



universität
wien

MASTERARBEIT

Titel der Masterarbeit

Aufbereitung und kartographische Visualisierung
adressgenauer thematischer Daten in Österreich und
Europa auf Basis regionalstatistischer Rastereinheiten

Beispielhaft anhand des Entwurfs einer Online-Applikation

Verfasser

Manuel Illmeyer, Bakk. techn.

angestrebter akademischer Grad
Master of Science (MSc)

Wien, 2015

Studienkennzahl lt. Studienblatt: A 066 856

Studienrichtung lt. Studienblatt: Kartographie und Geoinformation

Betreuer: Ass.-Prof. Mag. Dr. Karel Kriz

Inhaltsverzeichnis

Inhaltsverzeichnis	i
Abbildungsverzeichnis	iv
Tabellenverzeichnis	vi
Kurzfassung	vii
Vorwort.....	viii
1 Einleitung.....	1
1.1 Fragestellung und Methoden	2
1.2 Aufbau und Ziele.....	3
1.3 Inhalte und Begriffserläuterungen.....	5
2 Kartographische Visualisierung	10
2.1 Grundlagen der Kartographie	10
2.1.1 Bestandteile einer Karte.....	10
2.1.2 Kartographische Gestaltung und Visualisierung	12
2.1.3 Bildschirmkartographie	14
2.2 Thematische Kartographie	15
2.2.1 Die Grundkarte	15
2.2.2 Entwurf und Gestaltung thematischer Karten	16
3 Aufbereitung adressgenauer Daten durch Geokodierung	21
3.1 Einführung in Geokodierung.....	21
3.2 Methoden der Geokodierung.....	22
3.3 Geokodierungsqualität.....	23
3.4 Online-Geokodierungsdienste	25
3.5 Evaluierung der Online-Geokodierungsdienste	28
3.5.1 Fehler in Adressdatensätzen	29
3.5.2 Analytischer Hierarchieprozess.....	30
3.5.3 Kriterienbewertung	32

3.5.4 Analyse und Ergebnisse.....	34
3.5.5 Vergleich mit anderen Studien	45
3.5.6 Diskussion der Ergebnisse und Auswahl eines Geokodierungsdienstes	48
4 Daten auf Basis regionalstatistischer Raster in Österreich und Europa	51
4.1 Projektionen für Daten auf Rasterbasis	51
4.2 Daten in Österreich.....	52
4.3 Daten in Europa.....	54
4.3.1 Das GEOSTAT 1A Projekt	54
4.3.2 Der GEOSTAT 1A Datensatz	56
4.4 Datenschutzbestimmungen und Vertraulichkeitsaspekte	58
4.5 Entwicklungen.....	60
5 Regionalstatistische Raster	62
5.1 Geschichte	63
5.2 Administrative vs. geometrische Gebietsgliederungen	64
5.2.1 Bezugflächen	66
5.2.2 Zeitliche Vergleichbarkeit	66
5.2.3 Darstellungen auf Länderebene	68
5.2.4 Länderübergreifende Darstellungen	69
5.3 Zuordnung statistischer Daten zu Rastern.....	71
5.3.1 Zuordnung von Punktdaten zu Rasterzellen	72
5.3.2 Zuordnung von Flächendaten zu Rasterzellen.....	72
5.3.3 Zuordnung von linienbezogenen Daten zu Rasterzellen	73
5.4 Darstellungsarten und kartographische Aspekte	74
5.5 Anwendungsbereiche rasterbasierter Karten.....	75
5.5.1 Infrastrukturplanung und Verkehrsplanung.....	76
5.5.2 Geomarketing	77
6 Erstellung eines Prototypen der Online-Applikation	78
6.1 Konzept und Idee.....	78

6.2	Technische Umsetzung.....	81
6.3	Funktionsweise.....	83
6.4	Weiterentwicklungen.....	90
6.5	Beispiele	90
7	Zusammenfassung.....	96
8	Literatur.....	102
	Abkürzungsverzeichnis	107
	Lebenslauf	109

Abbildungsverzeichnis

Abb. 1: Übersicht der Teilauspekte der Webapplikation	4
Abb. 2: Ablauf und Abhängigkeiten der Arbeit.....	4
Abb. 3: Bestandteile einer Karte	11
Abb. 4: Kartographische Variablen der Objektmerkmale.....	13
Abb. 5: Arten thematischer Karten	17
Abb. 6: Fehlinformation durch Darstellung von Absolutwerten in einer Choroplethenkarte.....	18
Abb. 7: Methoden der Klassenbildung anhand der Verteilung der Einwohnerzahlen von 30 Städten	19
Abb. 8: Prozess der Offline- (links) und Online-Geokodierung (rechts).....	22
Abb. 9: Entscheidungshierarchie AHP	31
Abb. 10: Übersicht der Adresspunkte im Testdatensatz	35
Abb. 11: Geokodierungsschnelligkeit in Sekunden pro Anfrage.....	37
Abb. 12: Boxplot-Diagramme für Distanzen aller Geokodierungsergebnisse.....	40
Abb. 13: Boxplot-Diagramme für Distanzen der Geokodierungsergebnisse von Bing und Google.....	41
Abb. 14: Histogramm der Distanzen für CloudMade, MapQuest und OpenStreetMap	42
Abb. 15: Histogramm der Distanzen für Bing und Google.....	42
Abb. 16: Trefferraten aller Geokodierungsdienste.....	43
Abb. 17: Durchschnittliche Geokodierungszeit pro Anfrage in Sekunden	47
Abb. 18: Geokodierungsqualität von Bing und Google nach Genauigkeitsstufe in Prozent	47
Abb. 19: Altersgliederung der unter 15-jährigen Bevölkerung 2012 auf Basis 2km- Rasterzellen.....	53
Abb. 20: Verfügbare Georeferenzierung der Zensus Daten 2011.....	55
Abb. 21: Europaweiter Bevölkerungsraaster 2011	57
Abb. 22: Bevölkerungsdichetekarte Südschwedens 1856	64
Abb. 23: Absolute Einwohnerzahlen für das Wiener Umland.....	66
Abb. 24: Stand der Bezirksgrenzen Steiermark vor und nach der Zusammenlegung 2013.....	68

Abb. 25: Bevölkerungsdichetekarte Österreichs mit den Bezugsebenen Gemeinden und 2 km-Raster.....	69
Abb. 26: Bevölkerungsdichte im Grenzgebiet Österreich/Slowenien.....	71
Abb. 27: Zuordnung von Polygonflächen zu Rastern über Rasterzellenmittelpunkt.....	73
Abb. 28: Zuordnung von Polygonflächen zu Rastern über Rasterzellenschwerpunkte.....	73
Abb. 29: Zuordnung von linienhaften Daten zu Rasterzellen	74
Abb. 30: Nebenwohnsitze 2001 in NÖ und Wien auf Basis 2,5km-Rasterzellen.....	75
Abb. 31: Fußläufige Distanz zu Supermärkten	76
Abb. 32: Überblick der Webapplikation	78
Abb. 33: Workflow der Webanwendung	81
Abb. 34: Einstiegsseite der Webapplikation	83
Abb. 35: Upload von Adressdaten	84
Abb. 36: Einstellungen zur hochgeladenen Adressdatei	85
Abb. 37: Ergebnis der Geokodierung.....	85
Abb. 38: Upload von Punktdaten	86
Abb. 39: Einstellungen zur hochgeladenen Koordinaten-Datei.....	86
Abb. 40: Konfigurationsmöglichkeiten für rasterbasierte Karten	89
Abb. 41: Bevölkerungsdichte Österreichs 2011 auf Basis 1km Rasterzellen	92
Abb. 42: Bevölkerungsdichte Europas 2011 auf Basis 10km Rasterzellen	93
Abb. 43: Schulen in Österreich auf Basis von 2km-Rasterzellen	94
Abb. 44: Grundwasser- und Fließgewässer-Messstellen in Österreich auf Basis von 2km-Rasterzellen.....	95

Tabellenverzeichnis

Tab. 1: Vorteile und Nachteile von Online-Geokodierungsdiensten	23
Tab. 2: Zusammenfassung ausgewählter Geokodierungsdienste.....	28
Tab. 3: Auszug aus möglichen Adressdatensatz	29
Tab. 4: AHP-Skala zur Bewertung qualitativer Merkmale	32
Tab. 5: Vergleichsmatrix der Kriterien Schnelligkeit, Genauigkeit und Zuverlässigkeit	32
Tab. 6: Vergleichsmatrix der Subkriterien für Genauigkeit.....	33
Tab. 7: Vergleichsmatrix der Subkriterien für Zuverlässigkeit.....	33
Tab. 8: Lokale Gewichtung der Hauptkriterien.....	33
Tab. 9: Globale Gewichtung der Kriterien	34
Tab. 10: Auszug aus Test-Adressdatensatz.....	36
Tab. 11: Adressdaten in Europa	36
Tab. 12: Statistische Distanz-Parameter der originalen Geokodierungsergebnisse	38
Tab. 13: Statistische Distanz-Parameter der getrimmten Geokodierungsergebnisse	39
Tab. 14: Umgang der Geokodierungsdienste mit Fehlern in Quelldaten.....	44
Tab. 15: Ergebnismatrix der Bewertung	44
Tab. 16: Normalisierte Ergebnisse der Alternativenbewertung	45
Tab. 17: Gewichtete Ergebnisse der Alternativenbewertung	45
Tab. 18: Trefferraten der Geokodierungsergebnisse.....	46
Tab. 19: Abweichungen der Geokodierungsdienste in Meter	46
Tab. 20: Ergebnisse gruppiert nach Genauigkeitsstufe	48
Tab. 21: Anteil verschiedener Bevölkerungsmerkmale die aus Datenschutzgründen in Österreich nicht weitergegeben werden.....	59

Kurzfassung

Die vorliegende Masterarbeit beschäftigt sich mit der Aufbereitung und Visualisierung adressgenauer Daten auf Basis regionalstatistischer Rastereinheiten. Nach einer Einführung in die (thematische) Kartographie wurde der Prozess der Geokodierung näher behandelt. Dafür wurden Adressdaten geographische Koordinatenpaare zugeordnet, um diese auf einer kartographischen Darstellung abbilden zu können. Neben der Beschreibung des Geokodierungsvorgang an sich wurden ebenfalls einige Geokodierungsdienste evaluiert. Es wurde ermittelt welche rasterbasierten Daten in Österreich bzw. Europa bereits existieren und in weiterer Folge die Spezifikationen und Charakteristiken von rasterbasierten kartographischen Darstellungen genauer erläutert. Dazu wurde ein Vergleich zu Abbildungen mit administrativen Bezugseinheiten durchgeführt und die Einsatzmöglichkeiten rasterbasierter Karten aufgezeigt. Die Erkenntnisse des theoretischen Teils und der durchgeführten Untersuchungen flossen in die Umsetzung einer Webapplikation ein, mit welcher Adressdaten geokodiert und auf einer Karte dargestellt werden können.

Abstract

The present master thesis deals with the preparation and visualization of address accurate data based on regional statistical grids. After an introduction to thematic mapping, the process of geocoding is treated in more detail. Therefore geographic coordinates have been assigned to address data, to be able to display it on a map. In addition to the description of the process of geocoding, some geocoding services have been evaluated. It was determined which grid-based data in Europe and Austria already existed and the specifications and characteristics of grid-based cartographic representations discussed. For this purpose, a comparison to maps with administrative units was accomplished and the capabilities of grid-based maps have been shown. The outcome of the theoretical part and investigations have been integrated in the development of a web application, with the functionalities to geocode address data and display them on a map.

Vorwort

Die Kartographie gilt in der Wissenschaft als eine Disziplin, mit der man einfach Sichtweisen der breiten Masse beeinflussen kann. Heutzutage gibt es bereits viele „Hobby-Kartographen“, z.B. in den Medien, die Karten als Mittel zur Veranschaulichung von geographischen Daten benutzen. Unglücklicherweise werden bei der Erstellung thematischer Karten bzw. Kartogrammen die Grundprinzipien der Kartographie teilweise missachtet (s. MONMONIER 1996). Ein Beispiel dafür ist die Verteilung von absoluten Werten als Flächenfärbung auf unterschiedlich große Bezugseinheiten wie z.B. Landesgrenzen, Bezirksgrenzen, etc. Dies geschieht jedoch wahrscheinlich oft ohne Vorsatz des Kartenproduzenten.

Die vorliegende Arbeit beschreibt eine in der Kartographie noch wenig verwendete Darstellungsweise von Verteilungen auf Karten und erläutert deren Anwendung. Diese Art der Darstellung existiert zwar bereits seit Mitte des 19. Jahrhunderts, allerdings hat sie noch keine große Anwendung in der klassischen Kartographie gefunden. Die Regionalstatistik bedient sich bereits seit längerem rasterbasierter Bezugseinheiten.

Ein weiterer Eckpunkt dieser Arbeit ist die Darstellung von adressgenauen Daten in thematischen Karten. Dies erweist sich oftmals als schwierig, da Adressen in einem Geoinformationssystem (GIS), einem System zur Erfassung, Bearbeitung, Organisation, Analyse und Präsentation räumlicher Daten, nicht ohne Umwandlung in eine exakte geographische Lage gebracht werden können. In diesem Kontext ist der Prozess der Geokodierung erforderlich. Es gibt bereits eine Vielzahl von Anbietern, die Geokodierungstools kostenlos oder gegen eine Gebühr anbieten. Diese Arbeit konzentriert sich dabei auf kostenlose Online-Dienste, die für Österreich bzw. Europa von Bedeutung sind, wobei eine Evaluierung ebenfalls erfolgt.

Da, soweit bekannt, noch keine Anwendung existiert, die diese zwei Aspekte kombiniert, wird auf Basis der gewonnenen Erkenntnisse eine Online-Applikation entwickelt. Diese soll es ermöglichen Adressen Online zu geokodieren und anschließend auf Basis von Rastereinheiten einfache und ansprechende thematische Karten zu erstellen.

1 Einleitung

Die Masterarbeit mit dem Titel „Aufbereitung und kartographische Visualisierung adressgenauer thematischer Daten in Österreich und Europa auf Basis regionalstatistischer Rastereinheiten“ befasst sich mit thematischen Karteninhalten auf Basis regionalstatistischer Rastereinheiten und der Erstellung dieser Karten mit Hilfe einer Online-Applikation. Diese Arbeit beschäftigt sich vor allem mit dem Prozess der Aufbereitung adressgenauer Daten, welcher durch den Vorgang der Geokodierung beschrieben werden kann und in Abschnitt 1.3 bzw. in Kapitel 3 genauer erläutert wird. Darüber hinaus werden die Bearbeitung der gesammelten Informationen sowie die Visualisierung in Form einer einfachen und kartographisch ansprechenden Karte mit Hilfe einer Online-Applikation thematisiert. Die kartographischen Grundlagen dazu werden in Kapitel 2 beschrieben. Der geographische Fokus der Arbeit liegt in Europa, wobei Österreich im Zentrum des Interesses steht. Die Erkenntnisse des theoretischen Teils fließen unmittelbar in den praktischen Teil ein, in dem die Programmierung einer Webplattform zur Geokodierung von Daten sowie deren Aufbereitung und kartographische Visualisierung als rasterbasierte Karte umgesetzt werden.

Regionalstatistische Rastereinheiten sind quadratische Rastermaschen, die flächendeckend über ein zuvor bestimmtes Gebiet gelegt werden und dadurch ein räumliches Bezugssystem für Geodaten aller Art bilden. Aufgrund der Unabhängigkeit von Verwaltungsgrenzen erlauben derartige räumliche Einheiten eine stärkere sachbezogene Gebietsabgrenzung (vgl. KAMINGER 2010a).

Kartographische Abbildungen, deren Thematiken an administrative Einheiten gebunden sind, sind aufgrund der vorgegebenen Grenzen bzw. der Größe der Bezugsflächen wenig anpassungsfähig. Bspw. können Darstellungen verschiedener Zeiträume nicht miteinander verglichen werden, für den Fall, dass sich die Grenzen der Bezugsebene geändert haben (vgl. u.a. WONKA 2010; WONKA et al. 2007). Wegen der unterschiedlichen Größe und Form verschiedener administrativer Grenzen, wie bspw. Gemeinde-, Bezirks-, Landes- oder Staatsgrenzen, können keine neutralen Raumgliederungen dargestellt werden, wodurch ein flächenmäßiger Vergleich nicht möglich ist.

Im Zuge dieser Arbeit soll aufgezeigt werden, wie mit Hilfe von kartographischen Darstellungen auf Basis regionalstatistischer Rastereinheiten die oben angesprochenen Sachverhalte besser bearbeitet und die Inhalte anhand einer Webapplikation auf einer Karte dargestellt werden können. Adressgenaue Geodaten können dadurch als gleich große Rasterflächen repräsentiert werden, wodurch Vergleiche vereinfacht werden können. Bei Verwendung eines europaweit gleichmäßigen Rasters kann dies auch länderübergreifend gewährleistet werden, sofern adressgenaue thematische Daten auf Basis von Rasterzellen zur Verfügung stehen. Des Weiteren ist es bspw. möglich beliebige Daten von Unternehmen, Vereinen oder NGOs auf örtlichen, regionalen oder überregionalen Rasterkarten adäquat darzustellen.

1.1 Fragestellung und Methoden

Der vorliegenden Forschungsarbeit liegt folgende zentrale Forschungsfrage zugrunde:

„Wie können adressgenaue Daten in einer Karte, basierend auf regionalstatistischen Rastern, dargestellt werden und wie können diese Anforderungen mit Hilfe einer Webapplikation realisiert werden?“

Zur adäquaten Beantwortung dieser zentralen Forschungsfrage ist es notwendig das Forschungsvorhaben in bestimmten Themenbereich zu gliedern, wobei jeweils eine Teilfrage beantwortet werden muss.

- „Welche Anbieter im Bereich der Geokodierung von Adressdaten gibt es und wie genau und zuverlässig arbeiten diese?“
- „Wo sind adressgenaue bzw. thematische Daten auf Rasterbasis verfügbar und unter welchen Konditionen werden diese angeboten?“
- „Welche Unterschiede gibt es zwischen kartographischen Darstellungen auf Rasterbasis und Karten basierend auf administrativen Einheiten?“
- „Welche Arten von kartographischen Darstellungen auf Basis von Raster gibt es?“
- „Welche typischen Anwendungsbeispiele gibt es für Abbildungen auf Basis regionalstatistischer Rastereinheiten?“
- „Was sind die Anforderungen an eine Webapplikation, zur Darstellung regionalstatistischer Darstellungen und wie können diese umgesetzt werden?“

Ziel der Fragestellung ist es zu beweisen, dass die in der Kartographie bislang noch selten verwendeten kartographischen Darstellungen auf Basis regionalstatistischer Raster dazu geeignet sind, vor allem länderübergreifende thematische Inhalte besser darstellen zu können als Karten basierend auf administrativen Bezugsflächen. Um dies zu verdeutlichen wird eine Online-Applikation entwickelt, mit der rasterbasierte Karten einfach erstellt werden können.

