch durchführen muss, ist die GET-Abfrage simpel. Die fertige URL mit allen benötigten Parametern wird in einen beliebigen Browser eingegeben und der WFS bearbeitet dann die Abfrage. Das Ergebnis wird sofort sichtbar für den Nutzer ausgeliefert. Die Durchführung eines POSTs ist für einen Nutzer jedoch viel schwerer. Es bedarf hierbei eines weiteren Tools. Es gibt Webseiten, die online

⁹¹ Vgl. Wohlin et al. 2012, S. 120.

POST-Abfragen durchführen, jedoch nur im key-value-Schema und nicht über Dateien. Mit dem Skript ist es aber auch möglich Dateien unkompliziert per POST an den WFS zu senden. Die zu sendende XML-Datei wird dafür in den gleichen Pfad wie das Messwerkzeug abgelegt und in der Eingabedatei benannt. Danach können POST-Abfragen an einen WFS durchgeführt werden.

Für Benutzer, die mit XML oder dem WFS nicht vertraut sind, bietet das Skript noch eine weitere bequeme Funktion. Es ist möglich, ohne eine eigene XML-Datei einen POST abzuschicken. Hierbei wird überprüft, ob eine Datei vorliegt. Wenn das nicht der Fall ist, da beispielsweise durch einen Tippfehler der Dateiname nicht übereinstimmt oder vergessen wurde die Datei in den Pfad abzulegen, wird eine eigene XML-Datei generiert. Möchte ein in diesem Bereich unerfahrener Nutzer POST-Abfragen durchführen, ist dies auch eine Möglichkeit. Dadurch ist sichergestellt, dass in solchen Situationen trotzdem valide Ergebnisse vom WFS zurückkommen. Die automatisch generierte XML-Datei trägt den eingegebenen Dateinamen aus der Eingabedatei mit dem Kürzel ‚_xml‘ dahinter. Dadurch ist sichergestellt, dass die Ergebnisdatei und die Datei zur Abfrage unterschiedliche Namen aufweisen. In der XML-Datei sollten einige Parameter enthalten sein, damit der WFS die Anfrage sinnvoll bearbeiten kann. Das sind beispielsweise die Koordinaten und der Geometriotyp. Außerdem enthalten sein sollten die Version, die Operation und der Typename. Wenn das Skript eine eigene Datei erstellen muss, wird ein simples XML aufgebaut und mit den Daten aus der Eingabedatei befüllt und angepasst. Für den Geowebsservice sieht das dann folgendermaßen aus:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <wfs:GetFeature service="WFS" version="2.0.0">
3   <wfs:Query typeName="bfnsch_Schutzgebiet:Landschaftsschutzgebiete">
4     <ogc:Filter xmlns:ogc="http://www.opengis.net/ogc">
5       <ogc:BBOX>
6         <ogc:PropertyName>SHAPE</ogc:PropertyName>
7         <gml:Envelope xmlns:gml="http://www.opengis.net/gml" srsName="EPSG:25832">
8           <gml:lowerCorner>439435 5889500</gml:lowerCorner>
9           <gml:upperCorner>422628 5891386</gml:upperCorner>
10        </gml:Envelope>
11      </ogc:BBOX>
12    </ogc:Filter>
13  </wfs:Query>
14 </wfs:GetFeature>
```

Abbildung 8: automatisch generierte XML-Datei vom Messwerkzeug⁹²

In Abbildung 8 ist der Aufbau einer vom Messwerkzeug generierten XML-Datei zu sehen. Die vier Zeilen vor der Query wurden aus Platzgründen und Gehaltlosigkeit für diese Grafik

⁹² Eigene Darstellung

entfernt. Darin sind XML namespaces (xmlns), zu Deutsch XML-Namensräume, definiert. Damit werden Elemente und Attribute eindeutig identifiziert. Der abgefragte WFS ist der Geodienst des BfN. In der zweiten Zeile ist die Abfrage, nämlich GetFeature, zu erkennen. Wie bei einer Abfrage per URL sind die anderen Parameter ebenfalls zu übergeben wie die Version und der Service. Ab Zeile drei beginnt die Abfrage. In dieser Zeile ist auch der spezifische Typename enthalten. Für dieses Beispiel wurden die Landschaftsschutzgebiete wie in Zeile drei in der Abbildung 8 zu sehen ist, abgefragt. In Zeile fünf ist der abzufragende Geometriotyp zu sehen. Das Skript passt diesen automatisch an die Nutzereingabe an. In Zeile sieben ist das Koordinatensystem über das EPSG zu sehen. Darunter sind dann die Koordinaten der abzufragenden Bounding Box abgebildet. Diese sind in lower und upper Corner unterteilt. Das Messwerkzeug hat die meisten variablen Daten aus der Eingabedatei entnommen und die Koordinaten per Zufallsfunktion generiert. Es ist ebenfalls möglich die Koordinaten vorzugeben, indem der Parameter der Zufallsgeneration in der Eingabedatei den Wert false aufweist. Diese XML-Datei wird für jeden neuen Typename und jede neue Abfrage überschrieben. Sie dient als Hilfe für eine simple und trotzdem sinnvolle POST-Abfrage. Die Datei wird gespeichert, um Nutzern eine Vorlage zu bieten, die erweitert werden kann. Es ist möglich komplexe Bedingungen zu konzipieren wie beispielsweise mehrere Filter mit Vergleichen und Wildcards. Dies kann auch direkt im Programmcode angepasst werden, sodass aus den Standardabfragen komplexere XML-Dateien werden, mit denen das Skript den WFS abfragt. Es ist ebenfalls möglich die Datei nicht zu speichern, um Ressourcen zu sparen.

Die Implementierung von möglichen POST-Requests ist wichtig, um die WFS weiter zu testen. Aber auch, weil POST viele weitere Vorteile bietet, die für Geodaten von Bedeutung sind. Ein Vorteil ist die Unterstützung größerer Datenmengen, da diese im Body übertragen werden statt an der URL angehängt. Gerade bei komplexeren Anfragen mit vielen Koordinaten, Parametern und Filtern sollte auf POST zurückgegriffen werden. Die URL verliert durch viele Parameter die Übersichtlichkeit und lässt sich dementsprechend schwieriger lesen und erweitern. Auch die URL-Encodierung von Browsern kann zu unschönen Veränderungen führen, indem Sonderzeichen durch die Prozentdarstellung ersetzt werden. Nach der Abfrage werden alle Sonderzeichen wie beispielsweise Leerzeichen oder spitze Klammern ausgetauscht. Das erschwert die Les- und Wartbarkeit zusätzlich. POST-Abfragen bieten den Vorteil, dass verschiedene Inhaltsarten wie XML, JSON oder binäre Daten verschickt werden können. Durch diese Flexibilität können komplexe Datenstrukturen leichter übertragen werden.⁹³

⁹³ Vgl. Bressoud und White 2020, S. 632–633.

Tabelle 2: gekürztes Abfrageergebnis einer POST-Abfrage

Zeile	returned objects	upload size	method	status code	call
1	0	0	GET	200	getCapabilities
2	0	892	POST	200	GetFeature
3	3	895	POST	200	GetFeature
4	0	886	POST	200	GetFeature
5	0	888	POST	200	GetFeature
6	0	881	POST	200	GetFeature
7	0	878	POST	200	GetFeature
8	3	886	POST	200	GetFeature
9	2	892	POST	200	GetFeature

Quelle: Eigene Darstellung

In Tabelle 2 sind die fünf wichtigsten Spalten von der Abfrage zu sehen. Die zurückgelieferten Objekte sind vom Featuretype abhängig. Es ist zu erkennen, dass insgesamt acht Objekte zurückgegeben wurden. In der Spalte mit dem Namen upload size ist die Größe der versendeten Daten aufgelistet. Sie ist durch die variable Anpassung nicht konstant. Bei den GetCapabilities findet eine GET-Operation statt. Aus diesem Grund ist die übermittelte Datenmenge null wie in der Spalte upload size zu erkennen ist. Alle Zeilen weisen den Statuscode 200 auf, was bedeutet, dass alle Abfragen ohne Probleme abgearbeitet wurden.

5.2.2 Hochschule Merseburg

Das Messwerkzeug soll nicht nur wiederholbare, langanhaltende wissenschaftliche Messungen durchführen können, sondern auch Messungen für den tatsächlichen Gebrauch. Hierbei ist es wichtig, dass das Skript schnell durchläuft und nach einem Durchgang stoppt. Für diesen Anwendungsfall ist der WFS wieder der Geodienst des BfN. Der Zeitabstand zwischen den Messungen beträgt null, da nur eine Messung für diesen Anwendungsfall stattfindet. Die Koordinaten sind für dieses Beispiel vorgegeben. Die Koordinaten wurden mittels QGIS⁹⁴ ermittelt und über GoogleMaps⁹⁵ verifiziert. Die Version bleibt 2.0.0 und die HTTP-Abfrage ist GET. Um die Funktionalität und den Funktionsumfang des Messwerkzeugs zu demonstrieren, sind die Koordinaten nicht wie bei den anderen Experimenten in Längeneinheiten, sondern in Winkeleinheiten vorgegeben. Damit ist der EPSG-Code 4326. Der Geometrietyp bleibt eine Bounding Box. Testlauf und Zufallspunkt generieren sind deaktiviert. Mit diesen Parametern in der Eingabedatei kann der Messvorgang gestartet

⁹⁴ <https://www.qgis.org/de/site/>

⁹⁵ <https://www.google.com/maps>

werden. Die Koordinaten bilden hierbei eine Bounding Box, die das Hauptgebäude der Hochschule Merseburg einschließt. Über QGIS kann die BBox der Abfrage visualisiert werden:

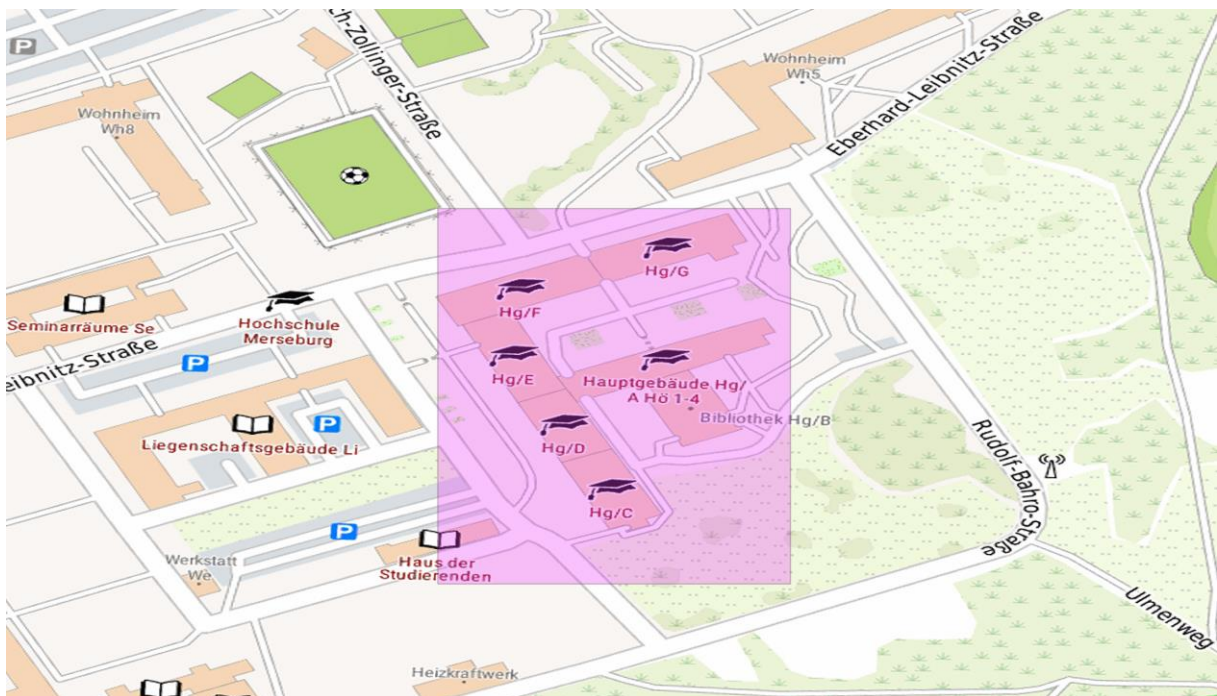


Abbildung 9: Bounding Box mit dem Hauptgebäude der Hochschule Merseburg ⁹⁶

Es wird die Hochschule Merseburg über ein System, welches Koordinaten verarbeiten kann (beispielweise Google Maps) ausgewählt und ein Koordinatenpaar erzeugt. Diese konkreten Koordinaten können nun das Skript über die Eingabedatei weitergegeben werden. Dadurch wird eine Box, wie in Abbildung 9 zu sehen, generiert, die der WFS verarbeiten kann. Sie umfasst das Hauptgebäude der Hochschule Merseburg. Das bedeutet, dass keine Schutzgebiete in der Abfrage auftauchen sollten.

Nachdem alle Parameter an die Eingabedatei übergeben wurden, kann die Messung ausgeführt werden. Es gibt zwei wichtige Dateien, die nach der Ausführung abgelegt werden. Zum einen die XML-Datei in der die Antworten des WFS gespeichert wird. In dieser Datei finden sich verschiedene Parameter und Informationen wieder. Ein Beispiel ist die Anzahl der zurückgesendeten Objekte, die sich im Kopf der XML befindet. Danach werden die Objekte mit ihren jeweiligen Informationen aufgelistet. Aus der Antwort kann beispielsweise auch der Name des betroffenen Naturschutzgebietes entnommen werden, falls die Koordinaten ein solches Gebiet schneiden beziehungsweise enthalten. In der CSV-Datei befinden sich wichtige Daten wie die Zeiten und der HTTP-Statuscode. Durch diesen kann abgeleitet

⁹⁶ Eigene Darstellung

werden, ob eine Anfrage erfolgreich ist. Nach dem Import in Excel und der Beseitigung einiger Spalten, können für diesen Anwendungsfall wichtige Schlussfolgerungen getroffen werden.