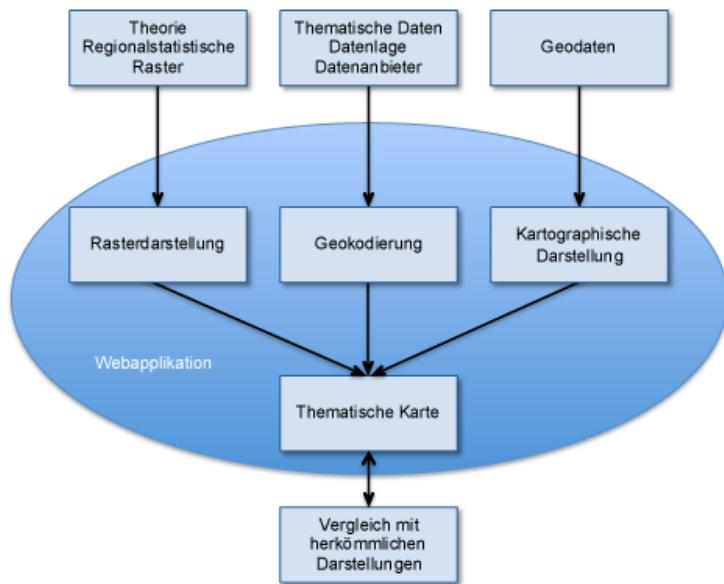
Der Großteil der Fragestellungen wird mit Hilfe einer theoretischen Abhandlung des Themas bearbeitet, wobei zur Veranschaulichung des Theorieteils kartographische Abbildungen den geschriebenen Text unterstützen. Die Evaluierung von Geokodierungsdiensten erfolgt mittels einer multikriteriellen Entscheidungsanalyse. Die Umsetzung der Webapplikation wird mit aktuellen Techniken der Webentwicklung durchgeführt. Als Quellen für den theoretischen Teil werden wissenschaftliche Arbeiten nationaler und internationaler Experten im Bereich der thematischen Kartographie und der Regionalstatistik herangezogen. Die theoretische Abhandlung über die Verwendung regionalstatistischer geographischer Raster basiert auf österreichischer Ebene unter anderem auf den Ausführungen von WONKA und KAMINGER, die dieses Thema in verschiedenen Publikationen behandelt haben. Auf europäischer Ebene gibt es, neben den Forschungsergebnissen der Experten in den einzelnen Ländern, Bestrebungen seitens der INSPIRE-Initiative, EUROSTAT und des EFGS Raster als Darstellungsform thematischer Daten einzuführen. Dafür wird die Verwendung eines einheitlichen Projektionssystems (flächentreue Lambert Azimutal-Projektion „ETRS-LAEA“) für Europa propagiert (vgl. STEINNOCHER et al. 2005).

1.2 Aufbau und Ziele

Der theoretische Teil der Arbeit gliedert sich in vier große Themenschwerpunkte (vgl. Abb. 1 und Abb. 2). Der erste Teil behandelt sämtliche Aspekte rund um die Einflüsse der Kartographie auf die Darstellung rasterbasierter Abbildungen. Der zweite Schwerpunkt beschäftigt sich mit der Geokodierung von Daten, wobei hauptsächlich auf Online-Geokodierungsdienste zurückgegriffen wird. In diesem Zusammenhang wird einerseits eruiert, welche Anbieter zur Verfügung stehen. Andererseits wird eine Evaluierung hinsichtlich Genauigkeit, Geschwindigkeit und Limitierungen durchgeführt. Die Ergebnisse dieser Bewertung werden mit bestehenden Ergebnissen verglichen. Das dritte Hauptaugenmerk der Arbeit liegt auf der Erfassung der Verfügbarkeit koordinatengebundener Daten und deren Nutzungskonditionen. Außerdem wird die Genauigkeit und Verwertbarkeit dieser Daten ermittelt. Des Weiteren wird der aktuelle Stand der Entwicklungen in Europa untersucht. Der letzte Teil der theoretischen Abhandlung befasst sich mit dem Thema „regionalstatistischer Rasterabbildungen“. Dabei wird untersucht wie die Größe der Rasterzellen in Abhängigkeit des Kartenmaßstabes gewählt werden soll, sowie welche Arten von Rasterdarstellungen es gibt und in welchen Projektionen diese Abbildungen dargestellt werden sollten. Darüber hinaus werden allgemein anerkannte kartographische Grundprinzipien auf Rasterkarten angewendet. Die graphischen Ergebnisse der regionalstatistischen Rasterdarstellungen werden im Anschluss herkömmlichen Abbildungen auf Basis administrativer Einheiten gegenübergestellt.

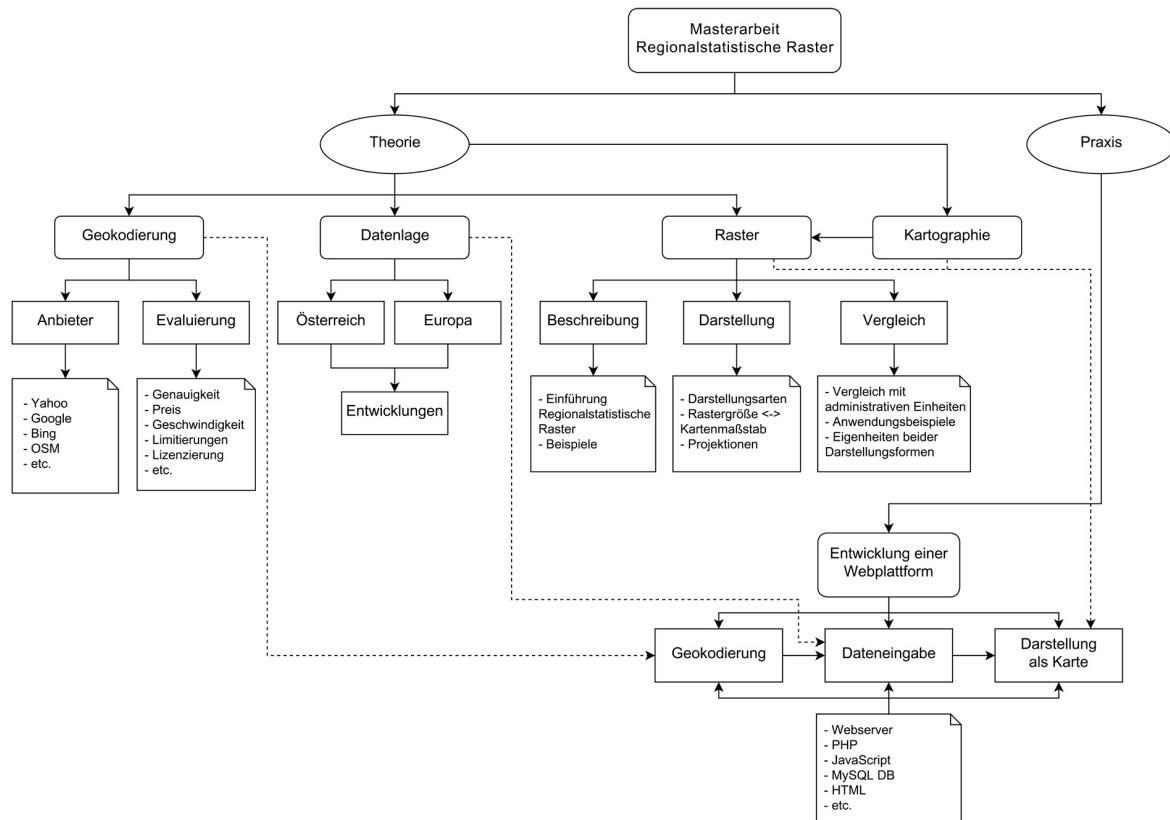
Im praktischen Teil der Arbeit wird die Umsetzung des Entwurfs einer Webapplikation beschrieben. In diesem Kontext werden die technischen Aspekte der Plattform erläutert sowie die verwendeten vorgefertigten Pakete beschrieben. Darüber hinaus wird die Anwendung des Programms auf Basis einiger Beispiele demonstriert. Des Weiteren wird erläutert, welche technischen Aspekte die Plattform beinhaltet, welche vorgefertigten Pakete verwendet wurden und wie das Programm verwendet werden kann.

Das Ziel der Arbeit ist es aufzuzeigen, dass die Anforderungen zur Darstellung von Daten auf Basis regionalstatistischer Raster und zur Aufbereitung adressgenauer Daten mit Hilfe einer Online-Anwendung umgesetzt werden können. Ferner soll gezeigt werden wie die Abbildung freier adressgenauer Daten durch die Verwendung von Rastereinheiten vereinfacht bzw. deren räumliche Vergleichbarkeit und Variabilität erhöht wird. Dies soll mit Hilfe der im Zuge der Arbeit entworfenen Internetapplikation veranschaulicht werden.



Quelle: ILLMEYER, 2012

Abb. 1: Übersicht der Teilespekte der Webapplikation



Quelle: ILLMEYER, 2012

Abb. 2: Ablauf und Abhängigkeiten der Arbeit

1.3 Inhalte und Begriffserläuterungen

Wie bereits erwähnt, gliedert sich diese Arbeit in die vier Themenbereiche (thematische) Kartographie, Geokodierung, Datenlage (in Österreich und Europa) und regionalstatistische Raster. Dabei liegt der Fokus darauf, die Ergebnisse der vier Abschnitte in den Entwurf einer Webanwendung einfließen zu lassen, welche die Erstellung von rasterbasierten Karten ermöglichen soll.

Die Kartographie, und in diesem Zusammenhang vor allem die thematische Kartographie, bilden die Basis um rasterbasierte Karten zu erstellen. In Kapitel 2 werden die Grundlagen zur Erstellung thematischer Karte näher gebracht und die notwendigen Bestandteile einer Karte beschrieben. Die wichtigsten Elemente finden sich ebenfalls in den Karten wieder, welche mit Hilfe der Webapplikation erstellt werden. So gibt es neben den thematischen Inhalten auch Elemente, welche im Kartenrand angezeigt werden. Dies sind im konkreten Fall der Titel der Karte, die Legende, ein Maßstabsbalken, Impressums- bzw. Quellenangaben und ein Nordpfeil, wobei die Verwendung des Nordpfeils auf Grund des kleinen Maßstabes umstritten ist.

Um thematische Daten laut geltenden kartographischen Prinzipien korrekt auf einer Karte darstellen zu können, muss eruiert werden, welche Methoden zur Erstellung von Karten es gibt. Dafür gibt es die sechs kartographischen Variablen Größe, Form, Füllung, Tonwert, Richtung und Farbe, welche auch im Rahmen der Erstellung von Karten mit Hilfe einer Webapplikation diskutiert und bearbeitet werden. Dafür wird ermittelt, welche Variablen überhaupt auf rasterbasierte Karten angewendet werden können. Um Karten auf einer ebenen Fläche abilden zu können, müssen mathematische Vorschriften zur Projektion der wahren Erdgestalt definiert werden. Die Wahl des Projektionssystems bzw. des Kartennetzentwurfs hängt in diesem Zusammenhang jedoch nicht nur von den kartographischen und geodätischen Eigenschaften unterschiedlicher Systeme ab, sondern auch von den zur Verfügung stehenden Daten und dem darzustellenden Gebiet. Des Weiteren müssen auch österreichische und europäische Empfehlungen hinsichtlich der Wahl und der Verwendung von Kartennetzentwürfen beachtet werden. Ein weiterer wichtiger Aspekt bei der Erstellung von rasterbasierten Karten im Kontext dieser Arbeit stellt die Optimierung des Kartenbildes für den Bildschirm dar. Da die Karten mit Hilfe einer Webapplikation erstellt werden, werden diese auch in einem ersten Schritt auf dem Bildschirm angezeigt. Aus diesem Grund müssen bekannte Limitierungen für die Anzeige von digitalen Karten auf Bildschirmen beachtet werden. Da es sich jedoch um keine dynamischen Webkarten handelt, sollen diese auch für den Druck optimiert werden.

Da rasterbasierte Karten der Gruppe an thematischen Karten zuzuordnen sind, werden diese im Rahmen der vorliegenden Arbeit auch näher behandelt. Dafür werden die unterschiedlichen Arten thematischer Karten aufgezeigt und kurz erläutert. Die Webapplikation wird in einer ersten Umsetzung jedoch lediglich die Erstellung von einschichtigen elementaranalytischen Choroplethenkarten ermöglichen. Choroplethenkarte, oder auch Flächenkartogramme, sind dabei thematische Karten, bei denen flächenhaft auftretende Sachverhalte im Verhältnis zur Bezugsebene dargestellt werden. Um die Daten

adäquat abbilden zu können, müssen diese jedoch in einem vorgelagerten Schritt klassifiziert werden. Dafür werden einzelne Werte so zusammengefasst, dass eine vorgegebene Anzahl an Klassen entsteht. Die Klassengrenzen können dabei mit Hilfe von statistischen Methoden ermittelt, oder, wenn gewünscht, auch manuell definiert werden. Für die Klassifizierung müssen bestimmte Grundsätze beachtet werden, da davon die Aussagekraft der Karte entscheidend abhängt. So müssen beispielsweise eindeutige Klassengrenzen gebildet werden, damit kein Wert in zwei unterschiedlichen Klassen abgebildet werden kann, und die Klassen müssen alle Werte der Datenreihe enthalten. Die Grundsätze werden in Kapitel 2.2.2 noch näher erläutert. Im Anschluss an die Klassenbildung werden diese oft mit einer bestimmten Färbung flächig ausgefüllt. Diese sollte, wenn möglich, so gewählt werden, dass direkte Assoziationen zu den Daten ermöglicht werden. Für die im Rahmen der Arbeit erstellte Webapplikation bedeutet dies, dass dem Benutzer eine Menge an Farbschemen zur Verfügung gestellt wird.

Als letzter wichtiger Bestandteil von thematischen Karten wird in Kapitel 2.2.1 die Grundkarte und deren Elemente diskutiert, welche im Zusammenhang mit der Erstellung von rasterbasierten Karten mit Hilfe einer Webapplikation ebenfalls von Bedeutung sind. Es wird beschrieben welche topographischen Elemente als Orientierungsgrundlage verwendet werden können. Die Aufbereitung der Daten für die Grundkarte der Webapplikation erfolgt bereits im Vorfeld der Entwicklung und kann in der Anwendung selbst nicht mehr verändert werden.

Um Adressdaten auf einer rasterbasierten Karte anzeigen zu können, müssen diese entsprechend aufbereitet und in geographische Koordinaten umgewandelt werden. Diesen Vorgang der Umwandlung nennt man Geokodierung. Eine genaue Erläuterung des Vorgangs bzw. Ablaufs einer Geokodierung von Adressdaten, sowie die verfügbaren Methoden und eine Evaluierung von Online-Geokodierungsdiensten erfolgt in Kapitel 3.

Der Geokodierungsvorgang kann generell in die Methoden der Offline- und Online-Geokodierung unterteilt werden. Im Zuge der Offline-Geokodierung sind spezielle Programme und Grundlagendaten nötig, um eine Adresse in Koordinaten umzuwandeln. Da bei der Erstellung der Webapplikation jedoch Online-Geokodierungsdienste verwendet werden, wird dieser Methode in der Arbeit erhöhte Aufmerksamkeit geschenkt. In Kapitel 3.2 werden außerdem die Vor- und Nachteile von Online-Geokodierungsdiensten zusammengefasst.

Ein weiterer wichtiger Bestandteil, um Adressen korrekt auf Basis regionalstatistischer Raster darstellen zu können, ist die Qualität des Geokodierungsvorgangs und die Qualität der Eingabedaten. Die Qualität der Eingabedaten hängt vom Benutzer des Geokodierungsdienstes bzw. der Webapplikation ab und liegt somit in dessen Einflussbereich. Auf Grund dessen werden, im Rahmen dieser Arbeit, die Online-Geokodierungsdienste hinsichtlich der Untersuchungskriterien „Genauigkeit“, „Schnelligkeit“ und „Zuverlässigkeit“ evaluiert. Um dies zu erreichen, müssen zuerst geeignete Dienste gefunden werden. Geeignet bedeutet in diesem Zusammenhang, dass diese Dienste für das Gebiet von Österreich bzw. Europa zur Verfügung stehen und zumindest teilweise kostenlos zur

Verfügung stehen. Die aufgefundenen Dienste werden in weiterer Folge hinsichtlich technischer Kriterien beschrieben und deren Limitierungen dargelegt.

Auf Basis der Ausführungen der jeweiligen Anbieter und anhand eigener Untersuchungen bzw. Bewertungen werden die Online-Geokodierungsdienste mit Hilfe der bereits erwähnten Kriterien evaluiert, wobei sich die Kriterien teilweise noch in Unterkriterien aufteilen lassen. Die Genauigkeit wird daher noch in die Subkriterien „Lagefehler“ und „Trefferraten“ unterteilt, die Zuverlässigkeit in die Unterkriterien „Fehlerhafte Datensätze“ und „Limitierungen“. Weiters werden typische Fehler in Adressdatensätzen aufgezeigt und wie diese von den Online-Diensten interpretiert werden sollten.

Die Evaluierung erfolgt anhand einer multikriteriellen Analyse, dem analytischen Hierarchieprozesses (AHP), welche ein sogenanntes Decision Support System (DSS) darstellt. Die Arbeitsschritte, welche dafür benötigt werden, sind in Kapitel 3.5.2 dargelegt und erläutert. Ziel des Evaluierungsprozesses ist es, einen für die entworfene Webapplikation geeigneten Dienst auszuwählen. Nach der Aufstellung der Kriterien müssen diese mit Hilfe eines paarweisen Vergleichs hinsichtlich deren Wichtigkeit bewertet werden um eine globale Gewichtung der Kriterien zu erhalten. Nach der Prioritätenberechnung der einzelnen Kriterien werden die Online-Geokodierungsdienste auf Basis dieser untersucht bzw. bewertet. Die Kriterien „Schnelligkeit“, „Lagefehler“ und „Trefferraten“ können dabei anhand von quantitativen Untersuchungsergebnissen bewertet werden. Die Kriterien „Limitierungen“ und „Datenfehler“ hingegen erfordern eine qualitative Analyse. Zur Evaluierung der Online-Geokodierungsdienste wird außerdem ein Adressdatensatz benötigt, welcher von den nationalen Statistik- und Kartographie-Instituten Europas zur Verfügung gestellt werden muss. Dabei ist wichtig, dass neben den Adressen auch die zugehörigen Koordinaten mitgeliefert werden, um etwaige Lagefehler nach der Geokodierung erkennen zu können.

Die Schnelligkeit von Online-Geokodierungsdiensten wird dadurch ermittelt, in dem die Adress-Testdaten mit dem in der Webapplikation integrierten Geokodierungs-Tool abgearbeitet werden. Dafür werden über einen bestimmten Zeitraum, die Adressen jeweils an alle verfügbaren Geokodierungsdienste übermittelt und auf deren Antworten gewartet. Dadurch erhält man relativ einfach einen aussagekräftigen Wert zur Bestimmung der Antwortzeiten und somit zur Ermittlung der Schnelligkeit. Der Lagefehler wird ermittelt, in dem die Adressdaten an den Geokodierungsdienst übermittelt werden und die zurückgegebenen Koordinaten mit den Koordinaten aus den Ursprungsdaten verglichen werden. Dafür muss die kürzeste Entfernung zwischen zwei Koordinatenpaaren ermittelt werden, welche entlang einer sogenannten „Orthodrome“ gemessen werden kann. Für das Kriterium der Trefferraten werden ebenfalls Adressdaten an die Online-Geokodierungsdienste gesendet und untersucht, wie viele der Anfragen ein gültiges Ergebnis zurück liefern. Im Rahmen der Evaluierung dieses Parameters ist es jedoch nicht von Bedeutung, ob das Ergebnis der Geokodierung lagerichtig ist.

Wie bereits erwähnt, werden die Kriterien „Fehlerhafte Datensätze“ und „Limitierungen“ anhand einer qualitativen Untersuchung beurteilt. Bei der Bewertung der Limitierungen werden die Angaben der Anbieter von Online-Geokodierungsdiensten gesammelt und

anhand einer Skala bewertet. Um zu ermitteln, wie die Dienste mit Fehlern in den Datensätzen umgehen, wird das zurückgelieferte Ergebnis mit den zu erwartenden Resultaten verglichen. Dafür wird, wie bereits erwähnt, in Kapitel 3.5.1 festgehalten, welche Resultate bei Fehlern zu erwarten sind. Um die erhaltenen Ergebnisse zu verifizieren, werden diese anschließend mit anderen, bereits durchgeführten, Studien verglichen, welche die gleichen Online-Geokodierungsdienste verwendet haben. Als Abschluss werden die Resultate nochmals diskutiert und zusammengefasst und ein, für die entworfene Webapplikation, geeigneter Dienst ausgewählt.

In Kapitel 4 der Arbeit wird aufgezeigt, welche adressgenauen bzw. rasterbasierten Daten in Österreich und Europa bereits existieren und welche Entwicklungen und Forschungen es für diese Art der Darstellung von Daten gibt. Neben den Daten der öffentlichen Institute sind beispielweise Kundendaten von Unternehmen zu nennen, insofern diese ihre Daten in adressgenauer Form vorhalten. Des Weiteren werden noch einmal Projektionen bzw. Netzentwürfe diskutiert, die für die Darstellung von Daten auf Basis von regionalstatistischen Rastern in Österreich und Europa in Frage kommen.

Hinsichtlich der Projektion von rasterbasierten Darstellungen und der zugrundeliegenden Rasternetze müssen, wie bereits erwähnt, ebenso Vorgaben und Bestimmungen der europäischen Länder bzw. der europäischen Gemeinschaft beachtet werden. So wurden beispielsweise österreichische Daten bisher in einem anderen Projektionssystem vorgehalten, als jenem, welches vom *European Forum for Geostatistics* (EFGS) für die rasterbasierte Darstellung vorgeschlagen wird. Auch seitens der *INSPIRE*-Initiative (Infrastructure for Spatial Information in the European Community) gibt es Vorgaben, wie Geodaten gespeichert und Bereitgestellt werden müssen. Um diesen Anforderungen gerecht zu werden, versuchen die nationalen Statistikinstitute die Geodatenhaltung in den jeweiligen Ländern zu normieren.

Auf österreichischer Ebene sind Adressdaten und rasterbasierte Daten vorwiegend bei der Bundesanstalt *Statistik Austria* erhältlich (siehe Kapitel 4.2). Auf europäischer Ebene werden die Daten, neben den nationalen Statistikinstituten, im Rahmen des *GEOSTAT*-Projektes gesammelt, welches vom EFGS geleitet wird (siehe 4.3). Ziel des Projektes ist es, Bevölkerungsdaten in allen teilnehmenden Ländern Europas auf Basis geokodierter Adressen zu erstellen. Dadurch kann ein einheitlicher europaweiter Bevölkerungsdichte-Datensatz auf Basis regionalstatistischer Raster erstellt werden.

Einen wichtigen Faktor bei der Weitergabe bzw. Darstellung von Adressdaten spielen Datenschutzbestimmungen. Es ist etwa nicht erlaubt, von den Merkmalsausprägungen auf einzelne Individuen schließen zu können. Dies muss ebenfalls bei der Entwicklung der Webapplikation beachtet werden, was sich zum Beispiel auf die Wahl der Rasterzellengröße auswirken kann.

Der letzte wichtige Faktor für die Darstellung von adressgenauen Daten auf rasterbasierten Karten ist die Definition des Begriffs der regionalstatistischen Raster und deren unterschiedlichen Darstellungsarten. Um begründen zu können, warum diese Art der Darstellung gewählt wurde, wird außerdem ein Vergleich zu Karten basierend auf

administrativen Gebietsgliederungen durchgeführt. Regionalstatistischen Raster sind demnach quadratische Rastermaschen, die flächendeckend über ein zuvor bestimmtes Gebiet gelegt werden und dadurch ein räumliches Bezugssystem für Geodaten aller Art bilden (vgl. KAMINGER 2010a). Der Vergleich der administrativen Einheiten gegenüber geometrischen Gebietsgliederungen erfolgt mit Hilfe bereits vorhandener Literatur im österreichischen und europäischen Raum (siehe Kapitel 5.2). Dadurch soll bewiesen werden, dass geometrische Gebietsgliederungen besser für den Vergleich von länderübergreifenden Darstellungen geeignet sind als administrative Bezugsflächen. Dafür werden in einem ersten Schritt Karten gegenübergestellt, welche auf vergleichbare Verwaltungsgrenzen bzw. Rasterzellen basieren. Weiters wird die zeitliche Vergleichbarkeit thematisiert, da sich koordinatengebundene Rasternetze, im Gegensatz zu administrativen Grenzen, im Laufe der Zeit nicht verändern. Ferner werden die Unterschiede bei Darstellungen auf Länderebene erläutert und die Vorteile bzw. Nachteile bei länderübergreifenden Karten beschrieben.

Um Geodaten auf einer rasterbasierten Karte darstellen zu können, müssen diese den jeweiligen Rasterzellen zugeordnet werden können. Während dies bei Punktdaten relativ einfach ermöglicht wird, gibt es bei der Zuordnung von Linien oder flächenhaften Elementen bereits mehrere Möglichkeiten, dies zu tun. Im Zuge der Erstellung von rasterbasierten Karten mit Hilfe der entwickelten Webapplikation ist es jedoch lediglich möglich, Punktdaten einzelnen Rasterzellen zuzuordnen. Um die die thematischen Daten in einer möglichst übersichtlichen Form darstellen zu können, werden weiters verschiedene Darstellungsarten rasterbasierter Karten aufgezeigt und deren jeweiligen Einsatzbereiche ausgewiesen. Außerdem werden noch wichtige Anwendungsbereiche aufgezeigt, in denen die erstellte Webapplikation Anwendung finden könnte.

Alle diese Erkenntnisse und Ergebnisse fließen zum Abschluss in den Entwurf einer Webapplikation ein. Diese Anwendung ermöglicht es, Adressdaten mit Hilfe eines Online-Geokodierungsdienstes in geographische Koordinaten umzuwandeln und diese in aggregierter Form auf einer Karte darzustellen. Als Ergebnis wird einer Karte, basierend auf regionalstatistischen Rastern, ausgegeben, welche durch den Benutzer auf Basis verschiedener Konfigurationseingaben erstellt wird.

Die Idee und das Konzept für die Online-Anwendung werden in Kapitel 6.1 näher erläutert. Des Weiteren werden die technische Umsetzung und die dafür verwendeten Pakete genauer beschrieben. In Kapitel 6.3 wird die Funktionsweise bzw. der Ablauf eines Durchgangs zur Erstellung einer Karte mit Hilfe von Screenshots der Applikation genau dokumentiert. Es werden ebenfalls die Konfigurationsparameter, z.B. für Klassifizierungseinstellungen, in diesem Abschnitt genau beschrieben werden. Da die Applikation einen ersten Entwurf darstellt, werden außerdem bereits Potenziale zur Weiterentwicklung identifiziert. Um die theoretischen Inputs in einen Kontext zu bringen, werden abschließend einige Beispiele, welche mit Hilfe der entwickelten Webanwendung erstellt wurden, präsentiert und die zugrundeliegenden Daten und Konfigurationsparameter dokumentiert.

2 Kartographische Visualisierung

Dieses einführende Kapitel dient dazu die Grundprinzipien und Elemente, die für die Erstellung von thematischen Karten von Bedeutung sind, darzulegen. In diesem Kontext werden zunächst die wichtigsten Bestandteile einer Karte erläutert, die bei der Erstellung von Karten benötigt werden. Die gestalterischen Elemente einer Karte werden mit Hilfe der kartographischen Variablen aufgezeigt.

In Bezug auf das Thema dieser Arbeit wird darauf Wert gelegt herauszufinden, welche gestalterischen Elemente bei der Erstellung von rasterbasierten Karten wichtig sind. Die daraus gewonnenen Erkenntnisse fließen in den Entwurf der Webapplikation zur Erstellung thematischer rasterbasierter Karten ein. In diesem Zusammenhang richtet sich das Augenmerk dieses Abschnittes besonders auf die Thematische Kartographie, da Karten auf Rasterbasis dieser Kategorie kartographischer Darstellungen angehören. Dabei soll einerseits aufgezeigt werden wie eine thematische Karte aufgebaut ist. Andererseits soll zum Vergleich zu Karten auf Rasterbasis exemplarisch aufgezeigt werden, welche anderen Darstellungsarten von Sachdaten es noch gibt. Darüber hinaus wird die Gestaltung thematischer Karten behandelt sowie bestehende Elemente einer Grundkarten beschrieben, die auch für die prototypische Erstellung der Online-Applikation von Bedeutung sind.

2.1 Grundlagen der Kartographie

Es gibt zahlreiche Definitionen zum Begriff Kartographie, die sich in Wesen, Inhalt und Umfang unterscheiden. Laut HAKE (1982: 13) bezeichnet Kartographie die „*Wissenschaft und Technik des Entwurfs, der Herstellung und des Gebrauchs kartographischer Darstellungen; diese vermitteln raumbezogene Informationen durch ein System geometrisch gebundener graphischer Zeichen*“. Neben der Karte als kartographische Ausdrucksform gibt es u.a. Luftbilder, Globen und Panoramen. (vgl. ebd.: 13f)

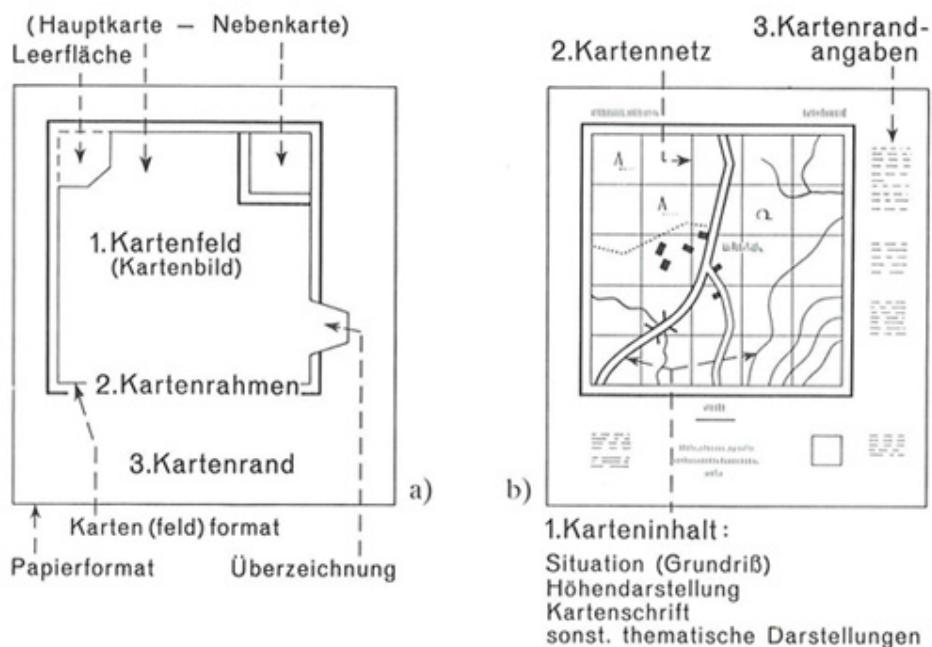
Kartographie hat die Aufgabe „*die theoretischen und praktischen Voraussetzungen dafür zu schaffen, dass kartographische Darstellungen entstehen, aus denen die Benutzer eine möglichst zutreffende Vorstellung oder Erkenntnis der vergangenen, gegenwärtigen oder geplanten Wirklichkeit gewinnen. Der Kartograph wird damit zum graphischen Dolmetscher raumbezogener Informationen*“ (ebd.: 14). Da sich die Umwelt ständig verändert, unterliegen kartographische Produkte ebenfalls ständigen Anpassungen. Diese Veränderungen fordern in weiterer Folge Fortschritte der Technik und Methodenlehre. (vgl. ebd.: 14)

2.1.1 Bestandteile einer Karte

Karten werden in formaler (äußerer) und sachlicher (innerer) Hinsicht in verschiedene Bestandteile gegliedert. Formale Bestandteile der Karte sind das Kartenfeld, der Kartenrahmen sowie der Kartenrand (siehe Abb. 3). Innerhalb des Kartenfeldes wird der

Karteninhalt dargestellt, der als Rahmenkarte oder Inselkarte gestaltet sein kann. Der Kartenrahmen bezeichnet die streifenförmige schmale Fläche zwischen Kartenfeld und äußerer Begrenzungslinie. Der Kartenrand wiederum beschreibt die Fläche außerhalb des Kartenrahmens und wird durch das Blattformat begrenzt. Der sachliche Karteninhalt beinhaltet die Gesamtheit der graphischen Strukturen und Informationen. Dieser besteht aus Situation, Höhendarstellung, Schrift und anderen thematischen Angaben. Unter Kartenrandangaben können alle textlichen und graphischen Darstellungen im Kartenrand und Kartenrahmen verstanden werden (z.B. Maßstab, Legende, Koordinatenwerte). (vgl. HAKE 1982: 26ff)

Diese formalen Regeln zu Erstellung von (thematischen) Karten werden auch bei der im Rahmen dieser Arbeit entwickelten Webapplikation eingehalten. Da die geographische Ausdehnung vorerst auf die Dimensionen von Österreich bzw. Europa beschränkt ist und die Datengrundlage nicht vollständig vorhanden ist, werden die Möglichkeiten zur Gestaltung der Inhalte und des Kartenrandes jedenfalls eingeschränkt sein.



Quelle: HAKE 1982: 27

Abb. 3: Bestandteile einer Karte

Um Karten auf der ebenen Fläche eines Papiers oder Bildschirms abbilden zu können, müssen zunächst mathematische Vorschriften zur Projektion der wahren Erdgestalt definiert werden. Dies wird durch Kartennetzentwürfe gewährleistet, die nach verschiedenen Gesichtspunkten eingeteilt werden können. Hinsichtlich der Parameter von Netzentwürfen kann zwischen geographischen und geodätischen Netzen unterschieden werden. Geographische Netze gelten als Repräsentation des globalen Koordinatensystems und sind für kleinmaßstäbige Karten geeignet. Geodätische Netze sind fix auf dem Rotationsellipsoid verankert und Grundlage großmaßstäbiger (ab ca. 1:500.000) Karten. Ferner können Kartennetze nach der Lage zwischen Erdkörper und Abbildungsfläche

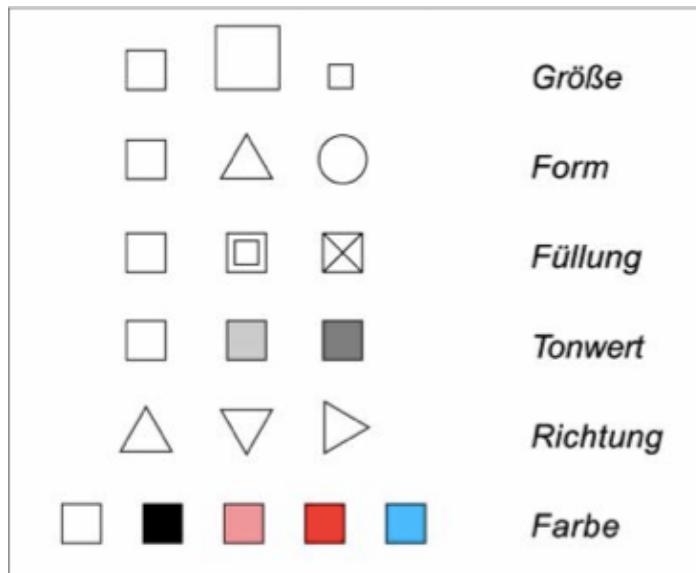
differenziert werden. In diesem Kontext sind normale (Erdachse fällt mit der Achse der Abbildungsfläche zusammen), transversale (Erdachse und Achse der Abbildungsfläche bilden rechten Winkel) und schiefachsige (beliebiger Winkel) Netzentwürfe zu unterscheiden. Die Art der Abbildungsfläche (konisch, azimutal oder zylindrisch) ist ebenfalls ein entscheidender Faktor für die Darstellung einer Karte. Hinsichtlich der Treueeigenschaften können noch längentreue (äquidistante), winkeltreue (konforme) und flächentreue (äquivalente) Abbildungen unterschieden werden. (vgl. HAKE et al. 2002: 53ff)

Die Wahl der Projektion für die Webapplikation hängt jedoch nicht nur von den benötigten Treueeigenschaften und der darzustellenden Fläche ab, sondern auch von den verwendeten Daten und von länderspezifischen oder regionsspezifischen Konventionen. Darum wird im Zuge dieser Arbeit für alle erstellten Karten das Referenzsystem „ETRS89 / LAEA“ verwendet, welches auch als „European Grid“ oder „European Reference System“ bekannt ist. Dieses Referenzsystem basiert auf der „Lambert Azimuthal Equal-Area“ Projektion (vgl. WIKIPEDIA 2015). In Bezug auf die oben beschriebenen Kriterien zur Einordnung von Kartennetzen ist dieser Entwurf ein geodätisches Netz mit azimutaler Abbildung, in normaler Lage und einer flächentreuen Eigenschaft. Zur Anwendung kommt es vor allem auch deshalb, weil dieses Referenzsystem von der Europäischen Kommission im Zuge der INSPIRE-Direktive empfohlen wird. Weitere Information zu INSPIRE werden im Kapitel 4 gegeben.