Tabelle 3: gekürztes Abfrageergebnis der Hochschule Merseburg

Zeile	returned objects	status code	call	typename
1	0	200	getCapabilities	-
2	0	200	GetFeature	bf_n_sch_Schutzgebiet:Nationale_Naturmonumente
3	0	200	GetFeature	bf_n_sch_Schutzgebiet:Fauna_Flora_Habitat_Gebiete
4	0	200	GetFeature	bf_n_sch_Schutzgebiet:Vogelschutzgebiete
5	0	200	GetFeature	bf_n_sch_Schutzgebiet:Biosphaerenreservate
6	0	200	GetFeature	bf_n_sch_Schutzgebiet:Nationalparke
7	0	200	GetFeature	bf_n_sch_Schutzgebiet:Naturparke
8	0	200	GetFeature	bf_n_sch_Schutzgebiet:Naturschutzgebiete
9	0	200	GetFeature	bf_n_sch_Schutzgebiet:Landschaftsschutzgebiete

Quelle: Eigene Darstellung

In Tabelle 3 sind die vier wichtigsten Spalten von der Abfrage zu sehen. Die zurückgelieferten Objekte sind in jeder Zeile null. Das bedeutet, dass zu jedem abgefragtem Typenamen kein Gebiet in der abgefragten BBox liegt. Die angegebenen Typenamen haben das zu überprüfende Gebiet jeweils im Namen. In Zeile vier wird beispielsweise auf Vogelschutzgebiete geprüft. In der Spalte call wird der Requesttyp aufgelistet. Die Abfrage beginnt jeweils mit einem GetCapabilities, um die Typenamen auszulesen. Aus diesem Grund gibt es in der ersten Zeile keinen Typenamen. Da der abgefragte WFS acht verschiedene Typenamen aufweist, werden acht GetFeature-Abfragen durchgeführt. Damit wird sichergestellt, dass die zu überprüfende Bounding Box für jedes Schutzgebiet getestet wird. In der Spalte status code werden die HTTP-Statuscodes aufgelistet. In diesem Fall kam bei jeder Abfrage der Code 200 zurück. Das bedeutet, dass die Anfrage erfolgreich ist. Bei einem GET mit einem Statuscode von 200 wird eine Repräsentation der angefragten Zielressource geliefert. Dementsprechend waren alle Abfragen erfolgreich und lieferten verwertbare Ergebnisse.

5.2.3 Aktivierung des Testlaufs

In diesem Abschnitt wird die Funktion des Testlaufs erklärt und anhand eines Beispiels veranschaulicht. Diese Funktion ist für das Sammeln von vielen Datensätzen auf einmal gedacht. Sie wird über die Eingabedatei über den Parameter trialrun aktiviert. Im Quellcode sind für diesen Parameter fünf Geometrietypen vorgesehen, die nacheinander abgearbeitet

werden. Diese Geometrietypen sind so gewählt, dass sie von allen oder den meisten WFS unterstützt werden. Für dieses Beispiel wird der Testlauf und die Zufallspunktgenerierung aktiviert. In Tabelle 4 befindet sich ein Ausschnitt aus der Abfrage.

Tabelle 4: gekürztes Abfrageergebnis des Testlaufs

Zeile	returned objects	call	type	coordinates
1	0	getCapabilities	BBox	-
2	8	GetFeature	Punkt	5735620 831535
3	8	GetFeature	Linie	5735620 831535 5889456 797430
4	0	GetFeature	BBox	5735620 831535 5889456 797430
5	8	GetFeature	Multipunkt	5735620 831535 5889456 797430
6	8	GetFeature	Multilinie	5735620 831535 5889456 797430

Quelle: Eigene Darstellung

In Tabelle 4 sind die vier wichtigsten Spalten der Abfrage zu sehen. In diesem Beispiel ist die Spalte `type` sehr wichtig. Sie zeigt, dass die Geometrietypen sich bei jeder Abfrage ändern. Dabei ist in diesem Ausschnitt der Typename gleich. Es werden die nationalen Naturmonumente abgefragt. Die Koordinaten passen sich ebenfalls an den Geometrietyp an und es werden über die Zufallsgeneration mehr Koordinaten erzeugt, wenn dies benötigt wird. Die zurückgelieferten Objekte unterscheiden sich in der Anzahl. Alle Geometrietypen außer die Bounding Box zeigen acht Objekte an. Mithilfe von QGIS wurde diese Abfrage und die nationalen Naturmonumente visualisiert. Dabei war zu erkennen, dass die gewählten Koordinaten kein Monument schneiden. Das bedeutet, die Anzahl der BBox ist korrekt. Aus diesem Grund wird für das Experiment ausschließlich die Bounding Box verwendet.

5.2.4 Abfrage eines ausländischen Web Feature Service

Dieses Kapitel soll zeigen, dass das implementierte Messwerkzeug auch WFS außerhalb von Deutschland abfragen kann. Durch die Standardisierung des OGC sind die Dienste im Ausland trotzdem messbar. Dieses Kapitel soll beweisen, dass auch außerhalb von Europa Web Feature Services automatisch abgefragt werden können. Hierfür wurde ein WFS aus den USA ausgewählt. Dieser WFS verfügt über Daten von aktiven Minen und Mineralienbetrieben in den USA⁹⁷ und stellt damit einen internationalen Dienst dar. Die enthaltenen Daten sind von Bergwerksanlagen und -betrieben für Rohstoffe, die vom National Minerals Information Center des United States Geological Survey (USGS) überwacht werden. Dies ist eine Behörde für amtliche Kartographie in den USA. Die Abfrage enthält einen Durchgang, bei dem die

⁹⁷ <https://mrdata.usgs.gov/mineplant/>

komplette Fläche des WFS per Bounding Box abgefragt wird. Dadurch werden alle Objekte, die der WFS enthält zurückgegeben.

Tabelle 5: gekürztes Abfrageergebnis des ausländischen WFS

Zeile	returned objects	status code	call	typename	measured time
1	0	200	getCapabilities	-	30,975443
2	6785	200	GetFeature	ms:mineplant	28,468866

Quelle: Eigene Darstellung

In Tabelle 5 sind fünf wichtige Spalten aus der CSV-Datei zu sehen. Außerdem sind es zwei Zeilen mit Daten aus den Abfragen. Das liegt daran, dass der getestete WFS nur einen einzigen Typenamen anbietet. Die Zeile 1 zeigt den Aufruf der GetCapabilities an. Dabei kommen keine Geoobjekte vom WFS zurück. Die Anfrage war jedoch, aufgrund des 200er Statuscodes, erfolgreich. Bei einer Abfrage der GetCapabilities wird kein Typename mitgegeben, deshalb ist dieses Feld leer. Die gemessene Systemzeit liegt bei dieser Abfrage bei fast 31 Sekunden. Danach wird in Zeile 2 die Operation GetFeature durchgeführt. Die Anzahl der zurückgegebenen Objekte beträgt 6785. Der Statuscode ist 200 und damit ist die Anfrage serverseitig erfolgreich. Der Typename lautet ‚ms:mineplant‘. Damit kann abgelesen werden, dass es sich um Bergbau handelt. Die gemessene Systemzeit beträgt 28,47 Sekunden. Das liegt an der großen Menge an Geodatensätzen, die der WFS zurückgibt. Jedoch musste das Messwerkzeug für eine erfolgreiche Messung angepasst werden. Bei diesem WFS heißt die Bounding Box nicht wie bei deutschen WFS üblich BBox sondern Box. Aus diesem Grund ist die erste Anfrage fehlgeschlagen. Nach der Anpassung im Quellcode liefen die Abfragen ohne Probleme erfolgreich ab. Durch diesen Versuch wird gezeigt, dass das Messwerkzeug nicht nur deutsche oder europäische, sondern weltweit WFS abfragen kann.

5.2.5 Experiment

Für diese Arbeit ist es notwendig, dass das Messwerkzeug Messungen über lange Zeiträume mit bestimmten zeitlichen Abständen dazwischen generiert. Die Aufgabe ist es einen ausgewählten WFS für einen bestimmten Zeitraum abzufragen und die Ergebnisse zu speichern. Außerdem ist es notwendig, dass weitere Daten wie Messzeiten und der Erfolg der Messung vorliegen. Der zu benutzende WFS ist der Geodienst des BfN. In Deutschland gibt es für jedes Bundesland einen eigenen WFS. Jedoch ist durch die Zufallsgenerierung der Koordinaten und die Größe der Umgebungs-BBox ein deutschlandweiter WFS besser geeignet. Die Zeit zwischen den Messungen beträgt eine Stunde, also 60 Minuten. Durch den zeitlich beschränkten Rahmen dieser Arbeit soll die Messung eine Woche laufen. Dadurch sind alle Wochentage enthalten und durch die stündliche Abfrage werden trotzdem genug

Datensätze für eine Auswertung generiert. Dadurch kann errechnet werden wie viele Durchläufe benötigt werden. Eine Woche mit sieben Tagen multipliziert mit 24 Stunden eines Tages ergeben 168 Durchläufe. Da im Vorfeld beim Testen von Abfragen die erste oder zweite Messung abbrach, wurde der Parameter auf 170 aufgerundet, um etwaige Abbruchfehler zu ersetzen. Außerdem wird die erste Messung sofort, also ohne Wartezeit durchgeführt. Aus diesem Grund würde die Messung eine Stunde eher abgeschlossen als erwartet. Im Kapitel 6.4 wird dieser Fehler erklärt und die Lösung erklärt. Die Koordinaten sollen zufällig generiert werden, daher bleibt der Wert leer. Die Version ist 2.0.0 und damit der aktuelle Standard zum Zeitpunkt des Experiments. Die HTTP-Methode ist GET. Der EPSG-Wert ist 25832 und damit ein Koordinatensystem, welches Metereingaben verarbeiten kann. Dies wird in der Abbildung 10 visualisiert.

```
1 * URL to be requested:
2 https://geodienste.bfn.de/ogc/wfs/schutzgebiet?
3
4 * time interval between measurements:
5 60
6
7 * number of runs:
8 1
9
10 coordinates separated with spaces:
11 504671 5730581 505671 5729581
12
13 version:
14 2.0.0
15
16 * http method:
17 get
18
19 EPSG:
20 EPSG:25832
```

Abbildung 10: Eingabedatei Teil 1 ⁹⁸

Der Pfad zur XML-Datei ist leer, da GET durchgeführt wird. Der Geometrietyp ist Bounding Box. Damit können Flächen wie beispielsweise Bauflächen getestet werden. Der Testlauf ist für dieses Experiment nicht nötig und wird daher auf false gesetzt. Die Dateinamen müssen vorgegeben werden. Die Umgebungs-BBox ist eine große BBox in der die Zufallspunkte generiert werden. Sie soll so gewählt werden, dass es keine Überschneidungen mit Nachbarländern oder der Ost- beziehungsweise Nordsee gibt. Dadurch gibt es zwei mögliche Bounding Boxen. Entweder ein langes, aber schmales Rechteck von Süden (zum Beispiel Stuttgart) bis Norden (z.B. Hamburg). Oder eine breite BBox von Osten (z.B. Dresden) nach Westen (z.B. Düsseldorf). Für diesen Versuch wird zweitere gewählt. Der erste Punkt dieser

⁹⁸ Eigene Darstellung

BBox hat die Koordinaten East: 839435m und North: 5669500m. Der zweite Punkt hat die Werte East: 422628m und North: 5891386m. Daraus ergibt sich eine Fläche von circa 106.000km². Damit nimmt die BBox rund 30% der Gesamtfläche von Deutschland ein. Zufallspunkte sollen generiert werden und deswegen wird die Zeile mit yes beschrieben. Damit ist sichergestellt, dass verschiedene Koordinaten und damit Ergebnisse für die Auswertung zur Verfügung stehen. Außerdem wird es durch die zufällige Auswahl der Punkte ein richtiges Experiment wie in Kapitel 3 erläutert. Es findet eine Randomisierung der Variablen der Koordinaten und damit auch der Bounding Boxen statt. In Abbildung 11 ist der zweite Teil der Eingabedatei für das Experiment zu sehen.

```
22  path to the xml file:
23  post.xml
24
25  * geometry type:
26  BBox
27
28  trial run?:
29
30
31  * name of the files:
32  filename
33
34  boundaries of the BoundingBox:
35  839435 5669500 422628 5891386
36
37  randomly generate centre point?:
38  yes
```

Abbildung 11: Eingabedatei Teil 2 ⁹⁹

Die Zeile 21 ist entfernt, da sie eine Leerzeile zur Trennung der Parameter zur besseren Lesbarkeit in der Eingabedatei darstellt. Mit dieser Eingabedatei kann das Experiment wiederholt werden. Aber durch die Zufallsfunktion der Koordinaten wird das Ergebnis jedes Mal unterschiedlich sein. Diese Messung wurde am Mittwoch, den 28.08.2023 17:20 Uhr gestartet. Am Mittwoch, den 30.08.2023 gegen 18:33 Uhr wurde das Skript erfolgreich beendet. Insgesamt sind es 4130 Datensätze, die in der CSV-Datei abgespeichert sind. Einige Spalten sind für die statistische Auswertung nicht relevant. Aus diesem Grund wurde die Exceldatei vor dem Import in Statistica¹⁰⁰ angepasst. Nur die Variablen, die für die Hypothesen überprüft werden müssen, werden importiert. Davor wird noch eine neue Spalte namens working time erstellt, die die Werte null und eins annehmen kann. Dieser Parameter gibt an, ob sich die Messung in der Kernarbeitszeit befindet oder nicht und bildet damit einen wichtigen

⁹⁹ Eigene Darstellung

¹⁰⁰ <https://www.statistica.com/en/software/tibco-data-science/-/tibco-statistica>

Bestandteil der dritten Hypothese. Die null steht hierbei für außerhalb der Arbeitszeit und die eins für innerhalb. Danach können die Berechnungen für die Statistik beginnen.

5.3 Überprüfung der Daten

Nach der Datenerhebung muss der Versuchsleiter prüfen, ob die Daten angemessen sind und ob sie korrekt erhoben wurden. Es ist wichtig zu überprüfen, ob das Experiment tatsächlich so durchgeführt wurde, wie es geplant war. Bei der Datenvalidierung kann es vorkommen, dass ein oder mehrere Probanden das Experiment sehr früh verlassen, und ihre Datenformulare müssen sorgfältig überprüft werden, um sicherzustellen, dass sie die Formulare in angemessener Weise ausgefüllt haben. Wenn es zu Problemen gekommen ist, kann es sein, dass die Daten von einem oder mehreren Messungen entfernt werden müssen.¹⁰¹ Bei der Durchführung des Experiments kam es zu keinen Problemen und es können alle Daten für die Auswertung genutzt werden.

¹⁰¹ Vgl. Wohlin et al. 2012, S. 121.

6 Analyse und Interpretation

Nach der Durchführung des Experiments fließen die gesammelten Daten in die Analyse und Interpretation ein. Nach der Phase der Datenerhebung stehen experimentelle Daten zur Verfügung. Das Ziel ist auf Grundlage dieser Datenreihen Schlussfolgerungen zu ziehen. Damit gültige Schlussfolgerungen gezogen werden können, müssen die Daten des Experiments richtig interpretiert werden. Die quantitative Auswertung kann in drei Schritten erfolgen. Im ersten Schritt werden die Daten mit Hilfe der deskriptiven Statistik charakterisiert, die die zentrale Tendenz, die Streuung usw. sichtbar macht. Im zweiten Schritt werden anormale oder falsche Datenpunkte ausgeschlossen, wodurch der Datensatz auf eine Reihe gültiger Datenpunkte reduziert wird. Im dritten Schritt werden die Daten durch Hypothesentests analysiert, bei denen die Hypothesen des Experiments mit einem bestimmten Signifikanzniveau statistisch ausgewertet werden.¹⁰²

6.1 Deskriptive statistische Auswertung

Die deskriptive Statistik befasst sich mit der Darstellung und numerischen Verarbeitung eines Datensatzes. Nach der Erhebung von Versuchsdaten kann die deskriptive Statistik dazu verwendet werden, interessante Aspekte des Datensatzes zu beschreiben und grafisch darzustellen. Das Ziel der deskriptiven Statistik ist es, ein Gefühl dafür zu bekommen, wie der Datensatz verteilt ist.¹⁰³

Für die erste Hypothese wird eine weitere Spalte in Excel angelegt. In dieser kann abgelesen werden, ob der Statuscode und die zurückgegebene Anzahl der Objekte valide sind. Eine eins steht hierbei für wahr, also beides ist erfolgreich und eine null steht dafür, dass eine oder beide Bedingungen nicht erfüllt werden. Die Formel prüft beim Statuscode auf die 200 und die Anzahl der zurückgegebenen Objekte muss größer oder gleich null sein.

Nachdem die Messung beendet wurde, können die Datensätze ausgewertet werden. Der Versuch startete an einem Mittwochabend und der Messprozess lief eine Woche wie in Kapitel 5.2.5 beschrieben. In dieser Zeit wurden insgesamt 4130 Datensätze zurückgesendet und in einer Datei abgespeichert. Diese Daten werden dann in Excel bearbeitet. Dabei werden Spalten mit Variablen gelöscht, die nicht relevant für die Beantwortung der Hypothesen sind. Danach werden in Excel zwei weitere Spalten manuell erstellt. Diese beinhalten die Arbeitszeit

¹⁰² Vgl. Wohlin et al. 2012, S. 123.

¹⁰³ Vgl. Wohlin et al. 2012, S. 123.

und die Variablenauswertung für die erste Hypothese als dichotome bzw. binäre Merkmale. Sie können also nur die Werte null und eins annehmen.

Tabelle 6: Deskriptive Statistik Teil 1

Variable	Deskriptive Statistik (Geodienste_Abfrage in Geodienste_Statistica - Ergebnis)				
	Gült. N	Mittelw.	Minimum	Maximum	Stdabw.
start time	4130	0,4933	0,014641	0,9806	0,285966
returned objects	4130	0,5610	0,000000	31,0000	1,559751
measured time	4130	0,1852	0,098226	10,0956	0,547536
status code	4130	199,5157	0,000000	200,0000	9,830625
working time	4130	0,2542	0,000000	1,0000	0,435484

Quelle: Eigene Darstellung

In Tabelle 6, die mit Statistica 13 erstellt wurde, sind in der linken Spalte die Namen der Variablen aufgelistet. Diese Namen und Werte kommen aus der CSV-Datei, die beim Durchlauf erstellt wurde. Die Spalten bezeichnen die berechneten Maße. Statistica 13 bietet für die deskriptive Statistik 33 Maße zur Berechnung an. Für die Auswertung in dieser Arbeit werden jedoch nicht so viele Kennzahlen benötigt. Für die Berechnungen werden nur die Variablen ausgewählt, die im Rahmen der zu untersuchenden Hypothesen von Bedeutung sind. In der nächsten Spalte befindet sich die gültige Anzahl und bezieht sich auf die Anzahl der tatsächlich verfügbaren und verwendeten Datenpunkte. In einigen Fällen können Daten fehlen oder unvollständig sein, was dazu führt, dass bestimmte Datenpunkte nicht für die Berechnungen berücksichtigt werden können. Der Mittelwert ist das arithmetische Mittel der gültigen Werte einer Variable. Die Spalte Minimum steht für den kleinsten Wert, den die Variable in der Datenreihe annimmt. Das Maximum steht dementsprechend für den größten Wert einer Variable im untersuchten Datensatz. In der letzten Spalte befindet sich die Standardabweichung der Variablen.

Die Variable measured time weist bei der gültigen Anzahl den Wert 4130 auf, was bedeutet, dass 4130 Datenpunkte in die Analyse einbezogen wurden. Der Mittelwert dieser Variable beträgt 0,1852. Das gibt den durchschnittlichen Wert der Variable an. Das Minimum ist 0,098226 und das Maximum ist 10,0956, was die Bandbreite der Werte zeigt, die von der Variable angenommen werden. Die Standardabweichung von 0,547536 gibt an, wie stark die Werte um den Mittelwert streuen. Je größer die Standardabweichung, desto größer ist die Streuung der Datenpunkte um den Mittelwert.

Ein weiterer Tabellenausschnitt mit fünf wichtigen Maßen aus Statistica 13:

Tabelle 7: Deskriptive Statistik Teil 2

Variable	Deskriptive Statistik				
	Median	Modus	Häufigk. f. Modus	Summe	Spannweite
start time	0,4812	Mehrere	10	2037,4	0,966
returned objects	0	0	3100	2317	31
measured time	0,1224	Mehrere	3	765	9,9974
status code	200	200	4120	824000	200
working time	0	0	3080	1050	1

Quelle: Eigene Darstellung

Die Variablen sind gleichbleibend wie bei Tabelle 6. Als erstes statistisches Maß in Tabelle 7 ist der Median zu sehen. Der Median ist ein statistischer Wert, der die Mitte einer geordneten Datenreihe repräsentiert. In der Spalte rechts daneben befindet sich der Modus. Dadurch wird der am häufigsten vorkommende Wert einer Datenreihe ermittelt. Die Häufigkeit für den Modus gibt an, wie oft der am häufigsten vorkommende Datenwert in der Datenreihe vertreten ist. Die Summe ist das Ergebnis der Addition aller einzelnen Werte einer Datenreihe für eine Variable. Sie gibt an, wie viel insgesamt von der betrachteten Größe vorhanden ist. Bei der Spannweite in der letzten Spalte wird der Unterschied zwischen dem größten und dem kleinsten Wert der Datenreihe eingetragen.

Für die Variable measured time ergeben sich die folgenden Werte. Der Median liegt bei 0,1224. Das bedeutet, dass die Hälfte der Werte für die gemessene Systemzeit kleiner als 0,1224 Sekunden ist und die andere Hälfte größer als dieser Wert ist. Der Modus hat für dieses Experiment den Wert mehrere, was bedeutet, dass es mehrere Werte in der Datenreihe gibt, die gleich häufig vorkommen. Da der Wert ‚Mehrere‘ wenig zufriedenstellend ist, wird mit Excel der Modalwert ermittelt. Die Funktion MODALWERT ist jedoch veraltet und wurde ersetzt. Die zu verwendende Funktion lautet MODUS. Sie unterteilt sich in den einfachen Modus, bei dem nur ein Datenwert zurückgegeben wird und dem multiplen Modus, bei dem alle Werte angegeben werden. Dementsprechend muss die Funktion MODUS.VIELF genutzt werden. Excel gibt dann sieben Werte aus, die den Modus abbilden. Die Werte sind im Bereich von 0,1066 bis 0,1102 und weisen darauf hin, dass die benötigte Zeit oft in diesen Wertebereich liegt. Die Häufigkeit des Modus ist drei. Das bedeutet die sieben ermittelten Werte kommen alle dreimal vor. Die Summe beträgt bei der measured time 765 Sekunden. Excel ermittelt den

Wert 765,02. Wenn diese Sekunden in Minuten umgerechnet werden ergeben sich daraus 12 Minuten und 45 Sekunden. Das bedeutet, dass die Zeit für alle 4130 Abfragen fast 13 Minuten betrug. Die Spannweite beträgt 9,9974 Sekunden. Das bedeutet, dass der größte Wert der Datenreihe fast zehn Sekunden größer ist als der kleinste enthaltene Wert. Die Summe ist für Variablen wie start time und den Statuscodes jedoch wenig aussagekräftig. Für die Zeiten und die zurückgegebenen Objekte ist das Ergebnis der Summe für Interpretationen und Rückschlüsse wichtig.

Ein Histogramm ist ein Diagramm, das verwendet wird, um die Verteilung von Daten darzustellen. Es besteht aus einer Reihe von Säulen, wobei jede Säule eine bestimmte Datenkategorie repräsentiert und die Höhe der Säule die Häufigkeit oder den Anteil dieser Kategorie in den Daten angibt. Die Werte aus dem Experiment sind meistens sehr nah beieinander und es gibt wenig Streuung und nur einzelne Ausreißer. Aus diesem Grund sehen viele Histogramme ähnlich aus. Im Folgenden soll der Unterschied zwischen einem aus Statistica13 und einem mit Excel generierten Histogramm veranschaulicht werden.

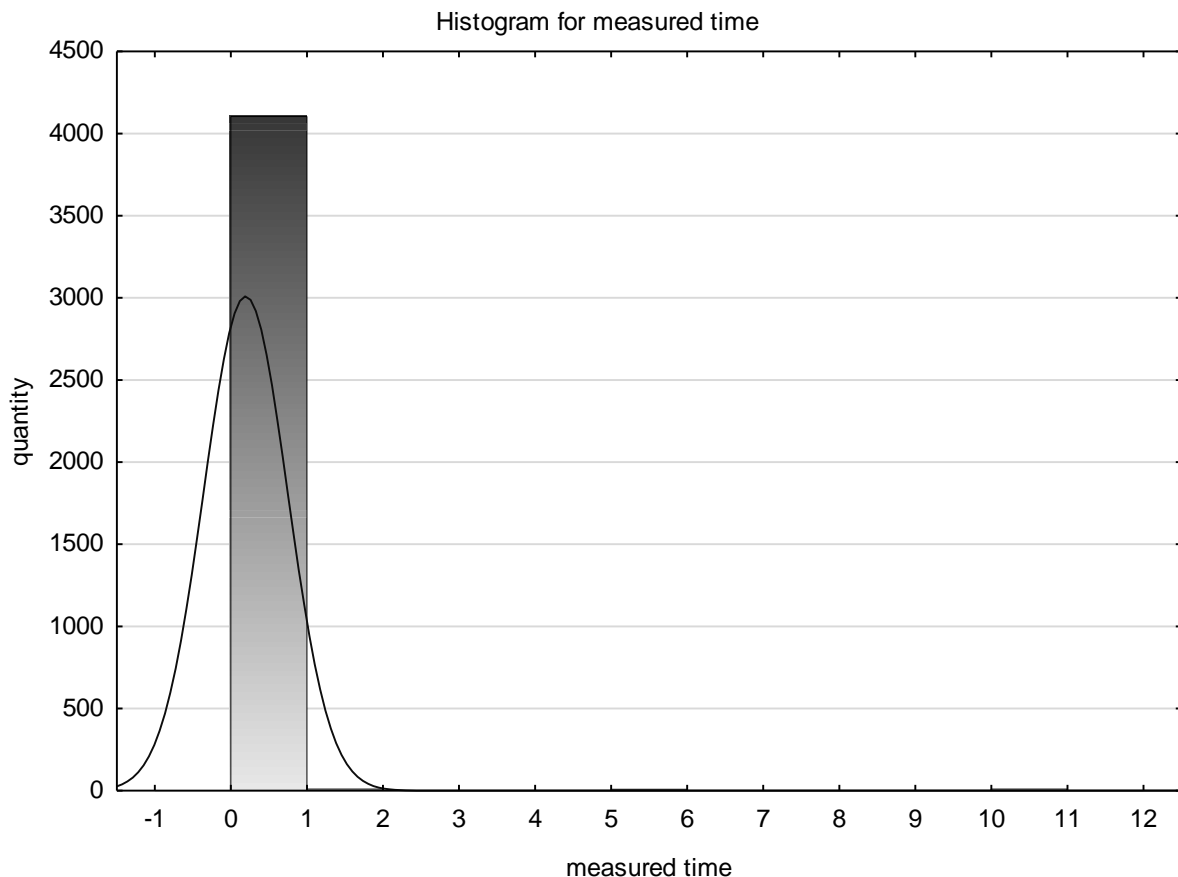


Abbildung 12: Histogramm für die gemessene Systemzeit ¹⁰⁴

¹⁰⁴ Eigene Darstellung

In Abbildung 12 wird ein Histogramm dargestellt. Es zeigt die Variable measured time. Die y-Achse zeigt die Anzahl der beobachteten Werte. Auf der x-Achse sind die Werte der gemessenen Systemzeit in Sekunden abgebildet. Es sind insgesamt vier Säulen zu sehen. Die erste Säule zwischen den Werten null und eins ist die größte und geht über 4000 bei der Anzahl. Das bedeutet der Großteil der Abfragen dauerte bis zu einer Sekunde. Neben dieser Säule befindet sich ein viel kleinerer im Wertebereich von eins bis zwei. Und dann noch in den Wertebereichen fünf bis sechs und zehn bis elf. Da diese Säulen sehr kurz sind, ist die Anzahl der Abfragen für diese Wertebereiche sehr gering. In diesem Diagramm ist ebenfalls zu erkennen, dass viele Daten um den Median herum konzentriert sind und der Mittelwert durch die Ausreißer verschoben ist. Die Daten sind in vier weit auseinander liegende Säulen eingeteilt. Das deutet auf eine Streuung hin. Daten mit drei, vier, sieben und acht Sekunden gibt es beispielsweise nicht für measured time. Die Kurve in der Abbildung ist gestreckt, das bedeutet, dass die Standardabweichung kleiner als eins ist und viele Werte in einer Kategorie sind.