2.1.2 Kartographische Gestaltung und Visualisierung

Karten werden mithilfe der Kartographie dargestellt, die als Gesamtheit der für Karten aller Art typischen Darstellungsweisen gilt. Die Kartenzeichen liefern Informationen über den Raumbezug sowie den Eigenschaften von Objekten. Um einzelne Zeichen voneinander trennen und bestimmte Sachinhalte darstellen zu können, werden die Variablen in graphischer Hinsicht variiert (siehe Abb. 4). Die sechs Variablen der Objektmerkmale können wie folgt beschrieben werden: (vgl. HAKE et al. 2002: 108f)

- *Größe (Breite): Aufzeigen unterschiedlicher Quantitäten; Betonen und Abschwächen.*
- *Form: Unterscheidung von Qualitäten; Erleichtert Assoziation bei bild- und symbolhaften Zeichen.*
- *Füllung: Geeignet für Qualität und Quantität.*
- *Tonwert: Beschreibung von Quantitäten durch Helligkeitsstufen; Vor allem für flächenhafte Relativzahlen; Meist in gestufter Weise.*
- *Richtung (Orientierung): Zur Aufgliederung von Merkmalen; Darstellung zeitlicher Inhalte.*
- *Farbe: Beschreibung verschiedener Qualitäten als Farbton; Beschreibung verschiedener Qualitäten als Farbsättigung und Farbhelligkeit; Beschreibung zeitlichen Verhaltens.*



Quelle: HAKE et al. 2002: 108

Abb. 4: Kartographische Variablen der Objektmerkmale

Diese sechs Variablen lassen sich jedoch nicht ausschließlich auf rasterbasierte kartographische Darstellungen umlegen, da hierbei nur großteils quantitative Themen dargestellt werden sollen. Es kann z.B. bei rasterbasierten Abbildungen die Größe oder die Form der einzelnen Rasterzellen verändert werden. Da diese Änderungen jedoch auf alle Zellen Anwendung finden, lassen sich daraus keine qualitativen oder quantitativen Aussagen ableiten. Ebenso spielt die Richtung keine Rolle. Im Gegensatz dazu haben die restlichen drei Variablen einen großen Einfluss auf das darzustellende Thema, wobei auf den Tonwert und die Farbe vermehrt Wert gelegt wird. So sollen dem Ersteller einer Karte bereits vorgefertigte Farbskalen, auf Basis der Anzahl der verwendeten Klassen, zur Verfügung gestellt werden. Diese sind so aneinander angepasst um harmonische Verläufe zu ermöglichen.

Signaturen können punkt-, linien- oder flächenhaft gestaltet sein. Im Zuge dieser Arbeit sind jedoch neben der Darstellung von Grenzen als Linien bzw. der Darstellung von Orten als Punkte für die Aussage des Themas hauptsächlich flächenhafte Signaturen von Bedeutung. Laut HAKE (1982: 210ff) gestatten Flächendarstellungen Aussagen über die Lage und Qualität flächenhafter Diskreta sowie Angaben über flächenbezogene Quantitäten und über Wertstufen eines Kontinuums.

Ein weiteres wichtiges Gestaltungsmittel im Zuge der Kartenerstellung ist der Prozess der kartographischen Generalisierung. Dabei werden zur Steigerung der Aussagekraft bestimmte Objekte der Natur auf Kosten der Lagegenauigkeit vereinfacht dargestellt. Jede Karte entspricht einer generalisierten Darstellung der Wirklichkeit. Im Zuge der Kartenstellung haben der Maßstab und die Zweckbestimmung einen maßgeblichen Einfluss auf die kartographische Abbildung. Der Generalisierungsprozess besteht aus verschiedenen elementaren Vorgängen, wobei die folgenden sieben Gruppen unterschieden werden können: (vgl. HAKE 1982: 223ff)

- *Vereinfachen (mit der wichtigen Untermenge glätten),*
- *Vergrößern (mit der wichtigen Untermenge verbreitern),*
- *Verdrängen (als Folge des Vergrößerns),*
- *Zusammenfassen,*
- *Auswählen,*
- *Klassifizieren bzw. Typisieren,*
- *Bewerten.*

Im Kontext dieser Arbeit kommt die kartographische Generalisierung ebenfalls zur Anwendung. Diese findet jedoch großteils bereits bei der Aufbereitung der Grundlagendaten statt, da die Karten auf Basis fixer Maßstäbe erstellt werden. So werden zum Beispiel Ländergrenzen vereinfacht oder Beschriftungen von kleineren Städten weggelassen, ohne dabei eine korrekte Darstellung der Inhalte zu gefährden. Eine Zusammenfassung von Daten findet jedoch auch während des Erstellungsprozesses statt. So kann der Benutzer der Anwendung bspw. auswählen, auf Basis welcher Rastergröße die adressgenauen Daten aggregiert werden sollen.

2.1.3 Bildschirmkartographie

Das Internet bietet eine große Anzahl an Möglichkeiten zur Visualisierung von Geodaten und der darauf basierenden Vermittlung von raumbezogenen Informationen. Voraussetzung für die Akzeptanz webbasierter multimedialer und interaktiver kartographischer Informationssysteme ist die Anwendung einer klaren, lesbaren sowie einer dem Maßstab und den technischen Anforderungen des Ausgabemediums angepassten Kartographik. Da die Anzeige der Karte auf dem Bildschirm nicht mehr lediglich ein Zwischenprodukt der gedruckten Karte ist, wird besonderes Augenmerk auf die Kartographik gelegt. Dafür müssen jedoch die Eigenheiten von Computerbildschirmen beachtet werden. Neben der Einhaltung der Mindestdimensionen muss jedenfalls eine geeignete Farb-, Form-, Richtungs-, Muster- und Schriftwahl beachtet werden. Bei der Farbwahl für die Darstellung am Bildschirm bestehen keine technischen Einschränkungen. Lediglich die Differenzierbarkeit verschiedener Farben sollte bedacht werden. Bei der Schriftwahl sollten seriflose Schriften eingesetzt werden, die möglichst parallel verlaufen. Hinsichtlich der Form-, Richtungs- und Musterwahl sind folgende Punkte zu beachten: (vgl. LECHTHALER und STADLER 2006: 248ff)

- *Verwendung von rechteckigen und quadratischen Punktsignaturen*
- *Ausrichtung nach der Bildpunktmatrix (keine schrägen Punktsignaturen oder Linien)*
- *Vermeidung von Linienmustern (z.B. strichliert, Doppellinie)*
- *Verwendung von Flächenfarben anstelle von Flächenmustern*
- *geringer Einsatz von linearen Graphikelementen (z.B. Umrandungen von Flächen)*

Das Ziel, der in dieser Arbeit entwickelten Online-Anwendung, ist die Erstellung von thematischen Karten, welche für den analogen Druck optimiert sind. Deswegen müssen hierbei sowohl die Empfehlungen für analoge als auch für digitale kartographische Produkte beachtet werden, wobei die Vorgaben für digitale Werke hinsichtlich deren

Gestaltung wesentlich restriktiver sind. Zwar stellen die hergestellten Karten keine interaktiven Webprodukte dar, jedoch müssen sie ebenfalls zur Verwendung im Internet geeignet sein.

2.2 Thematische Kartographie

Die Gesamtheit an Kartenwerke kann grundsätzlich in zwei Hauptgruppen, nämlich topographische und thematische Karten, untergliedert werden. Topographische Karten spiegeln im Wesentlichen die Form, Gliederung und Bedeckung des Geländes wider, wohingegen thematische Karten vorwiegend Sachverhalte nicht topographischer Art, die jedoch einen räumlichen Bezug besitzen, beinhalten. (vgl. IMHOF 1972: 12ff)

Thematische Karten können darüber hinaus in analytische und synthetische Karten untergliedert werden. Erstere stellen einzelne Betrachtungsgegenstände, wie z.B. die Bevölkerungsdichte, Niederschlagsmenge, usw. dar. Werden mehrere Einzeldarstellungen in einer Karte vereinigt (z.B. Darstellung von Isothermen, Reisanbaugebieten und Diagrammen lokaler Regenmengen in einer Karte) ist von komplexen Karten die Rede. Synthetische Karten beinhalten hingegen korrelative Gefüge von Landschaften oder Regionen (z.B. Darstellungen naturräumlicher oder kulturräumlicher Zustände). (vgl. ebd.: 14f)

Die Bestandteile von Karten, wie bspw. das Kartenfeld bzw. der Kartenrahmen und die Kartenrandangaben, etc., wurden bereits beschrieben und finden sich auch in thematischen Karten wieder. Inhaltlich betrachtet kann eine Unterscheidung zwischen den topographischen Geometrie- bzw. Orientierungsdaten und den Sachdaten des gewählten Themas getroffen werden. Die Online-Applikation unterstützt in einer ersten Version nur die Erstellung von einschichtigen analytischen thematischen Karten. Auf Grund der einfachen Erweiterbarkeit der Anwendung werden jedoch verschiedene Arten thematischer Karten erklärt.

2.2.1 Die Grundkarte

Die kritische Beurteilung und Bereitstellung von Kartengrundlagen für thematische Karten kann als Voraussetzung jeglicher thematisch-kartographischer Arbeit angesehen werden. Oftmals besteht in der Auswahl und Bearbeitung sogar der Hauptteil eines Kartenentwurfs. Neben den Kartengrundlagen sind hierbei auch die Grundlagen der thematischen Daten zu nennen. Häufig werden in Karten Ergebnisse statistischer Zählungen oder Forschungsergebnisse dargestellt, wobei entsprechende Daten zunächst gesammelt und aufbereitet werden müssen. (vgl. IMHOF 1972: 15f)

Die topographische Grundkarte hat für jede thematische Karte große Bedeutung, da sowohl die Genauigkeit der thematischen Inhaltsgliederung als auch die Interpretation durch den Kartenleser von der Qualität der Grundlage abhängig ist (vgl. KELNHOFER 1971: 89). Für die Erstellung thematischer Karten wird eine Kartengrundlage mit den wichtigsten topographischen Elementen benötigt. In diesem Kontext kommen gedruckte bzw. digitale

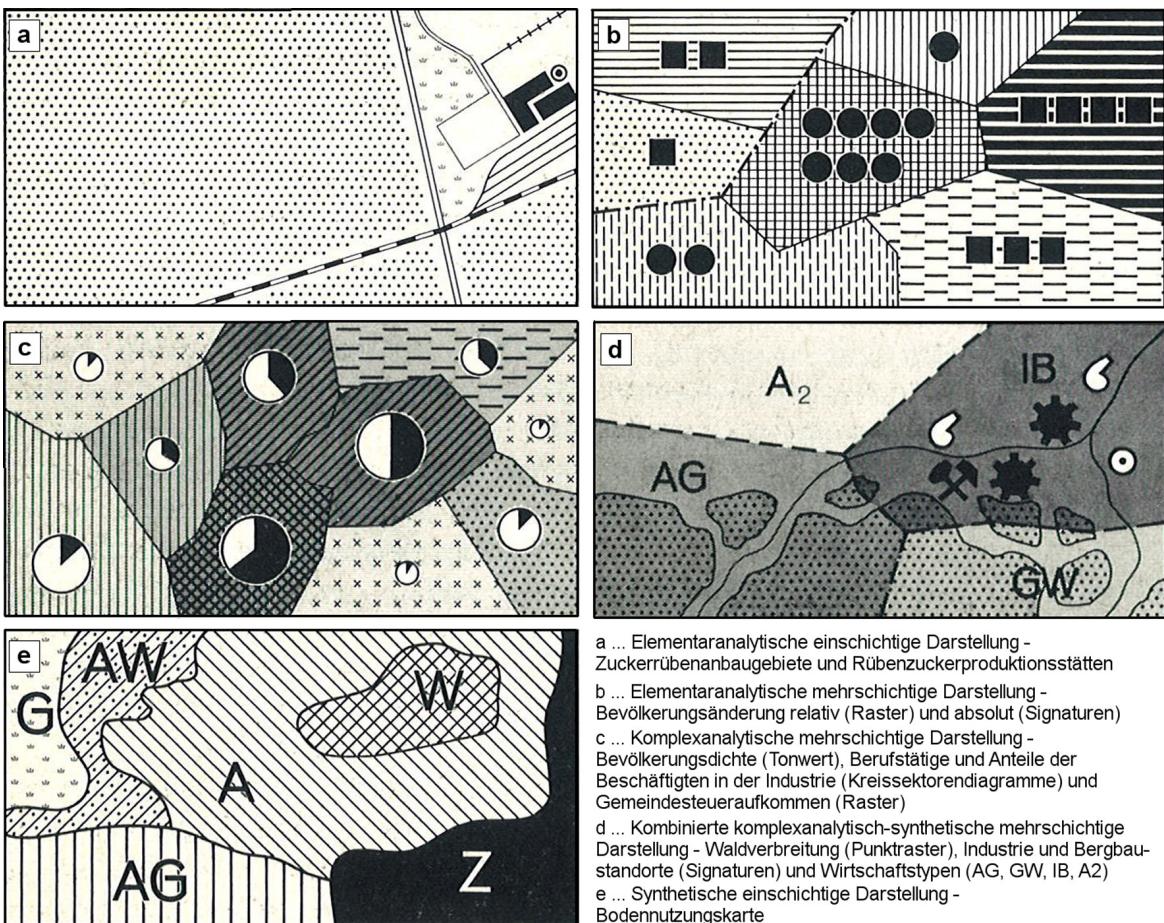
topographische oder thematische Karten wie auch Luftbilder oder Fernerkundungsdaten in Frage (vgl. OLBACH et al. 2002: 25f).

Unter diesen Gesichtspunkten wurde auch der Erstellung der jeweiligen Grundkarten für die Webapplikation erhöhte Aufmerksamkeit geschenkt. Der Benutzer kann bei der Erstellung der Karte zwischen verschiedenen Elementen wie der Beschriftung von Ländern oder Städten, der unterschiedlichen Darstellung von Gewässerflächen oder der Anzeige einer Geländeschummerung wählen. Die restlichen Objekte sind vordefiniert und können nicht geändert werden. Auf Grund von fehlenden Daten können in einer ersten Version keine weiteren Elemente für die Grundkarte angeboten werden.

Die vorhandenen Daten wurden jeweils für die Darstellung auf europäischer bzw. österreichischer Ebene bearbeitet und angepasst. So wurden z.B. Generalisierungen vorgenommen, die Position von Schriften definiert, und auch Symbolisierungen für Punkte und Linien vordefiniert.

2.2.2 Entwurf und Gestaltung thematischer Karten

Die Aussagekraft einer thematischen kartographischen Darstellung wird entscheidend beeinflusst durch die Wahl des Aussageweges. Die Darstellung und somit auch die Art der Wahrnehmung hängen im Wesentlichen davon ab, ob eine elementaranalytische, komplexanalytische, kombinierte komplex-analytisch-synthetische oder eine synthetische Aussage auswählt wurde. Hinsichtlich der Anzahl an dargestellten Themen kann zwischen einschichtigen und mehrschichtigen thematischen Karten unterschieden werden. Abb. 5 zeigt in diesem Zusammenhang einfache einfarbige Beispiele für verschiedenartige thematische Karten. (vgl. ARNBERGER 1977: 17)



Quelle: bearbeitet nach ARNBERGER 1977: 18

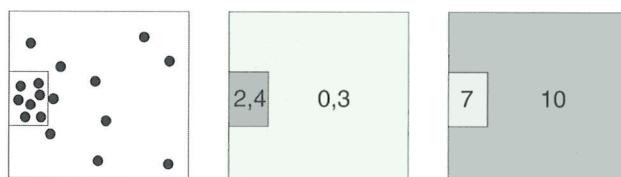
Abb. 5: Arten thematischer Karten

Es gibt eine Vielzahl an stilistischen Möglichkeiten, wie Daten in thematischen Karten dargestellt werden können. Für die Darstellung auf dem Bildschirm sind dies vor allem Standortkarten, Choroplethenkarten und Diagrammkarten (vgl. OLBRICH et al. 2002: 35f). Da im Zuge dieser Arbeit lediglich spezielle Choroplethenkarten erstellt werden können, wird diesen Darstellungsformen auch erhöhte Aufmerksamkeit geschenkt. Auf die Frage, warum es sich um spezielle Choroplethenkarten handelt, wird nach deren allgemeinen Beschreibung näher eingegangen.