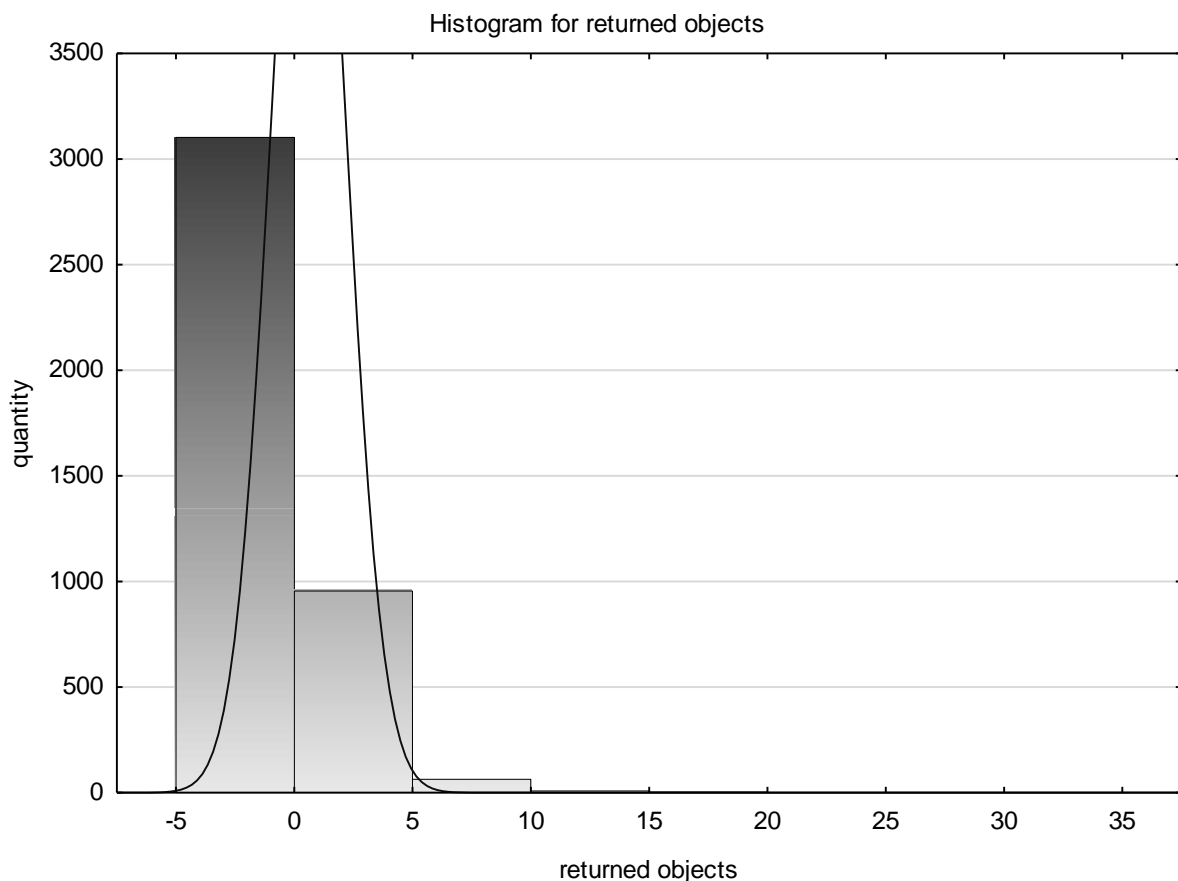


Abbildung 13: Histogramm für die Anzahl der Antwortobjekte ¹⁰⁵

¹⁰⁵ Eigene Darstellung

Ein weiteres Histogramm, welches mit Statistica erstellt wurde, ist das für die Antwortobjekte. Dabei reichen die Werte der x-Achse bis 35, weil es eine Messung mit dem Wert 31 gibt und dieser wegen der Skalierung nicht als Säule dargestellt werden kann. Das Diagramm enthält also sechs Säulen. Die erste Säule geht über den Wert 3000 auf der y-Achse. Das bedeutet, dass mehr als 3000 Abfragen null, also keine Antwortobjekte, zurückschickten. Fast 1000 Abfragen lieferten bis zu fünf Objekte zurück. Danach werden die folgenden drei Säulen sichtlich kleiner und kaum noch ablesbar. Das bedeutet, dass von den 4130 erfolgreichen Abfragen der Großteil keine bis fünf Antwortobjekte lieferte. Das liegt jedoch auch an den gewählten Distanzen für die Bounding Boxen. Die kleinste mit 100 Metern und die mittlere mit einem Kilometer Seitenlänge schneiden nur sehr selten Naturschutzgebiete. Bei dieser Variable sind die Daten ebenfalls stark um den Median und Mittelwert konzentriert. Außerdem gibt es bei diesem Histogramm Ausreißer, die weit vom erwarteten Wert entfernt sind. Diese sind jedoch nur in geringer Konzentration gegeben, ersichtlich durch die kleinen Säulen im rechten Bereich des Diagramms. Die gemessenen Daten streuen sehr wenig und konzentrieren sich eher in zwei Wertebereichen.

Im Gegensatz zu den zwei vorherigen Histogrammen, wurde das folgende mit Excel statt Statistica erstellt. Die untersuchte Variable ist die Startzeit. Im Gegensatz zu den anderen Variablen gibt es hier eine größere Streuung. Außerdem stellt dieses Diagramm die Uhrzeiten und die Anzahl der Abfragen dar. Daraus kann abgelesen werden zu welcher Uhrzeit Abfragen scheiterten.

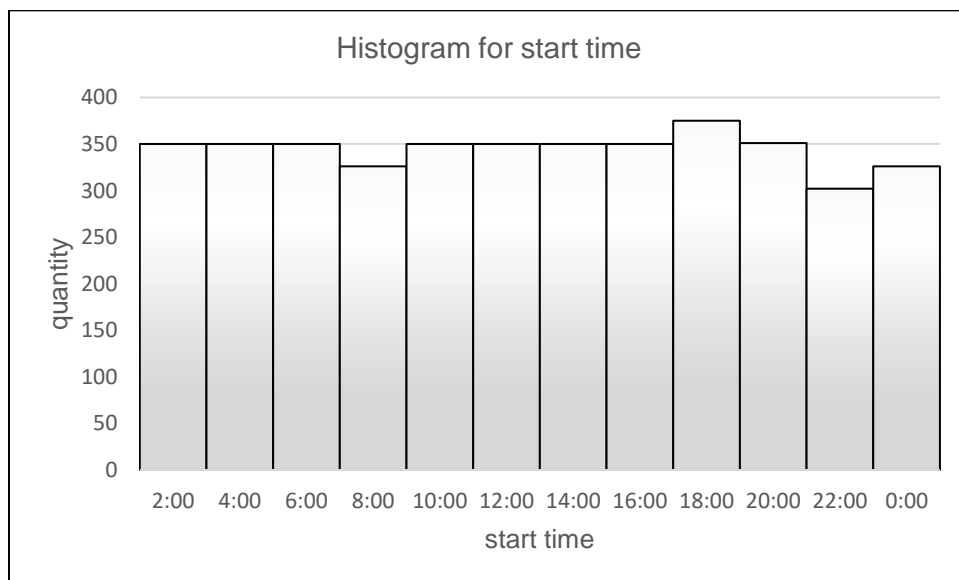


Abbildung 14: Histogramm für die Startzeit aus Excel ¹⁰⁶

¹⁰⁶ Eigene Darstellung

In der Abbildung 14 ist ein Histogramm zu sehen, das mit Excel erstellt wurde. Die Achsenbeschriftungen sind gleich. Die x-Achse bildet die Startzeiten der Messungen ab. Die Säulen bilden je zwei Stunden ab. Wenn alle Abfragen ohne Fehler erfolgreich gewesen wären, müsste jede Säule, exklusive der 18:00 Uhr Säule, den Wert 350 aufweisen. Das hat den Grund, dass zwei weitere Messvorgänge als Puffer in diesem Zeitraum eingeplant wurden. Die Säulen von 8:00, 22:00 und 0:00 Uhr sind alle niedriger als der Sollwert. Das bedeutet, dass in diesen Zeiträumen die Abfragen nicht erfolgreich waren. Dieses Diagramm veranschaulicht, dass in den drei genannten Zeiträumen Probleme auftraten. Es ist auch möglich die Kategorien weiter zu verändern. Es ist beispielsweise möglich die Kategorien, anstatt wie in Abbildung 14 aller zwei Stunden anzuzeigen, stündlich abzubilden. Das führt jedoch zu einer Verdopplung der Säulen und kann dementsprechend schnell unübersichtlich werden, wenn die Anzahl der Kategorien erhöht wird.

Für die Variable measured time wurde ein Box-Plot Diagramm erstellt. Der erste Versuch enthält alle gemessenen Werte der Datenreihe.

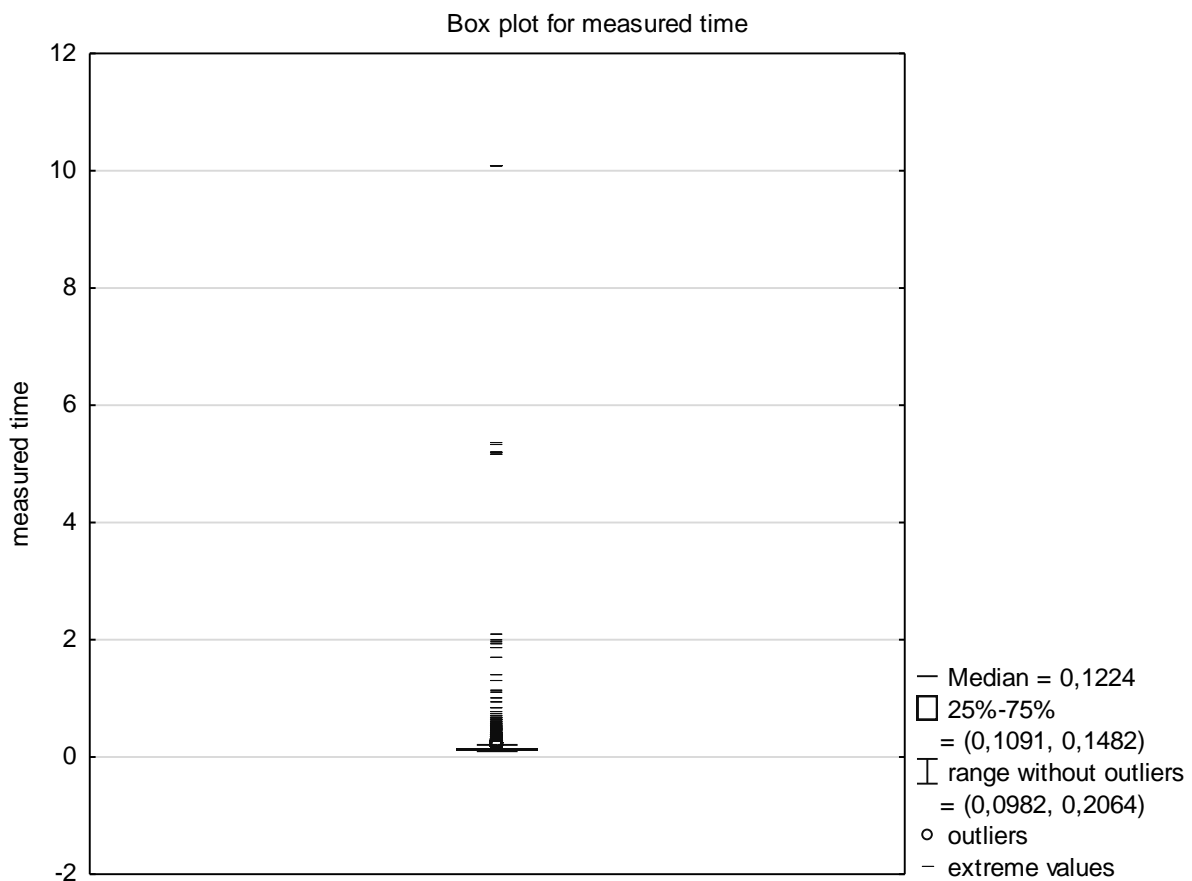


Abbildung 15: Boxplot für die gemessene Systemzeit ¹⁰⁷

¹⁰⁷ Eigene Darstellung

Es ist zu erkennen, dass die in Abbildung 15 dargestellte Box kaum sichtbar ist. Dieses Diagramm ist durch Extremwerte verschoben. Die Extremwerte beinhalten sowohl die gescheiterten Messungen mit den Werten über zehn Sekunden als auch erfolgreiche Messungen bis zu einer Sekunde. Neben Extremwerten gibt es noch Ausreißer, die als Kreise dargestellt sind, um sie von den Extremwerten zu unterscheiden. Da dieses Diagramm durch die Extremwerte stark beeinflusst ist, wird im Folgenden ein zweiter Box-Plot ohne Extremwerte gezeigt. Die Werte der y-Achse verändern sich dadurch drastisch. In Abbildung 15 gehen die Werte bis zwölf in Zweierschritten. Da die Box aber in dem Bereich von 0,1091 bis 0,1482 liegt, ist diese Einteilung ungeeignet.