In Standortkarten werden qualitative Sachverhalte mit Hilfe von punkt-, linien- oder flächenhaften Signaturen graphisch repräsentiert. Hinsichtlich flächenhaft auftretender Sachverhalte sind Choroplethenkarten, welche auch als Flächenkartogramme bezeichnet werden, die am häufigsten verwendete Darstellungsform. In diesem Zusammenhang ist zu beachten, dass die Grenzziehung häufig willkürlich oder unscharf ist, bspw. bei Bodenkarten, in denen der vorherrschende Bodentyp dargestellt wird, wobei das Vorhandensein anderer Bodentypen nicht kategorisch ausgeschlossen werden kann. In Choroplethenkarten werden vorwiegend qualitative Relativwerte abgebildet. Bei der Darstellung von Absolutwerten (z.B. Bevölkerungszahlen) in flächenhafter Form kann es auf Grund unterschiedlicher Bezugsgrößen zu Fehlinterpretationen kommen (siehe Abb. 6). Eine wesentliche Eigenschaft von Choroplethenkarten ist, dass die Daten meist

klassifiziert dargestellt werden, wobei bestimmte Grundsätze beachtet werden müssen: (vgl. OLBRICH et al. 2002: 38ff)

- *Vollständigkeit: Die Klassen müssen alle Werte der Datenreihe enthalten.*
- *Eindeutige Klassengrenzen: Für jeden Wert ist die Zuordnung zu genau einer Klasse möglich.*
- *Klassenzahl: Die Anzahl der Klassen sollte in einem angepassten Verhältnis zu den Ausgangsdaten gewählt werden.*
- *Methode der Klassenbildung: Das Verfahren ist an die Ausgangsdaten angepasst.*
- *Genauigkeit der Klassengrenzen: Die Methoden der Klassenbildung liefern oft Klassengrenzen mit einigen Kommastellen. Diese sollten gerundet werden.*
- *Wenig Informationsverlust: So viele Informationen wie möglich müssen erhalten bleiben.*



Quelle: bearbeitet nach OLBRICH et al. 2002: 39

Abb. 6: Fehlinformation durch Darstellung von Absolutwerten in einer Choroplethenkarte

Wie bereits erwähnt, handelt es sich, nach Ansicht des Autors dieser Arbeit, bei rasterbasierten Darstellungen um eine Spezialform von Choroplethenkarten. Die Grundsätze für die Klassifizierung der Daten müssen dabei jedoch in gleichem Ausmaß beachtet werden. Anbei werden zwei wichtige Gründe angeführt, warum es sich um spezielle Flächenkartogramme handelt:

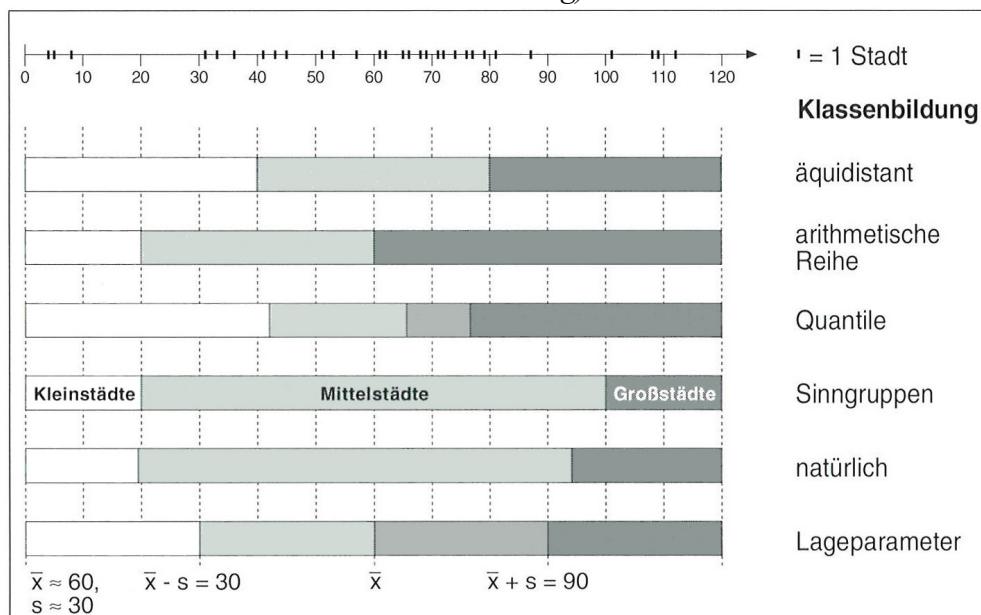
- Die Grenzziehung bei Karten auf Basis von Rastereinheiten erfolgt nach einem definierten Muster und ist somit immer unabhängig von reellen Objekten. Es können deshalb keine echten Grenzen von Verteilungen dargestellt werden.
- Bei Flächenkartogrammen müssen Absolutwerte immer in Bezug auf deren Fläche dargestellt werden. Da die Bezugsgrößen bei rasterbasierten Darstellungen immer gleich groß sind können auch Absolutwerte als Flächenfärbung dargestellt werden.

Vor einer Datenklassifizierung sollte festgelegt werden, wie viele Klassen gebildet werden sollen und wie die Klassengrenzen festzulegen sind. Die Wahl der Klassifizierungsmethode hat direkten Einfluss auf die Darstellung der Daten. Bei einfarbigem Druck sollten sechs bis sieben Klassen nicht überschritten werden. Bei einem mehrfarbigen Druck können bis zu zwölf Klassen gebildet werden (vgl. OLBRICH et al. 2002: 40 ff). Um die Gestaltungsmöglichkeiten nicht zu sehr einschränken zu müssen, jedoch die farbliche Wiedergabe der verschiedenen Klassen noch handhaben zu können, werden im weiteren Verlauf des Anwendungsentwurfs die Anzahl der Klassen auf drei bis sieben begrenzt.

Auch die Wahl der Klassengrenzen spielt eine große Rolle für die korrekte und aussagekräftige Darstellung thematischer Inhalte. Mit verschiedenen Klassifizierungen können Sachverhalte teilweise komplett anders dargestellt werden. Auch hier wird auf eine

relativ offene Gestaltung der Karte im Ablauf des Entwurfs eine rasterbasierten Karte gesetzt. Die Klassengrenzen können in der Online-Applikation frei definiert werden. Vorteil dieser Herangehensweise ist, dass ein relativ großer Handlungsspielraum bei der Erstellung der Karte gegeben ist. Ein Nachteil ist jedoch, dass eventuell andere Tools benötigt werden, um die Daten zu klassifizieren, bzw. die Klassengrenzen zu eruieren, oder um die Daten manuell zu analysieren. OLBRECH et al. (2002: 42ff) definiert und erklärt sechs gängige Methoden der Klassifizierung von Daten.

- *Äquidistante Klassen: Alle Klassen haben die gleiche Breite, d.h. den gleichen Abstand zwischen der oberen und unteren Klassengrenze.*
- *Mathematische Progression: Alle Verfahren, die auf Grundlage von mathematischer Reihenbildung entstehen (z.B. arithmetische oder geometrische Reihe).*
- *Quantile: Hier werden die Grenzen so gewählt, dass in jede Klasse gleich viele Werte fallen.*
- *Sinnklassen: Die Klassenbildung erfolgt auf Basis einer sachlogischen Begründung (z.B. Grenzwerte oder Klassenbildungen der amtlichen Statistik liegen vor).*
- *Natürliche Klassen („natural breaks“): Mit Hilfe von mathematischen Algorithmen werden Datenlücken genutzt, um Klassen zu definieren.*
- *Statistische Parameter: Klassenbildung mit Hilfe statistischer Parameter (z.B. arithmetisches Mittel und Standardabweichung)*



Quelle: bearbeitet nach OLBRECH et al. 2002: 45

Abb. 7: Methoden der Klassenbildung anhand der Verteilung der Einwohnerzahlen von 30 Städten

Im Folgenden werden noch kurz Diagrammkarten erklärt, da Karten auf Rasterbasis durch Diagramme ergänzt werden können. Die Definition ist jedoch absichtlich kurz gehalten, da sie in dieser Version der Applikation noch keine Anwendung finden. Nach OLBRECH et al. (2002: 47ff) sind Diagrammkarten thematische Karten, die Diagramme in räumlicher Anordnung enthalten. Sie verbinden die Vorteile der Karte und des Diagramms und

werden sowohl für die Darstellung der Raumstrukturen als auch zur Verdeutlichung der inhaltlichen Beziehungen verwendet. Es können vier verschiedene Typen von Diagrammkarten identifiziert werden: Diagramme für Einzelwerte, Balken- und Säulendiagramme, Kurvendiagramme und Kreisidagramme.

3 Aufbereitung adressgenauer Daten durch Geokodierung

Da Adressdaten nicht direkt auf einer Karte abgebildet werden können, müssen diese im Vorfeld der Kartenerstellung aufbereitet werden. Im Zuge der Aufbereitung werden Adressen in Koordinaten umgewandelt, welche dann in aggregierter Form oder als direkte Adresspunkte vorliegen und somit in einer Karte dargestellt werden können.

In diesem Kapitel wird erläutert, was der Begriff Geokodierung bedeutet bzw. welche Arten der Geokodierung es gibt, wobei die in dieser Arbeit vorwiegend verwendeten Methoden genauer beschrieben werden. Nach einer Übersicht von geeigneten Geokodierungsdiensten erfolgt eine Evaluierung derselben auf Basis zuvor definierter Parameter. Das Resultat wird mit vorliegenden Studien verglichen sowie um deren Erkenntnisse erweitert. Zum Abschluss dieses Kapitels werden die Ergebnisse diskutiert und interpretiert, sowie ein geeigneter Online-Geokodierungsdienst ausgewählt.

Die gewonnenen Erkenntnisse fließen direkt in die, im Zuge dieser Arbeit entwickelte, Online-Applikation ein, da diese ebenfalls Funktionalitäten zur Geokodierung von Adressen beinhaltet. Es werden auf dieser Plattform zwar mehrere Online-Geokodierungsdienste zur Verfügung stehen, der am besten geeignete Dienst soll jedoch als Standard-Service ausgewählt werden.

3.1 Einführung in Geokodierung

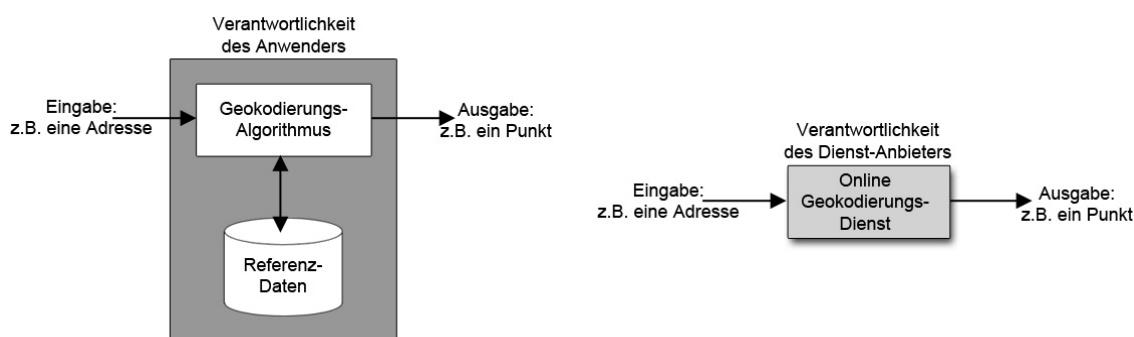
Die Verortung der Lage von Objekten zählt zu den grundlegenden Prozessen eines Geoinformationssystem (GIS), wodurch die Lagerichtigkeit der Kartenelemente garantiert werden soll. In diesem Zusammenhang müssen Adressen in geographische Koordinaten umgewandelt werden. Die Geokodierung von Adressen ist jener Vorgang, bei dem Postanschriften durch die Zuweisung von Lageangaben in ein räumliches Bezugssystem überführt werden. Für diesen Vorgang werden mindestens folgende drei Komponenten benötigt: Adresse, Geokodierer (Software, Algorithmus, Webplattform o.ä.) und Basisdaten mit bekannter Georeferenzierung (vgl. SCHULTE et al. 2010: 773). Der Prozess der Geokodierung gliedert sich in die Übersetzung von Adresseinträgen, die Suche nach der Adresse im Referenzdatensatz sowie der Rückgabe des am besten geeigneten Kandidaten. Dieser Prozessablauf beinhaltet im Detail die folgenden Schritte: (vgl. ZANDBERGEN 2008: 215)

1. *Die Zergliederung der eingegebenen Adressen in die einzelnen Komponenten. (Straßenname, Hausnummer, etc.)*
2. *Die Standardisierung von gängigen Abkürzungen.*
3. *Die Zuordnung der Adresselemente zu einer Kategorie.*
4. *Die Indizierung der Kategorien.*
5. *Die Durchsuchung der Referenzdaten.*
6. *Die Bestimmung der Wertung für jeden möglichen Kandidaten.*
7. *Die Filterung der Kandidatenliste basierend auf den minimalen Trefferraten.*
8. *Die Übertragung des besten Ergebnisses.*

Das Resultat der Geokodierung sind geographische Koordinaten, welche die exakte Position einer postalischen Adresse auf der Erde repräsentieren.

3.2 Methoden der Geokodierung

Der Geokodierungsvorgang kann grundsätzlich in zwei Methoden unterteilt werden – die konventionelle Offline-Geokodierung sowie die Online-Geokodierung. Während bei der konventionellen Geokodierung der Hauptteil der Arbeit und zugleich die Verantwortung (z.B. die Aktualisierung der Referenzdaten, die Nachbearbeitung der gelieferten Ergebnisse) beim Anwender liegt, ist im Gegensatz dazu im Zuge der Online-Geokodierung der Anbieter des Dienstes vorwiegend für die Qualität der gelieferten Ergebnisse verantwortlich (vgl. Abb. 8).



Quelle: nach ROONGIBOONSOPI und KARIMI 2010: 1084.

Abb. 8: Prozess der Offline- (links) und Online-Geokodierung (rechts)

Wie bereits erwähnt, ist die Geokodierung jener Prozess, bei dem geographische Koordinaten einem Ortsnamen durch Vergleich der Beschreibung mit ortsbezogenen Elementen in der Referenzdatenbank zugewiesen werden. Bei der Offline-Geokodierung wird die Software, die diesen Vorgang durchführt als „Geokodierer“ bezeichnet. Dieser kann in vier Komponenten unterteilt werden: Eingabe, Ausgabe, Bearbeitungsalgorithmus und die Referenzdatenbank. Die Eingabe beinhaltet die Beschreibung eines Objektes oder Ortes mit räumlicher Information. Die Ausgabe liegt in Form von geographischen Koordinaten vor. Der Bearbeitungsalgorithmus ist der wichtigste Teil bei der Offline-Geokodierung, da hierbei die Eingabe übersetzt und in einzelne Teile zerlegt sowie die geeignete Adresse in den Referenzdaten gesucht und ein geographischer Referenzcode zurückgegeben wird. Die Bezugsdatenbank besteht aus geographischen Elementen die vom Bearbeitungsalgorithmus benutzt werden, um ein Ergebnis zu erhalten. (vgl. ROONGIBOONSOPI und KARIMI 2010: 1083)

Für den Aufbau solcher Datenbanken stehen mehrere Adressmodelle, wie bspw. das Straßennetzwerk-Datenmodell, das Parzellengrenzen-Datenmodell (z.B. Postleitzahlen, Bezirke, Städte) oder das Adresspunkt-Datenmodell zur Verfügung (vgl. ZANDBERGEN 2008: 216ff).

Ein Online-Geokodierungsdienst ist ein über das Netzwerk erreichbares Service, das den Geokodierungsprozess automatisch durchführt. Die teilweise in GIS eingebauten Dienste sind meist über das Internet zugänglich, wobei die Eingabe über ein definiertes Kommunikationsprotokoll innerhalb von nur wenigen Sekunden bzw. Bruchteilen von Sekunden pro Eintrag abgearbeitet wird. Als Ergebnis werden geographische Koordinaten der Eingabe, die verwendete postalische Adresse und einen Genauigkeitswert ausgegeben. Der Algorithmus und die Referenzdatenbank sind bei Online-Geokodierungsdiensten fix definiert, können jedoch vom Nutzer eingesehen werden. (vgl. ROONGPIBOONSOPIT und KARIMI 2010: 1083)

Die Verwendung von Online-Geokodierungsdiensten unterscheidet sich in vielerlei Hinsicht von konventionellen Methoden der Geokodierung. In erster Linie ist es einfacher für den Anwender Online-Dienste zu verwenden, da keine Vorkenntnisse in Hinblick auf den verwendeten Algorithmus sowie die zu verwendenden Referenzdaten vorausgesetzt sind. Die Aktivität des Benutzers beschränkt sich lediglich auf die Eingabe der Adresse sowie die Interpretation des Ergebnisses. Des Weiteren muss sich der Anwender keine Gedanken über die Aktualisierung der Referenzdatenbank machen, da diese beim Anbieter gespeichert und gewartet wird. Dies kann sich jedoch auch als Nachteil erweisen, da keine Korrekturen der Bezugsdaten gemacht werden können. Die Anbieter solcher Geokodierungsdienste sind jedoch bestrebt ihre Software weiterzuentwickeln, um gegenüber anderen Services konkurrenzfähig zu sein. Eine kurze Zusammenfassung der Vorteile und Nachteile von Online-Geokodierungsdiensten bietet Tab. 1. (vgl. ebd.: 1083f)

Vorteile	Nachteile
1. Einfach zu verwenden	1. Keine Kontrolle über Referenzdaten
2. Schnelle Ergebnisse in Form von Koordinaten	2. Keine Kontrolle über Parameter des Geokodierungsprozess (z.B. Trefferraten, etc.)
3. Der Nutzer muss sich nicht um Anschaffung, Wartung und Aktualisierung der Referenzdaten kümmern	3. Unbekannte Qualität der Geokodierungsergebnisse
4. Der Anwender benötigt keine Software	4. Abhängig von Internet-Infrastruktur

Quelle: nach ROONGPIBOONSOPIT und KARIMI 2010: 1084.