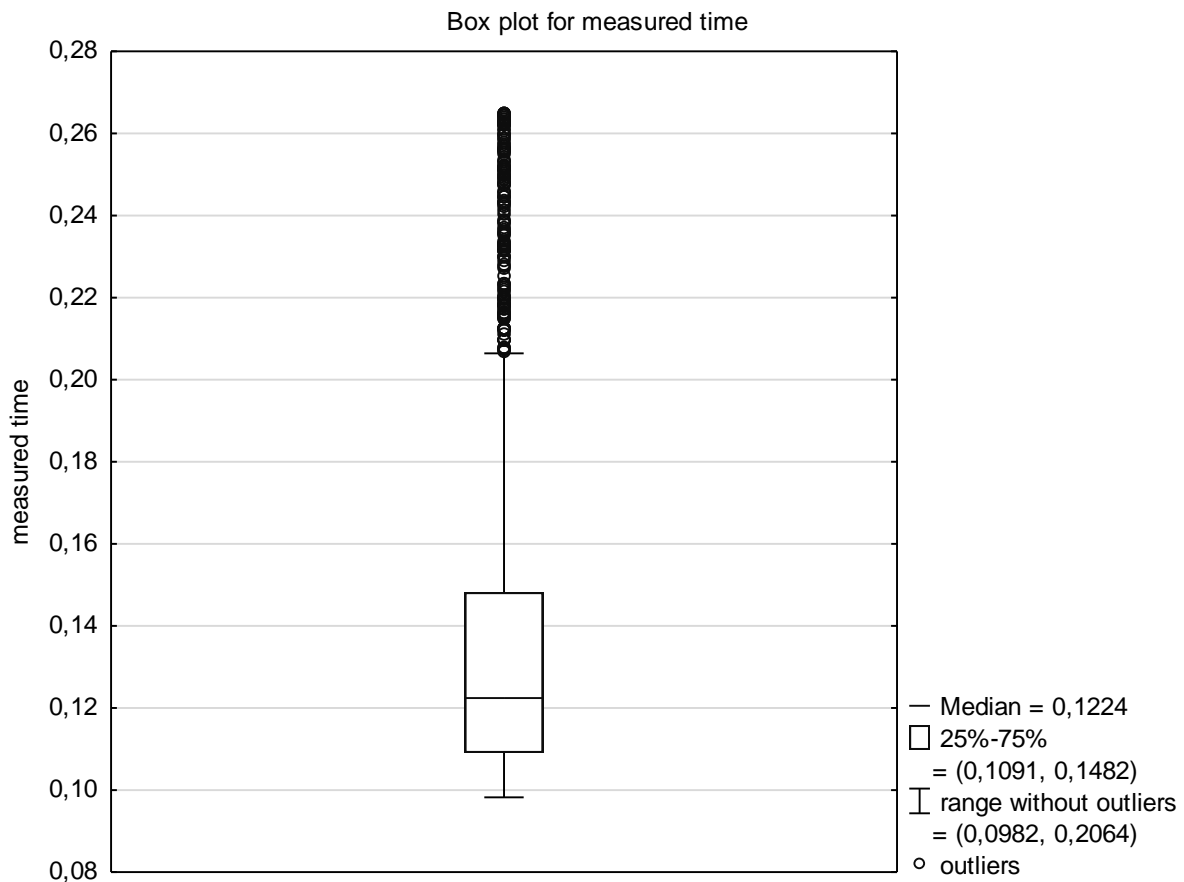


Abbildung 16: Boxplot für die gemessene Systemzeit exklusive der Extremwerte ¹⁰⁸

In der neuen Abbildung 16, bei der die Extremwerte exkludiert sind, ist die Box besser zu erkennen. Dies liegt an der neuen Einteilung der Werte auf der y-Achse. Die jetzt abgebildeten Messungen liegen alle unter 0,3 Sekunden. Die Box besteht wie aus der Legende abzulesen aus dem Interquartilabstand und wird von dem Median geteilt. Die Box repräsentiert die mittleren 50% der Daten. Die Whisker-Linien zeigen den Bereich der Daten außerhalb des

¹⁰⁸ Eigene Darstellung

Interquartilabstands an. Ausreißer sind nicht in diesen Linien enthalten. Die Punkte außerhalb der Whisker-Linien sind die Ausreißer und diese liegen im Wertebereich von 0,21 bis 0,27. Die Verteilung der Daten ist wie im Histogramm beobachtet nicht symmetrisch. Das sagen die unterschiedlich langen Whisker-Linien und die Lage der Box aus. Es ist aus Abbildung 15 zu erkennen, dass es neben Ausreißern Extremwerte gibt, was einer weiteren Untersuchung bedarf. Außerdem gibt es relativ viele Ausreißer bei dieser Variable.

Nach der univariaten Analyse, die einen Eindruck über Verteilungen und Streuung der Datenreihe gibt und diese veranschaulicht, wird im Folgenden die bivariate Analyse durchgeführt. Für die dritte Hypothese ist es notwendig die Variablen auf Korrelation zu untersuchen. In der nachfolgenden Tabelle sind die Korrelationen aus den Variablen der Hypothesen aufgeführt. In dieser Tabelle sind sowohl die Korrelationskoeffizienten als auch die p-Werte darunter enthalten. Der Korrelationskoeffizient misst die Stärke und Richtung der Beziehung zwischen den Variablen, während der p-Wert angibt, ob die Beziehung statistisch signifikant ist. Bei einem $p < 0,05$ gilt die Korrelation als statistisch signifikant und ist rot markiert.

Tabelle 8: Korrelationskoeffizienten der Variablen aus den Hypothesen

Variable	Korrelationen Markierte Korr. signifikant für $p < ,05000$ N=4130				
	start time	returned objects	measured time	status code	working time
start time	1,0000	,0005	,0053	,0043	,0100
	p= ---	p=,974	p=,732	p=,784	p=,521
returned objects	,0005	1,0000	,0538	,0177	,0014
	p=,974	p= ---	p=,001	p=,255	p=,928
measured time	,0053	,0538	1,0000	-,8910	-,0016
	p=,732	p=,001	p= ---	p=0,00	p=,916
status code	,0043	,0177	-,8910	1,0000	,0061
	p=,784	p=,255	p=0,00	p= ---	p=,693
working time	,0100	,0014	-,0016	,0061	1,0000
	p=,521	p=,928	p=,916	p=,693	p= ---

Quelle: Eigene Darstellung

In Tabelle 8, die mit Hilfe von Statistica erstellt wurde, sind die fünf Variablen der drei Hypothesen zu sehen. Die Variablen sind als Spalten- und Zeilenüberschriften gewählt und

können so einfacher und ohne Legende zugeordnet werden. Die Zellen enthalten die jeweiligen Korrelationen der beiden Variablen. Die Diagonale der Tabelle bildet den Test auf Korrelation mit der gleichen Variable. Aus diesem Grund ist die Diagonale der Tabelle der Wert 1,0. Die Variable start time weist eine sehr schwache positive Korrelation (0,0053 und damit größer als null) mit measured time auf. Diese Beziehung ist jedoch nicht statistisch signifikant, da der p-Wert zu groß ist. Bei der Variable measured time und status code handelt es sich um eine starke negative Korrelation. Diese Beziehung ist statistisch signifikant mit einem p-Wert von 0,00. Bei der für die dritten Hypothese untersuchten Korrelation zwischen working time und measured time besteht eine sehr schwache negative Korrelation, die nicht statistisch signifikant ist. Basierend auf diesen Ergebnissen kann festgestellt werden, dass die Variable measured time eine starke negative Korrelation mit status code aufweist und zwischen measured time und returned objects eine mäßige positive Korrelation besteht. Die anderen Variablen haben entweder keine signifikante Beziehung oder nur sehr schwache Beziehungen zueinander. Ein p-Wert von genau 0,000 bedeutet, dass die Wahrscheinlichkeit, dass der beobachtete Zusammenhang zwischen den Variablen auf reinen Zufall beruht, sehr gering ist.

6.2 Reduzierung des Datensatzes

Allen statistischen Methoden ist gemeinsam, dass das Ergebnis ihrer Anwendung stark von der Qualität der Eingabedaten abhängt. Wenn die Daten, auf die die statistischen Methoden angewendet werden, nicht das repräsentieren, was sie repräsentieren sollen, dann sind die Schlussfolgerungen, die aus den Ergebnissen der Methoden gezogen werden, nicht korrekt. Fehler im Datensatz können entweder als systematische Fehler oder als Ausreißer auftreten, was bedeutet, dass der Datenpunkt viel größer oder viel kleiner ist, als bei Betrachtung der anderen Datenpunkte erwartet. Eine effektive Methode zur Identifizierung von Ausreißern ist die Erstellung von Streudiagrammen oder Boxplots. Wenn Ausreißer identifiziert werden, ist es wichtig zu entscheiden, wie mit ihnen umgegangen werden soll. Dabei sollten nicht nur die Koordinaten im Diagramm berücksichtigt werden, sondern es ist wichtig, die Gründe für das Auftreten der Ausreißer zu analysieren. Wenn der Ausreißer auf ein seltsames oder seltenes Ereignis zurückzuführen ist, das nie wieder vorkommen wird, kann der Punkt ausgeschlossen werden. Wenn der Ausreißer auf ein seltenes Ereignis zurückzuführen ist, das wieder auftreten kann, ist es nicht ratsam, den Wert aus der Analyse auszuschließen, da viele relevante Informationen in dem Ausreißer enthalten sind. Im Allgemeinen ist es ratsam bei der Entfernung von Datenpunkten restriktiv vorzugehen, das heißt jede Entfernung von Daten soll gut begründet und dokumentiert sein.¹⁰⁹

¹⁰⁹ Vgl. Wohlin et al. 2012, S. 131–132.

Bei dem Experiment aus Kapitel 5.2.5 gibt es keine Ausreißer, die auf seltene Ereignisse zurückzuführen sind. Alle Ausreißer in den gespeicherten Datenreihen sind auf wiederkehrende Ereignisse zurückzuführen und daher auch für die Auswertung wichtig.

6.3 Hypothesentests

Das Ziel von Hypothesentests ist es, festzustellen, ob es möglich ist, eine bestimmte Nullhypothese auf der Grundlage einer Stichprobe aus einer bestimmten statistischen Verteilung zurückzuweisen. Das heißt, die Nullhypothese beschreibt einige Eigenschaften der Verteilung, aus der die Stichprobe gezogen wurde, und der Experimentator möchte diese Eigenschaften mit einer bestimmten Signifikanz zurückweisen.¹¹⁰

Die Nullhypothese sollte daher negativ formuliert werden, d. h. die Absicht des Tests besteht darin, die Hypothese zu verwerfen. Wird die Nullhypothese nicht abgelehnt, kann nichts über das Ergebnis gesagt werden, während bei Ablehnung der Hypothese festgestellt werden kann, dass die Hypothese mit einer bestimmten Signifikanz falsch ist. Bei der Durchführung eines Tests ist es in vielen Fällen möglich, die geringstmögliche Signifikanz (oft als p-Wert bezeichnet) zu berechnen, mit der die Nullhypothese abgelehnt werden kann.¹¹¹

6.3.1 Erste Hypothese

Nach diesen Erkenntnissen können die Hypothesen, die im vierten Kapitel aufgestellt wurden, beantwortet werden. Die erste Hypothese beschäftigt sich mit den Variablen Statuscode und der Anzahl der vom WFS zurückgegebenen Objekte.

Bei der Abfrage der Geowebsservices werden keine für das Verfahren relevanten Daten geliefert.

Mit der neu erstellten Spalte status code und time true kann die erste Hypothese beantwortet werden. Dazu müssen die false-Statements, also die Zellen mit dem Wert null, gezählt werden. Dies übernimmt die Excelfunktion ZÄHLENWENN, die auf die ganze Spalte angewandt wird. Das Ergebnis lautet zehn. Das bedeutet, dass von den 4.130 Datensätzen zehn den Anforderungen aus der ersten Hypothese nicht gerecht werden. Die zehn Datensätze entsprechen außerdem gerundet 0,242% der Gesamtanfragen. Das bedeutet, dass 99,758% der Anfragen an den WFS erfolgreich abgearbeitet wurden.

Bei der Verfügbarkeit von IT-Systemen gibt es eine Einteilung in verschiedene Klassen, die vom Bundesamt für Sicherheit in der Informationstechnik (BSI) bereitgestellt werden. Die

¹¹⁰ Vgl. Wohlin et al. 2012, S. 132.

¹¹¹ Vgl. Wohlin et al. 2012, S. 133.

Verfügbarkeitsklassen (VK) reichen von Klasse 0 bis Klasse 5. Klasse 0 bezeichnet Systeme ohne zugesicherte Verfügbarkeit und beinhaltet jegliche IT-Infrastruktur, die weniger als 99% Mindestverfügbarkeit aufweist. Die VK 1 bietet eine Mindestverfügbarkeit von 99% und wird als normale Verfügbarkeit bezeichnet. In dieser Klasse befindet sich der getestete WFS. Die Ausfallzeit des Dienstes ist somit weniger als acht Stunden im Monat und 88 Stunden pro Jahr. Ab 99,9% Mindestverfügbarkeit wird die Verfügbarkeitsklasse 2 erreicht. Die letzte VK bietet eine Verfügbarkeit unter extremen Bedingungen und ist somit Disaster-tolerant. Der überprüfte WFS ist mit den gesammelten Daten somit fast in VK 2 einzuteilen.¹¹²

Aus diesem Grund wird die Nullhypothese abgelehnt. Das bedeutet bei der Abfrage des Geowebsservices werden für das Verfahren relevanten Daten geliefert, wobei der Statuscode 200 und die Anzahl der Antwortobjekte größer gleich null ist. Dadurch ist sichergestellt, dass die Daten relevant sind. Es handelt sich nicht nur um eine Überprüfung der Erreichbarkeit des Services, sondern die Antwortdaten werden mit einbezogen. Die ermittelten Werte bezüglich der ersten Hypothese und Verfügbarkeit sind jedoch nur ein Ausschnitt und sollten mit weiteren Formeln für die Kenngrößen der Verfügbarkeit analysiert werden. Die Abfragen haben einen Abstand von einer Stunde und somit kann nicht sichergestellt werden, ob in der Zwischenzeit ein Ausfall des Services stattgefunden hat. Ein Beispiel für eine solche Kenngröße ist die ‚Mean Time to Failure‘. Dabei wird die Dauer des störungsfreien Betriebs beschrieben. Dafür benötigt es jedoch längere Messungen oder Zugriff auf den WFS selbst. Die hier bestimmten Werte können dementsprechend schwanken und sich sowohl verbessern (Aufstieg in Verfügbarkeitsklasse zwei oder höher) als auch verschlechtern (Verfügbarkeitsklasse eins wird nicht mehr erreicht).

¹¹² Vgl. Jasperneite und Lohweg 2022, S. 201–202.

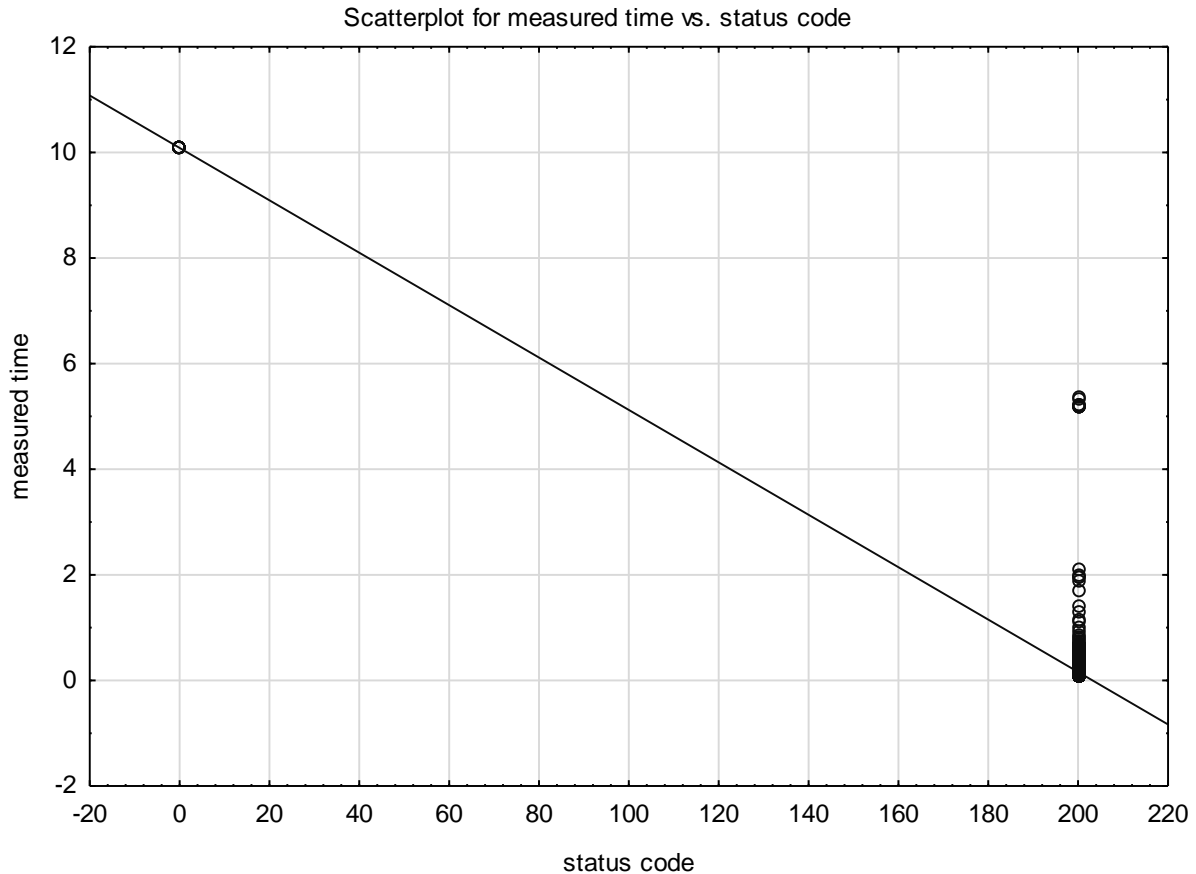


Abbildung 17: Streudiagramm mit der gemessenen Systemzeit und dem Statuscode ¹¹³

In Abbildung 17 sind die gemessene Systemzeit und der HTTP-Statuscode als Streudiagramm dargestellt. Die Datenpunkte sind in zwei Säulen eingeteilt und wenig gestreut. Der Statuscode nimmt nur die Werte 200 für eine erfolgreiche Abfrage oder null für eine gescheiterte Abfrage ein. Das Diagramm zeigt, dass bei erfolglosen Messungen die gemessene Zeit über zehn Sekunden geht. Bei erfolgreichen Abfragen liegen viele Zeiten unter 0,3 Sekunden und ein erheblicher Großteil der Messungen blieb unter zwei Sekunden. Es ist ebenfalls zu sehen, dass die Extremwerte mit ca. fünf Sekunden ebenfalls erfolgreich abgeschlossen wurden. Die Regressionslinie fällt deutlich von links oben nach rechts unten und weist damit einen starken negativen Anstieg auf. Dies ist die Visualisierung des Korrelationskoeffizienten aus der Tabelle der Korrelation. Der Wert beträgt $-0,8910$ und damit sind beide Variablen stark voneinander abhängig. Bei einer negativen Steigung weisen die Variablen eine abnehmende Beziehung auf. Das bedeutet, dass bei einer langen Zeit der Abfrage der Statuscode mit hoher Wahrscheinlichkeit null sein wird. Damit ist die Messung gescheitert. Dadurch kann ein Streudiagramm eine große Hilfe bei der Auswertung von großen Datenreihen sein.

¹¹³ Eigene Darstellung

6.3.2 Zweite Hypothese

Der überprüfte WFS ist derzeit technisch nicht in der Lage die benötigten Daten am Beispiel von Schutzgebietsdaten zu liefern.

Hierbei wird die Variable der gemessenen Systemzeit ausgewertet. Sie darf nicht größer gleich zehn Sekunden sein. Das Messwerkzeug misst sowohl die Zeit, die Curl zurückgibt als auch intern die Zeit, die die Anfrage benötigt. Die interne Zeit ist hierbei ausschlaggebend und für die Beantwortung der zweiten Forschungsfrage genutzt. Die gemessene Systemzeit wird für die Auswertungen benutzt, weil das die Zeit ist, die gebraucht wird, bis die Daten dem Messwerkzeug zur Verfügung stehen.

Die Zeiten sind ebenfalls in der Exceltabelle als Spalte verfügbar. Wie bei der ersten Hypothese unterstützt die Excelfunktion ZÄHLENWENN auch hier die Auswertung der Daten. Das Suchkriterium der Funktion wird auf größer als zehn gesetzt. Das Ergebnis lautet zehn Datensätze, die eine längere Laufzeit als zehn Sekunden benötigten. Eine weitere Möglichkeit, um die betroffenen Datensätze zu visualisieren, ist die bedingte Formatierung in Excel. Damit werden Zellen, deren Zeiten die zehn Sekunden überschreiten, farblich markiert.

Um die Hypothese zu überprüfen, muss ein Hypothesentest durchgeführt werden. Dabei werden die Anzahl n , die Standardabweichung s , das Signifikanzniveau α und der Mittelwert \bar{x} benötigt. Nach der Ermittlung der Freiheitsgrade, indem die Anzahl mit eins subtrahiert wird, kann in der Tabelle der t-Werte¹¹⁴ der kritische t-Wert ermittelt werden. Bei einem Freiheitsgrad von 4129 und einem Signifikanzniveau von 5% ist der kritische t-Wert 1,645. Da es sich um einen linksseitigen t-Test handelt, muss der ermittelte Wert negativiert werden, also -1,645.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{0,1852 - 10}{\frac{0,547536}{\sqrt{4130}}} = -1151,976981 < t_{(0,05;4129)} = -1,645$$

Formel 1: Berechnung des t-Werts und Entscheidung auf Signifikanz¹¹⁵

Die Nullhypothese wird für Werte, die größer oder gleich zehn Sekunden sind, angenommen. In Formel 1 ist zu sehen, dass der ermittelte t-Wert kleiner ist als der kritische t-Wert. Aus diesem Grund wird die Nullhypothese abgelehnt. Das bedeutet, dass der überprüfte WFS in der Lage ist die benötigten Daten im Beispiel von Schutzgebietsdaten zu liefern.

Da diese Hypothese nur eine Variable überprüft, ist kein Scatterplot möglich. Für dieses Diagramm werden zwei verschiedene Datenreihen benötigt.

¹¹⁴ https://www.uibk.ac.at/econometrics/einf/tab_t_stat.pdf

¹¹⁵ Vgl. Wohlin et al. 2012, S. 138.

6.3.3 Dritte Hypothese

Zwischen der (Kern-)Arbeitszeit von 9-15 Uhr und außerhalb der Kernarbeitszeit besteht kein Zusammenhang zur Antwortzeit.

Um diese Hypothese zu testen, reicht eine einfache Auswertung über eine Funktion in Excel nicht aus. Aus diesem Grund wird für weitere Berechnungen Statistica 13 verwendet. Es ist zwar möglich Regressionsanalysen in Excel zu erstellen, aber dieser Prozess ist zeitintensiver als bei Statistica. Es folgt eine beispielhafte Berechnung über Excel, um die Werte aus Statistica aus der Tabelle 8 zu bestätigen. In Excel gibt es ein Add-In namens Analyse-Funktionen, welches zuerst aktiviert werden muss. Damit ist es möglich eine Regressionsanalyse für zwei Variablen durchzuführen.

Tabelle 9: Regressionsanalyse mit Excel

<i>Regressions-Statistik</i>	
Multipler	
Korrelationskoeffizient	0,00164023
Bestimmtheitsmaß	2,6903E-06
Adjustiertes	-
Bestimmtheitsmaß	0,00023956
Standardfehler	0,54760193
Beobachtungen	4130

Quelle: Eigene Darstellung

Bei Tabelle 9 ist ein Ausschnitt der berechneten Regressionsanalyse aus Excel mit den Variablen aus der dritten Hypothese abgebildet. Die zwei Variablen sind *measured time* und *working time*. Der ermittelte Wert aus Tabelle 8 wird somit bestätigt. Beide Berechnungen kommen auf einen gerundeten Korrelationskoeffizienten von 0,0016. Statistica ermittelt parallel den p-Wert und kann somit entscheiden, ob der Wert statistisch signifikant ist oder nicht. Da der Korrelationskoeffizient sehr nah an null liegt, besteht kein Zusammenhang zwischen der Arbeitszeit und der gemessenen Systemzeit. Die Nullhypothese wird also angenommen.

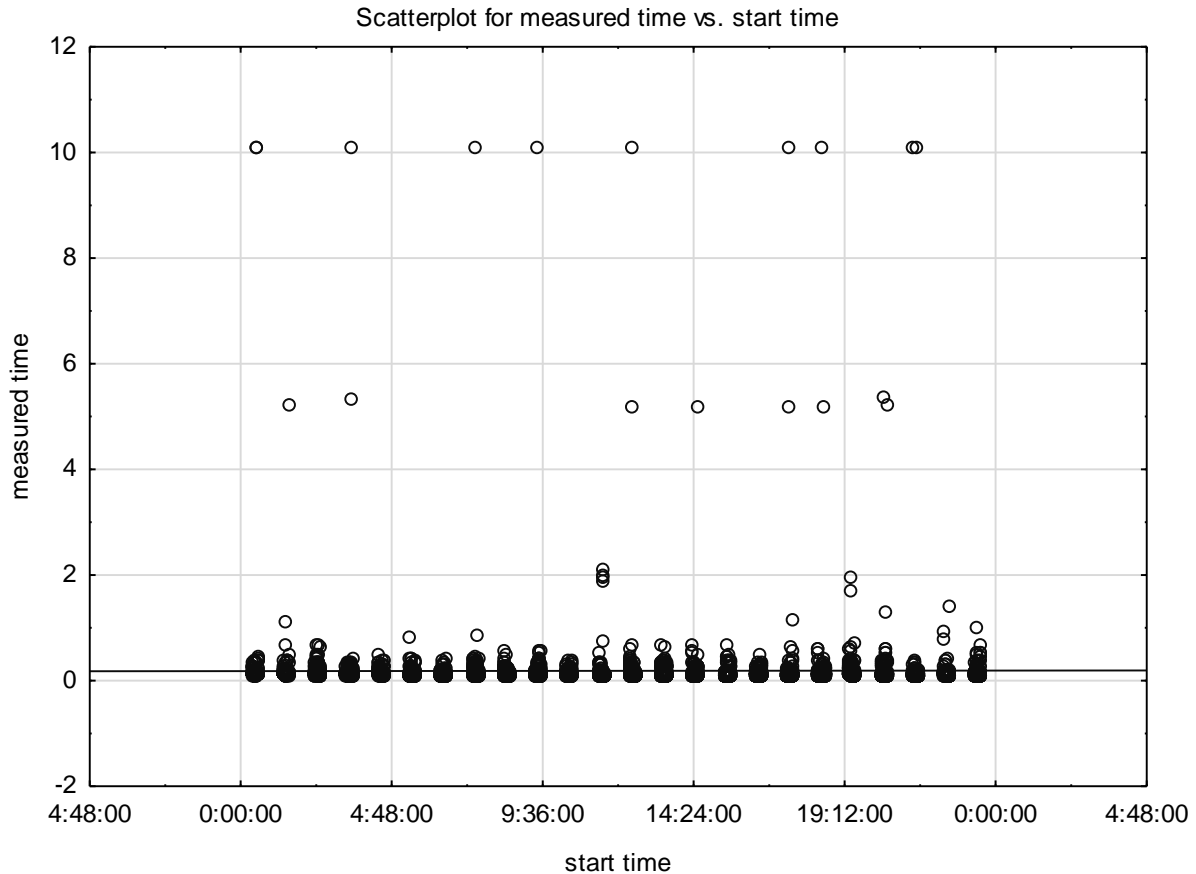


Abbildung 18: Streudiagramm mit der gemessenen Systemzeit und der Startzeit ¹¹⁶

Bei der Korrelation aus Tabelle 8 ist eine Korrelation nahe null für diese beiden Variablen abzulesen. Dies ist in Abbildung 18 ebenfalls durch die eingezeichnete waagerechte Trendlinie visualisiert. Dadurch lassen sich Vorhersagen für weitere Messungen bilden. Über die Steigung dieser Geraden wird der Zusammenhang der Variablen abgebildet. Da die Steigung dieser Linie fast null ist, bestätigt das Diagramm die Werte aus Tabelle 8. Es gibt keine Beziehung zwischen der gemessenen Zeit und der Startzeit der Anfrage. Die Datenpunkte sind sehr gleichmäßig verteilt. Auf der x-Achse finden sich 24 Säulen, weil die Messungen im Abstand von einer Stunde durchgeführt worden. Die Ausreißer sind jedoch auch gleichmäßig über die Startzeiten verteilt, es gibt also keinen Zusammenhang zwischen der Zeit wann die Abfrage getätigt wurde und der Dauer der Anfrage. Im Gegensatz zu diesem Diagramm findet im Folgenden eine Einteilung der Startzeiten in Arbeitszeit und keine Arbeitszeit statt. Dadurch kann untersucht werden, ob durch mehr Abfragen in der Arbeitszeit die Fehlerquote des WFS höher ist als außerhalb.