Tab. 1: Vorteile und Nachteile von Online-Geokodierungsdiensten

3.3 Geokodierungsqualität

Um ein Geokodierungsergebnis richtig interpretieren zu können, müssen bestimmte Qualitätsstandards eingehalten werden. Obwohl sich verschiedenste Wissenschaften bereits Geokodierungsdienste zu Nutze machen, wird den Fehlern, die bei der Geokodierung entstehen können, nur wenig Aufmerksamkeit geschenkt. I.d.R. werden in Bezug auf die Qualität eines Dienstes nur die Trefferraten thematisiert. Hohe Trefferraten können in diesem Kontext als signifikanter Parameter angesehen werden. Die Qualität eines geokodierten Datensatzes kann jedoch am besten mit Hilfe der drei Faktoren Vollständigkeit, Lagerichtigkeit und Wiederholbarkeit beschrieben werden. Die Vollständigkeit bzw. die Trefferrate, bezeichnet den Prozentsatz der Daten, die zuverlässig geokodiert werden können. Die Lagerichtigkeit kennzeichnet hingegen die Lagedifferenz zwischen dem

geokodiertem Punkt und der tatsächlichen adressbezogenen Position des Objektes. Unter Wiederholbarkeit können die Auswirkungen auf die Ergebnisse durch Änderungen des Straßennetzwerkes oder des Algorithmus der Software angesehen werden. Qualitativ hochwertige Geokodierungsergebnisse sind somit vollständig, lagerichtig und wiederholbar. Darüber hinaus können drei Fehlerquellen kategorisiert werden. (vgl. ZANDBERGEN 2008: 218)

- *Fehler in den Eingabedaten*
- *Fehler in den Referenzdaten*
- *Fehler im zugrundeliegenden Geokodierungsprozess*

Der einfachste Weg die Qualität eines Geokodierungsergebnisses zu verifizieren, ist die Ermittlung des prozentualen Anteils verlässlicher Ergebnisse. Da hinsichtlich der Trefferrate noch keine Standards vorliegen, variieren hier die Ergebnisse deutlich. Die Trefferraten eines Datensatzes können am einfachsten durch die Verbesserung bzw. Vervollständigung des Adressdatensatzes erhöht werden. In diesem Zusammenhang sind Interpretationen allerdings subjektiv. Wird der Prozentsatz, ab dem ein Resultat als valide angesehen werden kann, verringert, erhöht sich zwar die Trefferrate, wobei ungenaue Ergebnisse jedoch nicht auszuschließen sind (vgl. ebd.: 652f). Während bei den meisten Studien die Trefferraten stark variieren, wurde festgestellt, dass diese in ländlichen Gebieten viel geringer sind als in städtischen Bereichen. (siehe z.B. CAYO und TALBOT 2003; DEARWENT et al. 2001)

Die Lagerichtigkeit wird definiert als die euklidische Distanz zwischen einem geokodierten Punkt und der tatsächlichen Position des Ortes, der mit der angegebenen Adresse in Verbindung gebracht wird. In diesem Kontext können unterschiedliche Ursachen für das Auftreten von Lagefehlern ausfindig gemacht werden. (vgl. ZANDBERGEN 2009: 654f)

- *Falsche Straßensegment: Durch Fehler in der eingegebenen Adresse oder Fehler im zugrundeliegenden Referenzdatensatz könnte die Adresse einem falschen Straßensegment zugeordnet werden. Dies führt teilweise zu sehr großen Lagefehlern.*
- *Falsche Platzierung im Straßensegment: Die Interpolationsmethoden von Geokodierungsalgorithmen nehmen an, dass die Hausnummern im Referenzdatensatz gleichmäßig auf ein Straßensegment verteilt sind und es keine Unregelmäßigkeiten gibt. Diese Fehlerart hängt stark von der Länge des Segmentes ab.*
- *Falscher Abstand vom Straßensegment: Viele Geokodierungsalgorithmen nehmen einen Abstand von zehn bis 15 Meter eines Gebäudes von der Mittellinie einer Straße an, was meist nicht den tatsächlichen Abstand in der reellen Welt entspricht.*
- *Lagefehler im Straßensegment: Jegliche Lagefehler im Straßensegment können sich auf die Fehlplatzierung von geokodierten Orten auswirken.*

Da in den meisten Studien diese Fehler nicht einzeln ausgewiesen sind, setzt sich der Lagefehler aus den oben erwähnten vier Ursachen zusammen. Hinsichtlich der Lagerichtigkeit der Ergebnisse sind Kenntnisse über Verbesserungspotentiale unerlässlich. (vgl. ebd.: 655)

Konträr zur Lagerichtigkeit wird der Qualitätsfaktor hinsichtlich der Wiederholbarkeit eines Geokodierungsprozesses kaum Beachtung zuteil (vgl. ebd.: 219). Da in dieser Arbeit lediglich Online-Geokodierungsdienste in Anspruch genommen werden, liegt auch hierbei das Hauptaugenmerk nicht auf der Ermittlung der Wiederholbarkeit. Wie bereits erwähnt, bieten Online-Dienste weder Einfluss auf den Algorithmus noch auf die zugrundeliegende Referenzdatenbank.

3.4 Online-Geokodierungsdienste

Bisher wurde der Vorgang der Geokodierung i.d.R. mit spezieller Software von Experten durchgeführt. Die zunehmende Verbreitung von kartenbasierten Web-Anwendungen führte dazu, dass Online-Suchmaschinen selbst interaktive Geokodierungsapplikationen entwickelten und diese z.T. frei über ein API (application programming interface) zur Verfügung stellten. Wegen der Flexibilität in der Anwendung und der vielfältigen Implementationsmöglichkeiten sind Geokodierungsdienste heute einfach zu bedienen, weshalb deren Popularität stetig steigt. (vgl. DAVIS und ALENCAR 2011: 852)

Da in der vorliegenden Arbeit eine Webapplikation zur Durchführung und Evaluierung des Geokodierungsprozesses verwendet wird, liegt der Fokus auf Online-Geokodierungsdiensten. Die Anschaffung und Wartung einer Referenzdatenbank und somit die Geokodierung auf konventionelle Art würde den Rahmen dieser Untersuchung sprengen. Die hier verwendeten Dienste sind frei zugänglich und bis zu einem gewissen Grad kostenlos, wobei manchen Services bestimmte Lizenzierungsauflagen zugrunde liegen.

Im Folgenden werden die für die Arbeit relevanten und verfügbaren Online-Geokodierungsdienste für Europa aufgelistet:

- *Bing Maps Locations*
- *CloudMade Geocoding & Geosearch*
- *Google Maps Geocoding*
- *MapQuest Geocoding Service*
- *OpenStreetMap Nominatim*

In weiterer Folge werden die oben aufgelisteten Geokodierungsdienste hinsichtlich Kosten, Schnittstellen, Limitierungen und Lizenzierung untersucht. Die Ausführungen basieren im Wesentlichen auf den Angaben des jeweiligen anbietenden Unternehmens. Es handelt sich somit um eine rein deskriptive Aufzählung der Eigenschaften der Dienste. Eine Evaluierung hinsichtlich Genauigkeit und Performance wird im Kapitel 3.5 durchgeführt.

Die *Bing Maps Locations API* von Microsoft ermöglicht das Abrufen von geographischen Koordinaten, indem eine Adresse (z.B. Straße, Postleitzahl und Ort) über eine URL (Uniform Resource Locator) an den Dienst gesendet wird. Das Ergebnis der Anfrage beinhaltet die Koordinaten, den Typ des Ortes und die Region, in der sich die Adresse befindet. Das Ergebnis kann in den Formaten *JSON* (JavaScript Object Notation) und *XML* (Extensible Markup Language) angefordert werden. Die Limitierungen von Microsofts

Geokodierungsdienst sind in den Nutzungsbedingungen nicht eindeutig geklärt, wobei Entwicklern ein frei verfügbares Limit von 10.000 Anfragen pro Tag zur Verfügung gestellt wird. Des Weiteren ist die Verwendung der Geokodierungsergebnisse nur in Verbindung mit einer Bing-Karte zulässig. Darüber hinaus ist der Aufbau einer Datenbank auf Basis der ermittelten Koordinaten sowie die Bereitstellung dieser Daten als Download nicht gestattet. (vgl. MICROSOFT 2013)

Der *Bing Locations Dienst* verwendet mehrere Qualitätsparameter bei der Rückgabe von Geokodierungsergebnissen. Der „confidence“-Parameter (hoch, mittel, niedrig) wird aus mehreren Faktoren berechnet, wie bspw. der Wichtigkeit des geokodierten Ortes und der Position des Benutzers. Ein hoher Parameter bedeutet, dass eine oder mehrere „gute“ Übereinstimmungen für die Adresse gefunden wurden, wobei bei mehreren Ergebnissen diese nach Wichtigkeit gereiht werden. Der „matchCode“-Faktor (gut, mehrdeutig, hierarchisch) repräsentiert den Geokodierungslevel für jede Adresse im Ergebnis. Hierarchisch bedeutet in diesem Fall, dass kein Ergebnis für die eingegebene Adresse gefunden wurde und somit ein weniger präzises Resultat zurückgeliefert wurde. Ein weiterer Parameter zur Überprüfung der Genauigkeit eines Geokodierungsvorganges sind die Berechnungsmethoden, die an dieser Stelle kurz beschrieben werden. (vgl. MICROSOFT 2013)

- „*Rooftop*“: Der geokodierte Punkt wurde mit einem Gebäudedach in Verbindung gebracht.
- „*ParcelCentroid*“: Der geokodierte Punkt liegt in der Mitte der übereinstimmenden Parzelle.
- „*InterpolationOffset*“: Der geokodierte Punkt wurde auf Straßenebene interpoliert, wobei ein Offset miteinbezogen wurde, um den Punkt auf eine Straßenseite zu verschieben.
- „*Interpolation*“: Der geokodierte Punkt wurde auf Straßenebene interpoliert.

Der *CloudMade Geocoding & Geosearch Dienst* kann Geokodierungsanfragen sowohl in strukturierter als auch unstrukturierter Form annehmen, wobei für strukturierte Anfragen bestimmte Adressschlüsselwörter wie bspw. POI (Point of Interest), Postleitzahl, Stadt, etc. angegeben werden müssen. Um das Service nutzen zu können, ist allerdings eine Anmeldung erforderlich, wobei im Anschluss daran ein API-Schlüssel zur Verfügung gestellt wird. Damit ist es möglich pro Monat 100.000 Geokodierungsanfragen kostenlos zu stellen. Das Ergebnis einer Anfrage bei *CloudMade* kann in den Formaten *JSON*, *GeoJSON*, *Property List* oder *HTML* (Hypertext Markup Language) zur Darstellung auf einer Karte abgerufen werden. Die Benutzung der API erfolgt nach der *Creative Commons Attribution-ShareAlike 2.0 Lizenz*. (vgl. CLOUDEMADE 2013)

Mit Hilfe der *Google Geokodierungs-API* können pro Tag 2.500 Geokodierungsanfragen gestellt werden. Im Falle einer diesbezüglichen Überschreitung an Anfragen, kann der Zugriff auf den Dienst vorübergehend gesperrt werden. Der Geokodierungsdienst von *Google* ist frei und kostenlos verfügbar. Allerdings ist die Verwendung dieser API, ähnlich wie beim Bing Location Dienst, nur in Verbindung mit einer Google-Karte gestattet ist. (vgl. GOOGLE 2013)

Ergebnisse einer Anfrage können entweder im *JSON*-Format oder im *XML*-Format ausgegeben werden, wobei die Verwendung von *JSON* zu bevorzugen ist. Die beiden Ausgabeformate enthalten jeweils zwei Stammelemente „*status*“ und „*result*“. Das „*status*“-Feld enthält Informationen darüber, ob eine Anfrage erfolgreich war bzw. welche Fehler aufgetreten sind. Das „*result*“-Element beinhaltet hingegen das eigentliche Ergebnis, wobei dies nicht nur die resultierenden Koordinatenpaare sind, sondern auch die Parameter der geokodierten Adresse sowie die Genauigkeit der Geokodierung. *Google* verwendet in diesem Zusammenhang folgende vier Genauigkeitsstufen. (vgl. ebd.)

- „*Rooftop*“: Liefert präzise Ergebnisse mit Standortinformationen auf Straßenebene.
- „*Range Interpolated*“: Das Ergebnis stellt eine Annäherung dar, die zwischen zwei bekannten Punkten interpoliert wurde.
- „*Geometric Center*“: Das Ergebnis ist der Mittelpunkt eines Elements wie einer Linie (z.B. Straße) oder eines Polygons (z.B. Region)
- „*Approximate*“: Das Ergebnis stellt lediglich eine Näherung dar.

Die *MapQuest Geocoding API* ermöglicht einen einfachen Zugang zum Geokodierer mittels *HTTP*-Anfrage (Hypertext Transfer Protocol). Öffentliche Nutzer können pro Tag 5.000 Geokodierungsanfragen an den Dienst schicken. Bei der Adressgeokodierung können Anfragen in unstrukturierter (zusammengesetzte Adresse) oder strukturierter Form (Adresse zerlegt in einzelne Komponenten) erfolgen, wobei Ausgaben in den Formaten *JSON*, *XML* und *CSV* (Comma-separated Values) geliefert werden können. Neben der Ausgabe der Adresse sowie der Koordinaten, gibt es noch einen Parameter der die Geokodierungsqualität angibt. Dieser Code setzt sich aus fünf Zeichen zusammen. Die ersten beiden Zeichen repräsentieren den Addresslevel des Ergebnisses. Das dritte Zeichen gibt an, wie genau auf Straßenebene geokodiert wurde. Das vierte Zeichen enthält Informationen über die Genauigkeit auf Ebene von administrativen Einheiten und das letzte Zeichen liefert Angaben über die Genauigkeit auf Postleitzahlenebene. (vgl. MAPQUEST 2013)

Nominatim ist das Geokodierungswerkzeug von *OpenStreetMap* zur Umwandlung von Adressen in geographische Koordinaten. In diesem Kontext ist festzuhalten, dass im Zuge von Anfragen über *Nominatim* Suchbegriffe von links nach rechts abgearbeitet werden. Aus diesem Grund muss das Spezielle (bspw. der Straßename) am Anfang und das Allgemeine (bspw. das Land) am Ende einer Anfrage stehen. *Nominatim* unterstützt dabei die Ausgabeformate *HTML*, *JSON* und *XML*. Da die Ergebnisse auf *OpenStreetMap*-Daten basieren, unterliegen diese der *Creative Commons Attribution-ShareAlike Lizenz*. Weitere Nutzungsrichtlinien oder Limitierungen sind nicht bekannt, wobei eine zu intensive Nutzung zur Sperrung des Dienstes führen kann. (vgl. OPENSTREETMAP 2013)

Neben den genannten Geokodierungsdiensten gibt es noch einige weitere Dienste, die auf Grund verschiedener Faktoren nicht in die Liste möglicher Alternativen aufgenommen wurden. Einerseits ist die Abdeckung mancher Dienste nur auf bestimmte Gebiete beschränkt (z.B. *GeoCoder.us* für die USA, *OIORest* für Dänemark, *IGN OpenLS* für Frankreich), andererseits ist die Nutzung kostenpflichtig, wie etwa bei *Yahoo! Boss PlaceFinder* (vgl. YAHOO 2012).

In Tab. 2 werden abschließend alle Geokodierungsdienste gegenübergestellt. Im Folgenden werden diese hinsichtlich Anfragezahlen verwendete Standards und sonstige Limitierungen verglichen.

Dienst	Anfragen	Ausgabeformate	Sonstige Limitierungen
Bing Maps Locations	10.000 / Tag	JSON, XML	Ergebnisse nur in Verbindung mit Bing-Karten; kein Download der Ergebnisse
CloudMade Geocoding & Geosearch	100.000 / Monat	JSON, GeoJSON, Property List, HTML	
Google Maps Geocoding	2.500 / Tag	JSON, XML	Ergebnisse nur in Verbindung mit Google-Karten; kein Download der Ergebnisse
MapQuest Geocoding Service	5.000 / Tag	JSON, XML, CSV	
OpenStreetMap Nominatim	unbegrenzt (jedoch keine Massenanfragen)	JSON, XML, HTML	

Quelle: ILLMEYER, 2013

Tab. 2: Zusammenfassung ausgewählter Geokodierungsdienste

3.5 Evaluierung der Online-Geokodierungsdienste

Die Evaluierung der im vorherigen Kapitel beschriebenen Online-Geokodierungsdienste erfolgt auf Basis der Kriterien Genauigkeit, Schnelligkeit und Zuverlässigkeit. Die Schnelligkeit eines Dienstes kann ermittelt werden, indem ein identer Testdatensatz mit Hilfe einer eigens erstellten Webplattform abwechselnd mit einem der fünf Dienste geokodiert wird. Das Ergebnis spiegelt die Zeit wider, wie lange der jeweilige Dienst zur Abarbeitung des Datensatzes benötigt, wobei eine stabile und gleichbleibend schnelle Internetverbindung vorausgesetzt wird. Um etwaige Ausreiser zu vermeiden, wird der Test an mehreren Tagen wiederholt.

Die Genauigkeit eines Geokodierungsprozess wird im Rahmen dieser Evaluierung in zwei Unterbereiche gegliedert:

- Lagefehler: Da im Testdatensatz nicht nur die Adressen sondern auch deren geographischen Koordinaten vorliegen, ist eine Ermittlung des Lagefehlers möglich.
- Trefferraten: Es wird erhoben, wie viel Prozent des Testdatensatzes erfolgreich geokodiert werden konnten. (siehe Kapitel 3.3)

Die Zuverlässigkeit gliedert sich ebenfalls in zwei Unterbereiche:

- Fehlerhafte Datensätze: Es wird ermittelt, wie die Geokodierungsdienste auf Fehlern im Adressdatensatz reagieren. Eine Auflistung und Beschreibung der möglichen Fehlerquellen folgt in Kapitel 3.5.1.

- Limitierungen: Da viele Dienste die Anzahl der Anfragen für eine bestimmte Zeitperiode limitieren, wird auch dieser Faktor mit in die Bewertung einbezogen. (siehe Kapitel 3.4)

Im Zuge der Untersuchung werden die Kriterien gewichtet, wobei die Dienste mit Hilfe einer multikriteriellen Analyse, dem Analytischen Hierarchieprozess (AHP), evaluiert werden. Die Ergebnisse dieser Analyse werden im Anschluss mit Resultaten anderer Studien verglichen und gegebenenfalls um diese erweitert.

3.5.1 Fehler in Adressdatensätzen

Dieser Abschnitt der Arbeit zeigt einige typische Fehler bzw. Probleme auf, die in Adressdatensätzen häufig in Erscheinung treten. Zur Veranschaulichung der Probleme wurde eine bestimmte Adresse mehrfach verwendet, wobei jeweils bestimmte beschreibende Elemente weggelassen wurden.

Zeile	Straße	Nr.	PLZ	Ort	Land
1	Perfektastraße	81	1230	Wien	Österreich
2	Perfektastraße	81	1230	-	Österreich
3	Perfektastraße	-	1230	Wien	Österreich
4	Perfektastraße	81	1230	Wien	-
5	Perfektastraße	81	-	Wien	Österreich
6	Perfektastr.	81	1230	Wien	Österreich
7	Perfektastraße	81	-	-	Österreich
8	Perfektastraße	120	1230	Wien	Österreich
9	Perfektastraße	81	1230	Wien	Österreich

Quelle: ILLMEYER, 2013

Tab. 3: Auszug aus möglichen Adressdatensatz

Tab. 3 beinhaltet exemplarisch eine Auflistung von Adressanweisungen, die idealerweise von Geokodierungsdiensten einwandfrei interpretiert werden sollten. Im Folgenden werden die Unterschiede der einzelnen Adresszeilen erläutert.