¹¹⁶ Eigene Darstellung

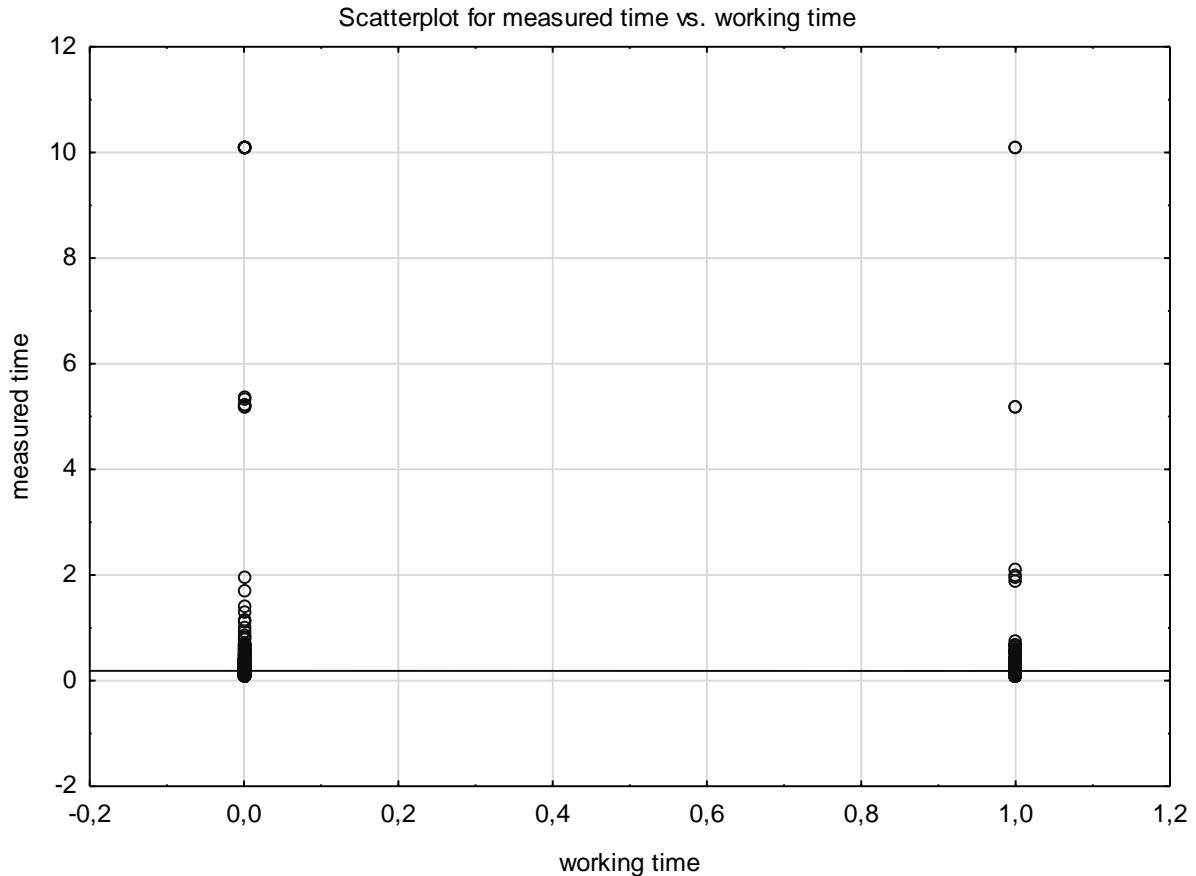


Abbildung 19: Streudiagramm mit der gemessenen Systemzeit und der Arbeitszeit ¹¹⁷

In Abbildung 19 ist auf der x-Achse die neue Variable working time abgebildet. Bei dieser Variable gibt es nur zwei Werte. Bei der null ist die Startzeit der Messung nicht in der definierten Arbeitszeit aus der dritten Hypothese. Aus diesem Grund befinden sich im linken Bereich des Diagramms mehr Werte. Für den Wert eins bedeutet das, dass die Startzeit in der Arbeitszeit liegt. Es ist zu sehen, dass es bei beiden Werten Ausreißer nach oben gibt. Bei beiden Säulen gibt es Abfragen, die gescheitert sind. Dies ist an den Punkten, die über zehn Sekunden liegen, zu erkennen. Nach dieser Zeit gilt die Anfrage als missglückt. Die Regressionslinie bestätigt den Korrelationskoeffizienten aus der Tabelle 8 mit dem Wert von $-0,0016$. Die Linie ist nahezu parallel zur x-Achse und weist somit keine Steigung auf. Das bedeutet, dass es keine Beziehung zwischen den beiden untersuchten Variablen gibt.

¹¹⁷ Eigene Darstellung

6.4 Diskussion der Ergebnisse

Nach der Analyse und Interpretation der Versuchsdaten müssen Schlussfolgerungen über das Versuchsergebnis gezogen werden. Wenn die Hypothesen abgelehnt werden, können Schlussfolgerungen über den Einfluss der unabhängigen Variablen auf die abhängigen Variablen gezogen werden, vorausgesetzt, das Experiment ist valide. Wenn das Experiment jedoch die Nullhypothese nicht ablehnen kann, können keine Schlussfolgerungen über den Einfluss der unabhängigen Variablen auf die abhängige Variable gezogen werden. In diesem Fall wurde lediglich gezeigt, dass es keinen statistisch signifikanten Unterschied gibt.¹¹⁸

Die dritte Hypothese wurde mit den Daten aus dem Experiment aus Kapitel 5.2.5 abgelehnt. Das bedeutet, dass keine Schlussfolgerungen gezogen werden können. Da die Arbeitszeit eine wichtige Variable der dritten Hypothese ist, können weitere Einschränkungen für die Messungen stattfinden. Im Experiment ist beispielsweise Sonntag enthalten, der für viele Berufe kein regulärer Arbeitstag ist. Aus diesem Grund sollte diese Variable weiter untersucht werden. Bei dieser Untersuchung wurde ebenfalls nicht überprüft, ob während der Arbeitszeit tatsächlich mehr Abfragen vom WFS bearbeitet wurden oder nicht. Diesen Wert kann das Messwerkzeug nicht feststellen. Dementsprechend können nur die Betreiber des WFS Auskunft über die Anzahl der Abfragen geben. Ohne diesen Wert ist das Ergebnis der dritten Hypothese nicht eindeutig interpretierbar, da diese Variable Auswirkungen auf die Fragestellung haben kann.

Wenn statistisch signifikante Unterschiede festgestellt wurden, besteht das Ziel darin, allgemeine Schlussfolgerungen über die Beziehung zwischen unabhängigen und abhängigen Variablen zu ziehen. Das Ergebnis kann jedoch nur auf Umgebungen verallgemeinert werden, die der Versuchsumgebung ähnlich sind. Selbst wenn das Versuchsergebnis statistisch signifikant ist, bedeutet dies nicht zwangsläufig, dass es von praktischer Bedeutung ist. Es kann auch vorkommen, dass Versuchsergebnisse statistisch nicht signifikant sind oder eine geringe statistische Signifikanz aufweisen, aber dennoch von praktischer Bedeutung sind. Die Tatsache, dass eine Nullhypothese auf einem bestimmten Signifikanzniveau nicht abgelehnt werden kann, bedeutet nicht, dass die Nullhypothese wahr ist. Es können Probleme bei der Versuchsplanung auftreten, wie beispielsweise zu wenige Datenstichproben. Wenn eine signifikante Korrelation zwischen Variable A und Variable B gefunden wird, kann im Allgemeinen nicht auf eine kausale Beziehung zwischen A und B geschlossen werden. Es

¹¹⁸ Vgl. Wohlin et al. 2012, S. 149.

könnte einen dritten Faktor C geben, der die messbaren Auswirkungen auf A und B verursacht.¹¹⁹

Insgesamt wurden 4130 Datensätze übermittelt. Laut der Berechnung der Anzahl der zu erwartenden Datensätze müssten es aber 4200 Datensätze sein. Das bedeutet 70 Abfragen konnten nicht durchgeführt werden. Bei einer gescheiterten Abfrage der GetCapabilities stehen keine Typenames im Skript zur Verfügung. Aus diesem Grund scheitern dann alle GetFeature-Abfragen für diese Messung. Mit abgespeicherten Typenames während der Laufzeit würde dieses Problem umgangen werden. Das führt jedoch zu einem höheren Aufwand bei der Entwicklung, da die Typenames vor dem Speichern mit den alten Zuständen verglichen werden müssen. Und andererseits ist bei den Anwendungsfällen, die keine langen Messungen darstellen, die Anzahl der Messungen eins. Das bedeutet, dass eine Speicherung nicht nötig ist. Für Messungen mit mehreren Durchläufen ist das Abspeichern der Typnamen auch nicht erwünscht, weil jede Abfrage als unabhängig zur vorherigen betrachtet wird. Das bedeutet, dass nach der Wartezeit eine neue, eigenständige Abfrage durchgeführt wird. Wenn die Typenames jedoch abgespeichert werden, gleicht das einer Manipulation des Experiments. Außerdem gab es im während der Implementierung und Testens des Messwerkzeugs Fehler. Beispielsweise gab es in Ausnahmen den Statuscode 302, was dazu führte, dass die Abfrage scheitert. Dieser Fehler wurde vor dem Start des Experiments behoben. Außerdem gab es Probleme, die zu einem Abbruch der Messung führten, wenn sich der Nutzer aus dem Betriebssystem ausloggt. Bei der Durchführung des Experiments traten keine Fehler mehr auf. Aus diesem Grund ist das vorherige Testen von Software wichtig.

Mithilfe der Version 3.5 von ChatGPT¹²⁰ wurde versucht valide Abfragen an einen WFS zu generieren. Um eine ähnliche Anfrage wie aus dem Experiment vom Modell zu erhalten, müssen viele Parameter wie bei der Eingabedatei an den Dienst gereicht werden. Jedoch kann ChatGPT noch nicht aktive Abfragen über das Internet durchführen, aus diesem Grund muss die Ausführung vom Nutzer oder einer Schnittstelle mit der ausgegebenen URL durchgeführt werden. Da keine Abfragen durchgeführt werden können, stehen die Typenames ChatGPT nicht zur Verfügung und müssen an das Tool gegeben werden. Das bedeutet, dass der Nutzer den WFS vorher selbst abfragen muss. Für viele Messungen hintereinander wie beim Experiment eignet sich ChatGPT auch nicht. In Zukunft soll der Dienst Zugriff auf das Internet haben und kann dann mit hoher Wahrscheinlichkeit Abfragen durchführen und das Ergebnis filtern.

¹¹⁹ Vgl. Wohlin et al. 2012, S. 149.

¹²⁰ <https://openai.com/chatgpt>

7 Zusammenfassung und Ausblick

In dieser Masterarbeit wurde das Verhalten von Geowebsservices mit einem dafür entwickelten Messwerkzeug analysiert. Für die Bewertung der ermittelten Ergebnisse wurden drei Forschungsfragen formuliert. Zur Beantwortung dieser Fragen wurde ein Experiment zur Datenerhebung mit einem Web Feature Service durchgeführt. Für jede Frage wurden Hypothesen aufgestellt, die statistisch überprüft wurden. Die erste Forschungsfrage ‚Sind die geprüften WFS grundsätzlich für das Verfahren geeignet?‘ konnte durch das Experiment nicht eindeutig geklärt werden. Dazu muss eine Grenze für die Verfügbarkeitsklassen gezogen werden. Im Experiment erreichte der getestete WFS die VK 1 und liegt damit bei über 99% Erreichbarkeit. Die zweite Forschungsfrage beschäftigt sich statt der allgemeinen Erreichbarkeit des Dienstes mit der Validität der gesendeten Daten. Sie lautet ‚Sind die derzeit verfügbaren WFS technisch in der Lage, bei synchronen Aufrufen die für die RPA benötigten Inputdaten am Beispiel von Schutzgebietsdaten zu liefern?‘. Hierbei wurden die Daten der gemessenen Zeit und die Antwortobjekte ausgewertet und durch einen Hypothesentest überprüft. Das Ergebnis ist, dass der getestete WFS in der Lage ist relevante Daten am Beispiel von Schutzgebieten zu liefern. Die letzte Forschungsfrage beschäftigt sich, ob die Uhrzeit einen signifikanten Einfluss auf die Antwortzeit des WFS hat. Aus diesem Grund wurde durch Korrelationskoeffizienten die Frage ‚Gibt es einen signifikanten Unterschied in der Antwortzeit des WFS während der Kernarbeitszeit von 9-15 Uhr im Vergleich zur Freizeit (außerhalb der Kernarbeitszeit)?‘ beantwortet. Das Ergebnis aus der Stichprobe des Experiments ist, dass es keinen signifikanten Unterschied gab. Es ist jedoch zu beachten, dass die Messung für eine komplette Woche, also auch Sonntag, durchgeführt wurde. Dies kann das Ergebnis beeinträchtigen. Aus diesen Ergebnissen lässt sich schließen, dass die WFS für eine Automatisierung geeignet sind. Das Messwerkzeug kann jeden WFS abfragen, auch außerhalb Europas wie in Kapitel 5.2.4 bewiesen wurde. Falls der Web Feature Service vom Standard des OGC abweicht, ist es nötig das Messwerkzeug anzupassen, um valide Abfragen zu erzeugen. Mit dem entwickelten Messwerkzeug lassen sich so die Dienste für einen langen Zeitraum abfragen, um Daten für Auswertungen zu sammeln. Das kann für weitere Forschungsarbeiten oder für die Betreiber dieser Dienste sinnvoll sein. Es ist jedoch ebenfalls möglich wie aus den Ergebnissen hervorgeht, mit dem Messwerkzeug automatisierte Abfragen für echte Bauanträge durchzuführen. Die Antwortzeiten des überprüften WFS lagen im Median bei ca. 0,12 Sekunden. Das bedeutet, dass bei der Automatisierung die Daten sehr schnell im System verfügbar sind und für weitere Verarbeitungen genutzt werden können.