- Zeile eins stellt einen optimalen Adresseintrag mit Straße, Hausnummer, Postleitzahl, Ort und Land dar.
- In Zeile zwei fehlt die Ortsangabe, wobei der Geokodierungsdienst die PLZ als Ortsangabe verwenden sollte.
- In Zeile drei fehlt die Hausnummer, wobei abzuklären ist, wie die Dienste bei fehlender Hausnummer die Geokodierung vornehmen.
- Beim vierten Eintrag fehlt lediglich die Angabe des Landes, wodurch es zu mehrdeutigen Ergebnissen kommen kann, insofern dieser Ort mit zugehöriger Adresse in einem anderen Land ebenfalls existiert.
- Bei Adresse fünf wurde die PLZ weggelassen. Dieses Problem sollte durch Zuhilfenahme der Ortsangabe gelöst werden.
- In Zeile sechs wurde der Straßenname mit der im deutschen Sprachraum üblichen Abkürzung „-str.“ versehen.

- In Zeile sieben fehlen sowohl die Ortsangabe als auch die PLZ. Dadurch können u.U. keine eindeutigen Ergebnisse rückgemeldet werden.
- Beim Straßennamen der achten Adresse wurde eine Hausnummer angegeben die allerdings in der Realität nicht existiert.
- Der letzte Eintrag beinhaltet einen Rechtschreibfehler im Straßennamen.

3.5.2 Analytischer Hierarchieprozess

Der Analytische Hierarchieprozess (AHP) ist eine einfach anzuwendende Methode zur Unterstützung von Entscheidungen, ein sogenanntes Decision Support System (DSS). Beim AHP handelt es sich um ein multikriterielles Verfahren durch das komplexe Entscheidungen strukturiert werden, um dadurch zu einer systematischen, optimalen und rational nachvollziehbaren Entscheidung zu gelangen. Im Hinblick auf ein Gesamtziel werden Kriterien formuliert, die zur Strukturierung des Entscheidungsproblems beitragen. Schlussendlich wird die zur Lösung des Problems am besten geeignete Alternative ausgewählt (vgl. MEIXNER und HAAS 2012: 167f). Der idealtypische Entscheidungsprozess gliedert sich in folgende Phasen. (vgl. ebd.: 77)

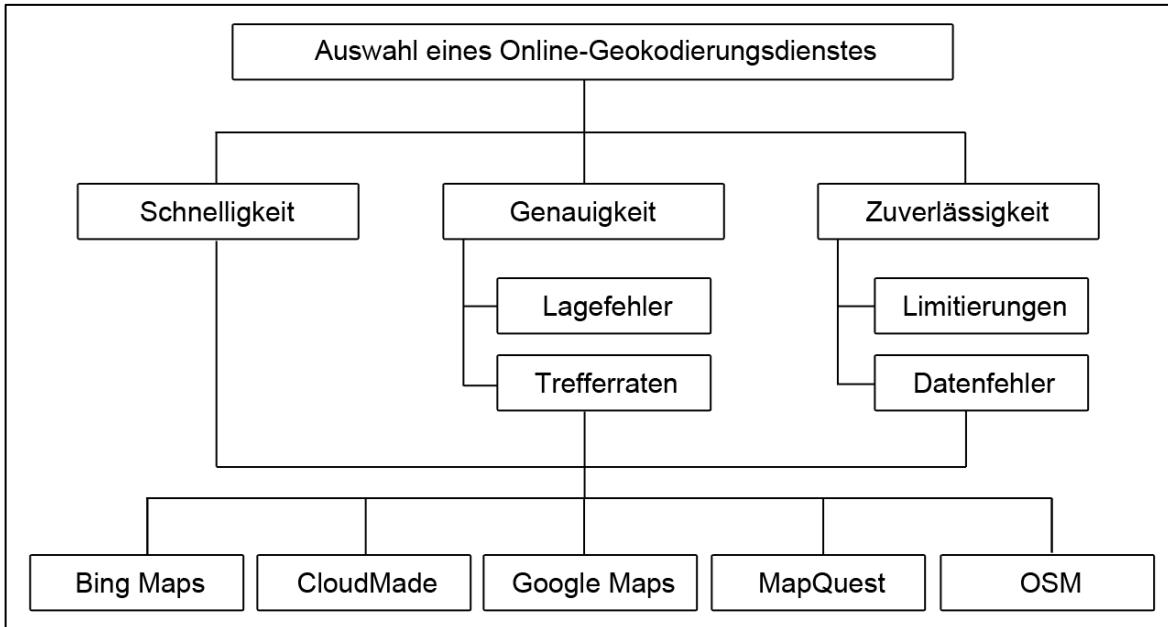
- *Präzise Formulierung des zu lösenden Problems*
- *Festlegung der Ziele*
- *Formulierung der Kriterien*
- *Suche nach Handlungsalternativen*
- *Bewertung der Handlungsalternativen*
- *Entscheiden (=Auswahl der optimalen Alternative)*
- *Umsetzung der Entscheidung*
- *Kontrolle und Feedback*

Der hier behandelte Entscheidungsprozess, der die Auswahl eines geeigneten Geokodierungsdienstes umfasst, unterscheidet sich geringfügig vom oben erwähnten idealen Ablauf. Am Anfang des Projekts steht die Suche bzw. die Definition frei verfügbarer Geokodierungsdienste.

In der Entscheidungstheorie wird die Objekt- von der Metaphase unterschieden. In letzterer werden unabhängig von der Methode oder dem angewandten Modell wichtige Vorrarbeiten geleistet, wie bspw. die Definition des Problems sowie der Kriterien bzw. die Suche nach Alternativen. Die Anwendung eines DSS bezieht sich auf die Objektphase und beinhaltet somit die nachfolgend aufgeführten Arbeitsschritte. (vgl. MEIXNER und HAAS 2012: 78)

1. Die Gewichtung der Kriterien.
2. Die Bewertung der Alternativen.
3. Die Ableitung der optimalen Lösung, basierend auf Urteilen und Einschätzungen, die in das Modell einfließen.
4. Die Beurteilung, ob die optimale Lösung akzeptiert wird.

Am Anfang des AHP wird die Ist-Situation sorgfältig analysiert. Daraus ergibt sich bei der Auswahl eines Online-Geokodierungsdienstes folgende Entscheidungshierarchie (siehe Abb. 9).



Quelle: ILLMEYER, 2013

Abb. 9: Entscheidungshierarchie AHP

Das Hauptziel stellt die Auswahl eines Online-Geokodierungsdienstes dar. Als alternativen können die fünf beschriebenen Geokodierungsdienste aufgezählt werden. Die dazwischenliegende Hierarchieebene umfasst die drei Hauptmerkmale (Schnelligkeit, Genauigkeit und Zuverlässigkeit), die zur Evaluierung der Szenarien benötigt werden.

Diese Merkmale werden zudem weiter unterteilt in maximal zwei Subkriterien, anhand derer die Szenarien schlussendlich bewertet werden. Damit der AHP angewendet werden kann ist es notwendig, bestimmte Voraussetzungen zu akzeptieren. Zwei Kriterien müssen miteinander paarweise verglichen werden. Außerdem muss das Entscheidungsproblem als Hierarchie formuliert werden können, wobei alle Kriterien und Alternativen, die Einfluss auf das Entscheidungsproblem haben können, in der Hierarchie enthalten sein müssen. Der AHP kann anhand eines vierstufigen Prozesses dargestellt werden. (vgl. MEIXNER und HAAS 2012: 183f)

- *Stufe 1: Aufstellung einer Entscheidungshierarchie mit Untergliederung in Kriterien.*
- *Stufe 2: Bewertung der Entscheidungshierarchie.*
- *Stufe 3: Prioritätenschätzung – Errechnung der Prioritäten der Entscheidungselemente und Überprüfung auf Konsistenz.*
- *Stufe 4: Aggregation der Prioritäten der Entscheidungselemente, um zu einer Rangreihung der Alternativen zu gelangen.*

Prinzipiell verläuft der AHP immer nach einem bestimmten Schema. Zuerst wird das Problem genau definiert und in einer hierarchischen Struktur modelliert (siehe Abb. 9).

Nach Erstellung der Hierarchie müssen die Hierarchieelemente einer Ebene paarweise verglichen werden. Aus diesen Paarvergleichen lassen sich die Partialgewichtungen für die Kriterien einer Ebene errechnen. Anschließend werden alle Partialgewichte einer Ebene zu Gesamtgewichtungen verdichtet. In weiterer Folge kann mittels einer Sensitivitätsanalyse überprüft werden, wie stabil die Prioritäten bei geänderter Merkmalsgewichtung sind. Abschließend werden die Alternativen in Hinsicht auf die Kriterien bewertet und ein Ranking der Alternativen errechnet. (vgl. ebd.: 195ff)

Um qualitative Informationen zu verarbeiten, ist es notwendig, eine geeignete Skala zu verwenden. Da bei der Gewichtung der Kriterien Paarvergleiche durchgeführt werden, muss es anhand der Skala möglich sein, den Paarvergleichen metrische Zahlenwerte zuzuweisen. Die AHP Skala verwendet dazu neun Abstufungen, deren Bedeutung in Tab. 4 näher erläutert wird. Ein umgekehrtes Verhältnis wird durch die Reziprokwerte ausgedrückt. Anhand dieser Skala können alle Paarvergleiche im AHP durchgeführt werden. (vgl. ebd.: 200f)

AHP-Werte	Beschreibung	AHP-Werte (reziprok)	Beschreibung
1	gleiche Bedeutung	1	gleiche Bedeutung
3	etwas größere Bedeutung	1/3	etwas geringere Bedeutung
5	erheblich größere Bedeutung	1/5	erheblich geringere Bedeutung
7	sehr viel größere Bedeutung	1/7	sehr viel geringere Bedeutung
9	absolut dominierend	1/9	absolut unterlegen
2, 4, 6, 8	Zwischenwerte	1/2, 1/4, 1/6, 1/8	Zwischenwerte

Quelle: bearbeitet nach MEIXNER und HAAS 2012: 202f

Tab. 4: AHP-Skala zur Bewertung qualitativer Merkmale

3.5.3 Kriterienbewertung

Aufgrund der bisherigen Ergebnisse und Erfahrungen mit Online-Geokodierungsdiensten sowie in Hinblick auf den zukünftigen Einsatz im Zusammenhang einer Webplattform zur Erstellung von Karten auf Basis regionalstatistischer Raster, werden die Merkmale im Folgenden bewertet. Dabei werden Paarvergleiche für die Kriterien (siehe Tab. 5) und Subkriterien durchgeführt und in Form einer Matrix dargestellt. Dabei ist zu erkennen, dass für die Umsetzung der geplanten Webapplikation die „Genauigkeit“ der Geokodierung am bedeutungsvollsten ist. Im Vergleich zum Kriterium der „Zuverlässigkeit“ ist auch der „Schnelligkeit“ eine höhere Priorität zuzuweisen.

Kriterien	Schnelligkeit	Genauigkeit	Zuverlässigkeit
Schnelligkeit	1	1/3	4
Genauigkeit	3	1	5
Zuverlässigkeit	1/4	1/5	1

Quelle: ILLMEYER, 2013

Tab. 5: Vergleichsmatrix der Kriterien Schnelligkeit, Genauigkeit und Zuverlässigkeit

Im Anschluss wird die Vergleichsmatrix für die Subkriterien des Merkmals „Genauigkeit“ gebildet (siehe Tab. 6). Den „Trefferraten“ ist im Vergleich zu „Lagefehlern“ eine etwas größere Bedeutung beizumessen, um ein vollständiges und korrektes Kartenbild generieren zu können.

Kriterien	Lagefehler	Trefferraten
Lagefehler	1	1/3
Trefferraten	3	1

Quelle: ILLMEYER, 2013

Tab. 6: Vergleichsmatrix der Subkriterien für Genauigkeit

Um Vollständigkeit gewährleisten zu können, werden auch die Subkriterien „Limitierungen“ und „Datenfehler“ miteinander verglichen (siehe Tab. 7). Da Fehler in den Eingangsdaten durch genaue Kontrolle verhindert werden können, ist das Kriterium der „Limitierungen“ von Online-Geokodierungsdiensten als erheblich wichtiger einzuschätzen.

Kriterien	Limitierungen	Datenfehler
Limitierungen	1	4
Datenfehler	1/4	1

Quelle: ILLMEYER, 2013

Tab. 7: Vergleichsmatrix der Subkriterien für Zuverlässigkeit

Anschließend werden die Matrizen normalisiert, damit für alle Kriterien vergleichbare Werte erhalten werden. Das bedeutet, dass diese auf eine vergleichbare Basis gebracht werden. Dieser Prozess wird mit den Hauptkriterien sowie den jeweiligen Subkriterien durchgeführt, sodass vorerst lokale Prioritäten für jede Hierarchieebene entstehen. Die Normalisierung erfolgt dabei exemplarisch für den Vergleich von Genauigkeit und Schnelligkeit mit der Rechnung $3 / (1+3+0,25) = 0,71$. Für die Berechnung der lokalen Gewichtung wird zunächst die Summe der Zeile des Kriteriums gebildet: $0,71+0,65+0,50 = 1,86$. Abschließend wird die Summe durch die Anzahl der Kriterien dividiert: $1,86 / 3 = 0,62$ (siehe Tab. 8).

Kriterien	Schnelligkeit	Genauigkeit	Zuverlässigkeit	Summe	Lokale Gewichtung
Schnelligkeit	0,24	0,22	0,40	0,85	0,28
Genauigkeit	0,71	0,65	0,50	1,86	0,62
Zuverlässigkeit	0,06	0,13	0,10	0,29	0,10

Quelle: ILLMEYER, 2013

Tab. 8: Lokale Gewichtung der Hauptkriterien

Dieser Schritt wird analog mit den Subkriterien durchgeführt, wodurch hierbei die lokale Gewichtung der Genauigkeit für Lagefehler 0,25 und für die Trefferrate 0,75 beträgt. Des Weiteren betragen die lokalen Gewichtungen für Limitierungen 0,8 und für den Umgang mit Datenfehlern 0,2.

Zur Bestimmung der globalen Gewichtungen müssen die lokalen Prioritäten aggregiert werden. Dies geschieht mittels Multiplikation der Gewichtung der Subkriterien (falls vorhanden) mit den dazugehörigen Gewichtungen der Hauptkriterien. Dadurch werden

Zahlenwerte für die globale Gewichtung aller im Entscheidungsprozess vorkommenden Kriterien ermittelt (siehe Tab. 9).

Kriterien	Lokale Gewichtung 1. Ebene	Lokale Gewichtung 2. Ebene	Globale Gewichtung
Schnelligkeit	0,28		0,28
Genauigkeit	0,62		
Lagefehler		0,25	0,155
Trefferraten		0,75	0,465
Zuverlässigkeit	0,10		
Limitierungen		0,8	0,08
Datenfehler		0,2	0,02

Quelle: ILLMEYER, 2013

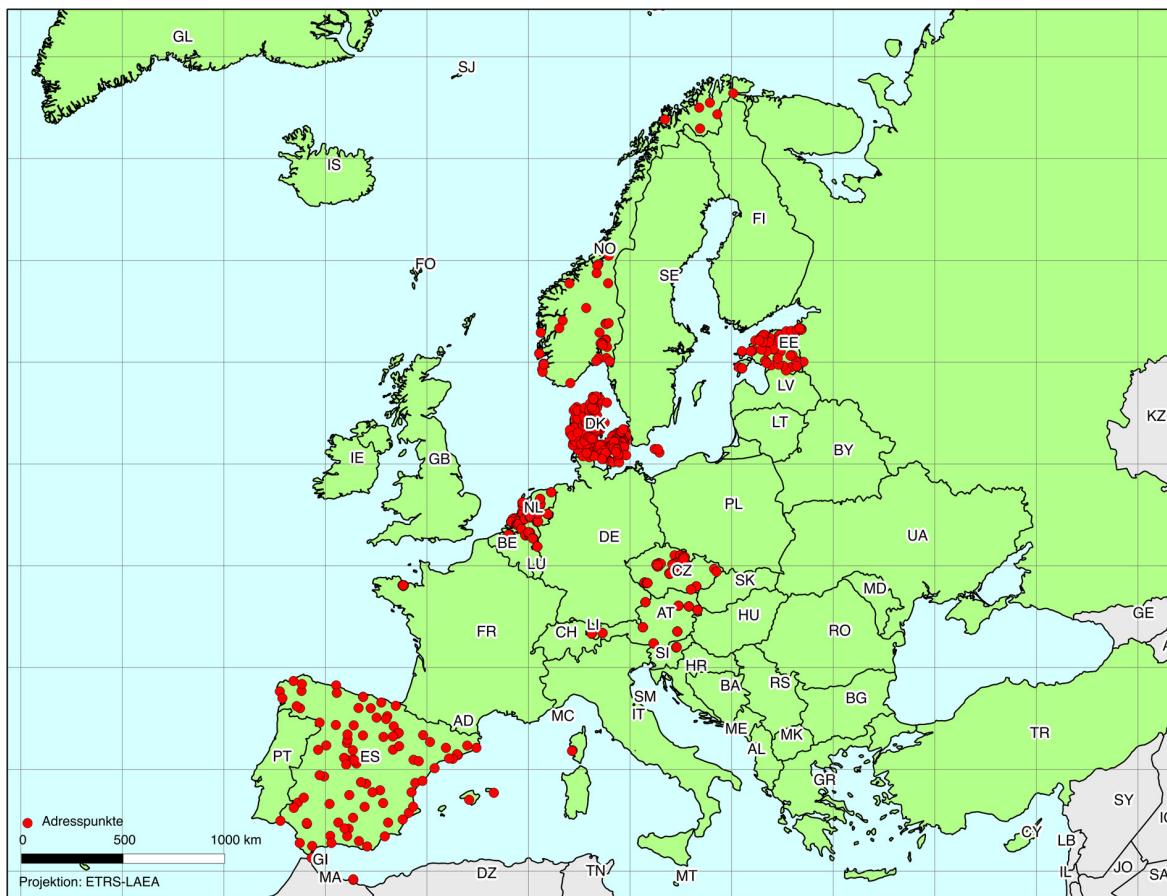
Tab. 9: Globale Gewichtung der Kriterien

3.5.4 Analyse und Ergebnisse

Nach der Prioritätenberechnung werden die Alternativen in Hinblick auf die Kriterien bewertet. Die Ergebnisse dieser Evaluierung werden anschließend mit der Gewichtung der Kriterien angepasst. In diesem Zusammenhang spiegelt das Ergebnis den Geokodierungsdienst wider, der aufgrund der Prioritäten der Kriterien sowie auf Basis der Bewertung als am geeigneten erachtet wird.

Die Kriterien „Schnelligkeit“, „Lagefehler“ und „Trefferraten“ können quantitativ bewertet werden. Die Kriterien „Limitierungen“ und „Datenfehler“ erfordern hingegen eine qualitative Bewertung.