Da bei dieser Arbeit und dem Experiment der zeitliche Rahmen begrenzt war, gibt es noch weiteren Forschungsbedarf und Handlungsempfehlungen für das Messwerkzeug. Eine Aufgabe für zukünftige Forschungen ist die Optimierung der Eingabe der Daten und die Verbesserung von internen Prozessen des Messwerkzeugs. Ein Beispiel ist das Schreiben in eine Datenbank, was sowohl Vorteile als auch Nachteile mit sich bringt. Dadurch müssen weitere Komponenten für den Start einer Messung auf dem Server bereitgestellt werden. Die Webseite aus Kapitel 4.7 bietet ebenfalls Optimierungspotential bezüglich der IT-Sicherheit und der Überprüfung der Eingaben. Die Webseite ist für unerfahrene Nutzer ein gutes Hilfsmittel und sorgt für technisch korrekte Eingabedaten. Aus diesem Grund kann eine Weiterentwicklung von Vorteil sein.

Weiterer Forschungsbedarf ergibt sich aus dem Vergleich von verschiedenen Web Feature Services. Für diese Arbeit wurde ausschließlich der WFS des BfN abgefragt und untersucht. Dieser WFS enthält Daten für ganz Deutschland. Es gibt jedoch auch andere WFS, die auf einzelne Bundesländer eingeschränkt sind. Diese könnten ebenfalls untersucht und untereinander oder mit dem deutschlandweiten WFS verglichen werden. Es ist auch möglich WFS aus verschiedenen Ländern zu vergleichen. Ein weiterer Punkt wäre die Umrechnung aus nicht metrischen Koordinatensystemen. Für das Experiment wurde ausschließlich EPSG 25832 gewählt, um eine Berechnung der Bounding Boxen zu ermöglichen. Das führt zu einer weiteren Fragestellung, die weiter untersucht werden kann. Das Koordinatensystem könnte Einfluss auf die Zeiten der Abfragen haben. Eine gute Funktion bezüglich des EPSGs wäre die Überprüfung, ob der zu untersuchende Dienst das gewünschte Koordinatensystem unterstützt. Dies ist in den GetCapabilities enthalten und kann so vor den konkreten Abfragen ausgeführt werden.

Weitere Untersuchungen mit dem Testlauf aus Kapitel 5.2.3 können relevant sein, weil durch passende Konfigurationen so das Fehlerhandling des WFS überprüft werden kann. Somit können bewusst fehlerhafte Abfragen gesendet und die Antworten gespeichert werden. Ein Beispiel ist, dass ein einzelner Punkt abgefragt werden soll, aber die Abfrage schickt drei Koordinaten. Daraus ergibt sich die Frage wie der WFS mit dieser Anfrage umgeht, ob beispielsweise die erste oder die letzte Koordinate verarbeitet wird oder ein Fehler entsteht.

Eine weitere Empfehlung ist die Anpassung der Dateien. Die CSV-Datei weist im aktuellen Stand nur Variablenwerte auf. Diese können dann mit Excel verarbeitet oder für weitere Berechnungen genutzt werden. Es ist jedoch auch möglich valide Formeln, die Excel berechnen kann, in die CSV per Messwerkzeug zu schreiben. Dadurch werden diese Formeln beim Öffnen mit Excel sofort berechnet und das Ergebnis dieser Berechnung erscheint dann in der jeweiligen Zelle. Das Messwerkzeug könnte das Ergebnis ebenfalls berechnen, jedoch birgt das den Nachteil, dass die Rechnung für Nutzer nicht nachvollzogen werden kann. Die

Eingabedatei ist eine Textdatei. Dies ermöglicht jedem Nutzer eine problemlose Anpassung des Inhalts an die jeweiligen Bedürfnisse. Für Textdateien sind keine weiteren Informatikkenntnisse notwendig. Außerdem kann jedes Betriebssystem Textdateien öffnen und bearbeiten. Es kann jedoch von Vorteil sein die Eingabedatei in maschinenlesbaren Formaten wie XML oder JSON bereitzustellen. Dadurch können Systeme die Eingabedatei automatisch erstellen und besser parsen. Der Nachteil ist, dass es für Menschen ohne Kenntnisse nicht mehr möglich ist diese Dateien zu editieren, das bedeutet es werden Grundkenntnisse vorausgesetzt.

Das Ergebnis, das vom WFS bereitgestellt wird, ist standardmäßig im Format gml/xml. Dies wird in einer XML-Datei gespeichert und steht dem Nutzer dann zur Verfügung. Es gibt jedoch auch Fälle in denen andere Datenformate wie beispielsweise JSON gewünscht sind. Dies kann ebenfalls konfigurierbar gemacht und mit in Eingabedatei aufgenommen werden. Die Auswertung der Ergebnisse ist für die Automatisierung wichtig. Für diese Arbeit waren nur Teile der Antwort des WFS von Bedeutung. Besonders die Zahlen wie Anzahlen oder Zeiten, die statistisch ausgewertet können, wurden verwendet. Bei den betroffenen Gebieten einer Abfrage werden jedoch auch die jeweiligen Namen gespeichert. Diese Namen geben mehr Auskunft und sollten daher auch gefiltert und in die CSV-Datei gespeichert werden.

Mit diesen Vorschlägen kann das Messwerkzeug und der Prozess der Messung weiter optimiert und weitere wissenschaftliche Untersuchungen durchgeführt werden. Die Ergebnisse zeigen, dass die Integration von Geowebsservices in Prozesse wie das digitale Baugenehmigungsverfahren erhebliche Vorteile bietet. Die gezeigten Vorteile sind erhebliche Zeitersparnisse bei Abfragen im Vergleich zu manuell getätigten Abfragen und dadurch eine verbesserte Effizienz und Genauigkeit bei der Standortanalyse. Es gibt jedoch auch Hindernisse wie Datenschutzbedenken, technische Kompatibilitätsprobleme und dem Verhalten, wenn der WFS keine Antwort liefert. Diese Punkte müssen vor einer erfolgreichen Automatisierung des Prozesses geklärt werden.

Literaturverzeichnis

- Bajwa, Imran Sarwar; Sibaliya, Tatjana; Jawawi, Dayang Norhayati Abang (Hg.) (2020): Intelligent Technologies and Applications. Second International Conference, INTAP 2019, Bahawalpur, Pakistan, November 6–8, 2019, Revised Selected Papers. 1st ed. 2020. Singapore: Springer Singapore; Imprint Springer (Springer eBook Collection, 1198).
- Becher, Margit (2021): XML. DTD, XML-Schema, XPath, XQuery, XSL-FO, SAX, DOM. 2. Auflage. Wiesbaden, Heidelberg: Springer Vieweg (Lehrbuch).
- Bressoud, Thomas; White, David (2020): Introduction to Data Systems. Building from Python. 1st ed. 2020. Cham: Springer International Publishing; Imprint Springer.
- Cardille, Jeffrey A.; Crowley, Morgan A.; Saah, David; Clinton, Nicholas E. (2023): CLOUD-BASED REMOTE SENSING WITH GOOGLE EARTH ENGINE. Fundamentals and. [S.l.]: SPRINGER INTERNATIONAL PU.
- Deinum, Marten; Cosmina, Iuliana (2021): Pro Spring MVC with WebFlux. Web Development in Spring Framework 5 and Spring Boot 2. 2nd edition. Erscheinungsort nicht ermittelbar, Boston, MA: Apress; Safari. Online verfügbar unter <https://learning.oreilly.com/library/view/-/9781484256664/?ar>.
- Fahrmeir, Ludwig; Heumann, Christian; Künstler, Rita; Pigeot, Iris; Tutz, Gerhard (2016): Statistik. Der Weg zur Datenanalyse. 8., überarbeitete und ergänzte Auflage. Berlin, Heidelberg: Springer Spektrum (Springer-Lehrbuch). Online verfügbar unter <http://www.springer.com/>.
- Fauzan, Reza; Siahaan, Daniel; Rochimah, Siti; Triandini, Evi (2019): Use Case Diagram Similarity Measurement: A New Approach. In: Proceedings of 2019 International Conference on Information & Communication Technology and System (ICTS). Surabaya, July 18th, 2019. 2019 12th International Conference on Information & Communication Technology and System (ICTS). Surabaya, Indonesia, 7/18/2019 - 7/18/2019. Institut Teknologi Sepuluh Nopember. Piscataway, NJ: IEEE, S. 3–7.
- Friesike, Sascha; Sprondel, Johanna (2021): Träge Transformation. Welche Denkfehler den digitalen Wandel blockieren. Was bedeutet das alles? 1. Originalausgabe. Ditzingen: Reclam Verlag (Reclams Universal-Bibliothek – /Was bedeutet das alles?]). Online verfügbar unter <http://nbn-resolving.org/urn:nbn:de:bsz:24-epflicht-1975940>.
- Huerta, Joaquín; Schade, Sven; Granell, Carlos (Hg.) (2014): Connecting a digital Europe through location and place. Association of Geographic Information Laboratories for Europe. Cham: Springer (Lecture notes in geoinformation and cartography).
- Jasperneite, Jürgen; Lohweg, Volker (Hg.) (2022): Kommunikation und Bildverarbeitung in der Automation. Ausgewählte Beiträge der Jahreskolloquien KommA und BVAu 2020. Springer-Verlag GmbH; Jahreskolloquium Bildverarbeitung in der Automation. Berlin, Heidelberg: Springer Vieweg (Technologien für die intelligente Automation, 14).
- Klenk, Tanja; Nullmeier, Frank; Wewer, Göttrik (Hg.) (2020): Handbuch Digitalisierung in Staat und Verwaltung. Wiesbaden, Heidelberg: Springer VS (Springer reference).
- Lange, Norbert de (2020): Geoinformatik in Theorie und Praxis. Grundlagen von Geoinformationssystemen, Fernerkundung und digitaler Bildverarbeitung. 4., grundlegend überarbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer Spektrum (Lehrbuch).
- Lü, Guonian; Batty, Michael; Strobl, Josef; Lin, Hui; Zhu, A-Xing; Chen, Min (2019): Reflections and speculations on the progress in Geographic Information Systems (GIS): a geographic perspective. In: *International Journal of Geographical Information Science* 33 (2), S. 346–367. DOI: 10.1080/13658816.2018.1533136.

- Müller, Heidrun; Scheffler, Arne (2017): Potenziale von IT-Standards in den Bereichen Baugenehmigung und Planung. In: *VM* 23 (2), S. 99–111. DOI: 10.5771/0947-9856-2017-2-99.
- Ogryzek, Marek; Tarantino, Eufemia; Rzaša, Krzysztof (2020): Infrastructure of the Spatial Information in the European Community (INSPIRE) Based on Examples of Italy and Poland. In: *IJGI* 9 (12), S. 755. DOI: 10.3390/ijgi9120755.
- Rumpe, Bernhard (2012): Agile Modellierung mit UML. Codegenerierung, Testfälle, Refactoring. 2. Aufl. Berlin, Heidelberg: Springer Vieweg (Xpert.press).
- Rumpe, Bernhard (2016): Modeling with UML. Language, Concepts, Methods. Cham: Springer (Springer eBook Collection Computer Science).
- Seidl, Martina; Scholz, Marion; Huemer, Christian; Kappel, Gerti (2015): UML classroom. An introduction to object-oriented modeling. Cham, Heidelberg, New York, Dordrecht, London: Springer (Undergraduate topics in computer science).
- Shekhar, Shashi (Hg.) (2008): Encyclopedia of GIS. With 90 tables. New York, NY: Springer (Springer reference). Online verfügbar unter <http://www.loc.gov/catdir/enhancements/fy0824/2007933825-d.html>.
- Shekhar, Shashi; Xiong, Hui; Zhou, Xun (Hg.) (2017): Encyclopedia of GIS. Second Edition. Cham: Springer (SpringerLink Bücher).
- Shklar, Leon; Rosen, Rich (2012): Web application architecture. Principles, protocols and practices. 2. ed., reprinted. Chichester: Wiley.
- Starbuck, Craig (2023): The Fundamentals of People Analytics. With Applications in R. 1st ed. 2023. Cham: Springer International Publishing; Imprint Springer.
- Stember, Jürgen; Eixelsberger, Wolfgang; Spichiger, Andreas; Neuroni, Alessia; Habel, Franz-Reinhard; Wundara, Manfred (Hg.) (2021): Aktuelle Entwicklungen zum E-Government. Neue Impulse und Orientierungen in der digitalen Transformation der öffentlichen Verwaltung. Springer Fachmedien Wiesbaden. Wiesbaden, Heidelberg: Springer Gabler (Edition Innovative Verwaltung).
- Stenberg, Daniel (o.D.): curl.1 the man page. Online verfügbar unter <https://curl.se/docs/manpage.html>, zuletzt geprüft am 25.10.2023.
- Streicher, Hans Werner (2020): Digitale Transformation in der öffentlichen Verwaltung. Praxishandbuch für Projektleiter und Führungskräfte. Berlin, Heidelberg: Springer Gabler.
- Vretanos, Panagiotis (2010): OpenGIS Web Feature Service 2.0 Interface Standard. Hg. v. Open Geospatial Consortium Inc. Online verfügbar unter <https://www.ogc.org/standard/wfs/>, zuletzt aktualisiert am 02.11.2010, zuletzt geprüft am 14.10.2023.
- Walmsley, Priscilla (2013): Definitive XML schema. Covers latest usage and new XML Schema 1.1. Second edition. Upper Saddle River, NJ: Prentice-Hall/Pearson (Always learning).
- Wohlin, Claes; Runeson, Per; Höst, Martin; Ohlsson, Magnus C.; Regnell, Björn; Wesslén, Anders (2012): Experimentation in Software Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit eigenständig und ausschließlich unter Verwendung der angegebenen Literatur und sonstigen Hilfsmittel verfasst habe. Textpassagen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Merseburg, 15.11.2023

Tim Bernutz