Der Adressdatensatz, der zur Evaluierung der Online-Geokodierungsdienste verwendet wird, besteht aus insgesamt 1.467 Adressen, die sich über neun europäische Länder erstrecken. Dies stellt die Anzahl der Adressen dar, die dem Autor dieser Arbeit für die notwendigen Forschungszwecke zur Verfügung gestellt wurden. Zur Akquirierung der Daten für den Testdatensatz wurden nationale Statistik- sowie amtliche Landesvermessungsinstitute kontaktiert. Die Adressen wurden im Sinne der Forschung teilweise kostenlos zur Verfügung gestellt. Eine Abdeckung des gesamteuropäischen Raums war aufgrund von bestimmten Datenschutzbestimmungen nicht möglich (siehe Abb. 10).



Quelle: ILLMEYER, 2013

Abb. 10: Übersicht der Adresspunkte im Testdatensatz

Um ausgewogene und aussagekräftige Geokodierungsergebnisse zu gewährleisten, wurde eine homogene Verteilung der Adressen auf die jeweilige Staatsfläche angestrebt. Dieses Vorhaben war allerdings angesichts verschiedener nationaler Vergabesysteme nicht immer möglich. Darüber hinaus war es im Zuge der Ergebnisevaluierung wichtig, dass die Adressen die zugehörigen geokodierten Koordinaten enthalten, um Lagefehler in den Ergebnissen erkennen zu können. Da die Adressen in den einzelnen Ländern mit verschiedenen Koordinatensystemen geokodiert wurden, mussten, um einen homogenen europäischen Datensatz zu erhalten, die Koordinaten in ein einheitliches System transferiert werden. In Rahmen dieser Arbeit wurden diesbezüglich geographische Koordinaten auf Basis des geodätischen Referenzsystems WGS84, basierend auf dem GRS80 verwendet. Zur besseren Veranschaulichung bietet Tab. 10 einen Auszug aus dem verwendeten Adressdatensatz.

Straße	Nr.	PLZ	Stadt	Land	Geogr. Länge	Geogr. Breite
Rue Alexander Fleming	15	22000	Saint Brieuc	France	-2,80076	48,51770
Calle Cardencha	48	28223	Madrid	Spain	-3,77996	40,39685
Liivaluite	7		Tallinn	Estonia	24,72892	59,39335
Snorresvej	3	3600	Frederikssund	Denmark	12,04907	55,85089
Horovskeho	146	16300	Praha	Czech Republic	14,31472	50,07314
Ghegagasse	29	8020	Graz	Austria	15,41919	47,07407
Kovaska cesta	4	3205	Vitanje	Slovenia	15,29930	46,37860
Hamrasveien	11	4016	Stavanger	Norway	5,73549	58,94200
Hagedoornstraat	26		Rotterdam	Netherlands	4,47368	51,94320

Quelle: ILLMEYER, 2013

Tab. 10: Auszug aus Test-Adressdatensatz

Tab. 11 liefert einen Überblick über datenbereitstellende Länder, inkl. dazugehöriger Quellen sowie Anzahl der Adressen.

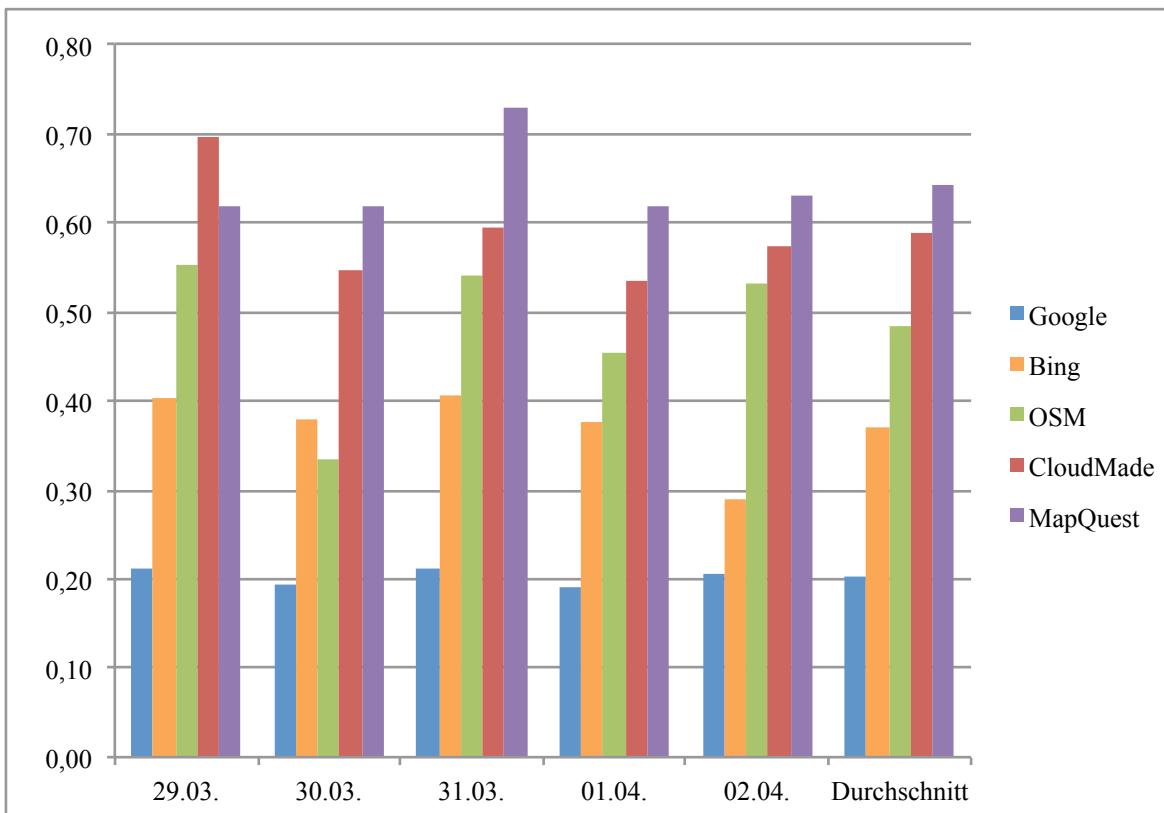
Land	Anzahl Adressen	Quelle
Dänemark	231	LIND: 2013
Estland	165	VALGMA: 2013
Frankreich	40	LIPATZ: 2013
Niederlande	273	VAN LEEUWEN: 2013
Norwegen	46	KARTVERKET: 2013
Österreich	543	BEV: 2013
Slowenien	17	REJC: 2013
Spanien	43	CARTOCIUDAD: 2013
Tschechische Republik	109	ŠMÍDOVÁ: 2013

Quelle: ILLMEYER, 2013

Tab. 11: Adressdaten in Europa

Die „Schnelligkeit“ von Online-Geokodierungsdiensten wird ermittelt, in dem ein Testdatensatz mit Adressen in Europa, mit Hilfe des in dieser Arbeit erstellten Webdienstes, jeweils durch einen der fünf Geokodierungsdienste abgearbeitet wird. Die Webapplikation basiert auf *PHP*, wobei jeweils eine Adresse an den Dienst gesendet und auf das Ergebnis im *JSON* Format gewartet wird. Während eines Geokodierungsvorgangs wird die gesamte Zeit, die der Prozess benötigt hat, gespeichert und diese dann durch die Anzahl der Adressen dividiert. Die Durchgänge wurden jeweils an fünf aufeinanderfolgenden Tagen (29.03.2013 – 02.04.2013) zu unterschiedlichen Tageszeiten durchgeführt, wobei auf Basis dessen ein Durchschnittswert berechnet wurde. Dabei war zu beachten, dass jeweils gleiche Bedingungen für die einzelnen Durchgänge vorherrschen. Das bedeutet, dass die Netzwerkverbindung gleich belastet wird und keine weiteren Prozesse auf dem Testrechner ausgeführt werden. Abb. 11 stellt in diesem Zusammenhang die Zeit dar, die an den jeweiligen Tagen bzw. im Durchschnitt die Geokodierungsdienste für die Bearbeitung eines Adresseintrages benötigt haben. Mit einer mittleren Geokodierungszeit von 0,2 Sekunden liegt der Dienst von *Google* bei dieser Evaluierung vor *Bing* (0,37 s) an erster Stelle. Beim *Google* Service konnten zudem die konstantesten Ergebnisse über alle fünf Untersuchungstage hinweg beobachtet werden. Die größten Schwankungen (0,33 s - 0,55 s) hinsichtlich der Geokodierungszeiten wies der *OpenStreetMap*-Dienst auf, der im

Durchschnitt mit 0,48 s im Mittelfeld der Bewertung liegt. Die beiden langsamsten Dienste sind *CloudMade* (0,59 s) und *MapQuest* (0,64 s).



Quelle: ILLMEYER, 2013

Abb. 11: Geokodierungsschnelligkeit in Sekunden pro Anfrage

Der „Lagefehler“ eines einzelnen Adresseintrages wird durch die Entfernung der bekannten Koordinaten aus den Quelldaten zu den resultierenden Koordinaten eines Geokodierungsvorganges berechnet. Hierzu wird die *Haversine-Formel* (vgl. MOVABLE TYPE 2013a) verwendet. Haversine beschreibt genau genommen die Hälfte des umgedrehten Sinus eines gegebenen Winkels (half versed sine). Es handelt sich dabei um eine gängige Methoden zur Berechnung von Distanzen auf der Erdoberfläche auf Basis von zwei koordinativ bekannten Punkten. Die kürzeste Luftliniendistanz (Orthodrome) wird hierbei entlang eines Großkreises berechnet. SINNOTT (1984) erläuterte erstmals die Berechnung von Distanzen auf der Erdkugel mit Hilfe der *Haversine-Formel*, auf Basis der Kenntnis, dass die zuvor üblich verwendeten Berechnungsmethoden der sphärischen Trigonometrie (Sphärisches Kosinus-Gesetz) dann ungenaue Ergebnisse liefern, wenn die zwei Punkte nahe beieinander liegen. Da die *Haversine-Methode* auf der Annahme basiert, dass die Erde eine Kugel mit einem fixen Radius sei, treten allerdings auch hier minimale Fehler auf. Im Allgemeinen liegt die Fehlerrate unter 0,3 %, in Abhängigkeit der geographischen Breite sowie der Messrichtung. Im Rahmen dieser Arbeit ist eine derartige Genauigkeit jedoch ausreichend, da in den meisten Fällen ein Fehler von nur wenigen Dezimetern erwartet wird (vgl. MOVABLE TYPE 2013a). Exaktere Ergebnisse mit geodätischen Genauigkeiten innerhalb von 0,5 mm würde die *Vincenty-Formel* liefern, die jedoch weitaus komplexer ist und dadurch sowohl mehr Rechenaufwand als auch

Rechenzeit benötigt (vgl. MOVABLE TYPE 2013b). Im Folgenden wird die *Haversine-Formel* präsentiert sowie der dazugehörige Rechnungsweg erläutert.

$$\text{haversin}\left(\frac{d}{R}\right) = \text{haversin}(\varphi_2 - \varphi_1) + \cos(\varphi_1) * \cos(\varphi_2) * \text{haversin}(\lambda_2 - \lambda_1)$$

wobei: $\text{haversin}(x) = \sin^2\left(\frac{x}{2}\right)$

Bezogen auf die Formel steht d stellvertretend für die Distanz, wobei R der Erdradius ist (im Mittel 6371km). Ferner bezeichnet φ die geographische Breite und λ die geographische Länge. Durch Auflösung der *Haversine-Formel* ergibt sich folgender Rechnungsweg. (vgl. MOVABLE TYPE 2013a)

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1) * \cos(\varphi_2) * \sin^2\left(\frac{\Delta\lambda}{2}\right)$$

$$c = 2 * \arctan2\left(\sqrt{a}, \sqrt{1-a}\right)$$

$$d = R * c$$

Der Abstand zwischen den resultierenden Koordinaten eines Geokodierungsdurchgangs und den Koordinaten aus den Primärdaten kann somit mit Hilfe der *Haversine-Formel* berechnet werden. Da die Ergebnisse der Geokodierung in Form einer MS Excel Tabelle vorliegen, wird auch die Distanz in Excel berechnet. Die umgewandelte Formel sieht innerhalb von MS Excel folgendermaßen aus:

$$\begin{aligned} & (\text{ARCCOS}(\text{SIN}(\text{BOGENMASS}(\text{lat1})) * \text{SIN}(\text{BOGENMASS}(\text{lat2}))) + \\ & \cos(\text{BOGENMASS}(\text{lat1})) * \cos(\text{BOGENMASS}(\text{lat2})) * \cos(\text{BOGENMASS}(\text{lon2}) \\ & - \text{BOGENMASS}(\text{lon1}))) * R * 1000 \end{aligned}$$

Mit Hilfe dieser Formel wird der Abstand der beiden Adresspunkte in Metern ausgegeben. Dabei müssen die geographische Länge und Breite zuvor in Radiant umgerechnet werden. Tab. 12 stellt in diesem Kontext die statistischen Parameter „Mittelwert“, „Minimum“, „Maximum“ und einige „Quantile“ bezogen auf die fünf Geokodierungsdienste dar.

Parameter	Bing	CloudMade	Google	MapQuest	OSM
Min	0,0 m	3,4 m	0,0 m	3,8 m	3,9 m
Max	156.298,0 m	14.599.893,1 m	101.836,8 m	226.044,6 m	225.857,6 m
Mittelwert	374,6 m	120.676,9 m	173,9 m	2.708,1 m	3.077,5 m
Quantile					
50%	1,3 m	229,3 m	5,2 m	196,4 m	185,0 m
75%	19,2 m	977,1 m	21,1 m	555,1 m	515,7 m
90%	190,2 m	9.810,1 m	95,4 m	1.197,1 m	1.176,2 m
95%	1.101,0 m	96.833,9 m	298,9 m	4.346,5 m	9.810,1 m

Quelle: ILLMEYER, 2013

Tab. 12: Statistische Distanz-Parameter der originalen Geokodierungsergebnisse

Auf Grund bestimmter Ausreißer können allerdings keine gehaltvollen Aussagen über statistische Parameter wie etwa dem Mittelwert gemacht werden. Die Quantile sind in

diesem Fall besser geeignet, um Auskunft über die Verteilung der Daten zu geben, da diese Maßzahlen unabhängig von Ausreißern sind. Quantile geben an, wieviel Prozent der Daten unter dem genannten Wert liegen. So weisen z.B. beim 75 %-Quantil des *Bing*-Dienstes drei Viertel der Daten eine Distanz von unter 19,2 m auf. Das 50 %-Quantil spiegelt den Median, den mittleren Wert einer Verteilung, wider. Die Gegenüberstellung zeigt, dass der *Bing*-Dienst für die Hälfte der Daten bessere Ergebnisse liefert als der *Google*-Dienst. Bei Vergleich der 95 %-Quantile von *Google* (<300 m) und *Bing* (>1000 m) wird augenscheinlich, dass *Google*, bezogen auf den gesamten Datensatz vergleichsweise genaue Ergebnisse liefert.

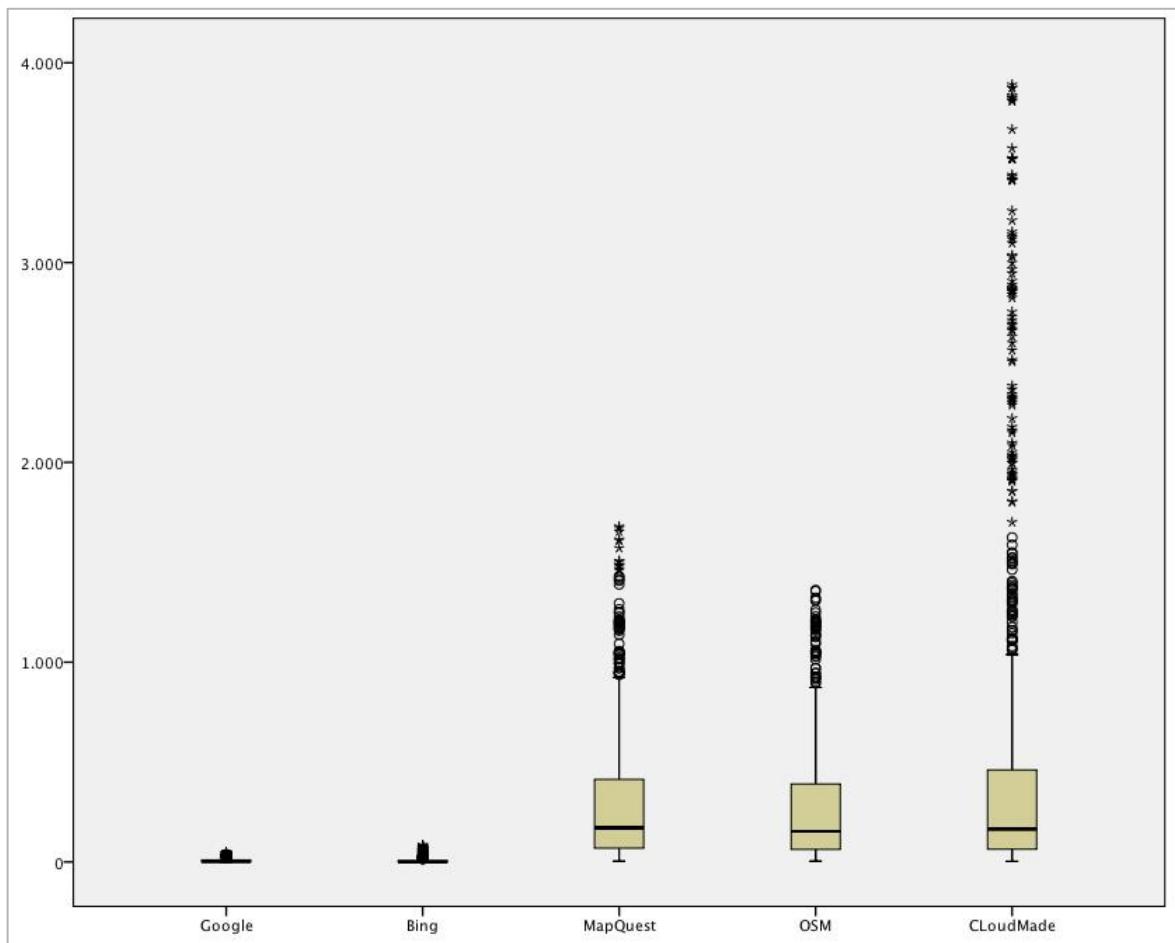
Zur Lösung des vorliegenden Problems der statistischen Parameter wurden die Ausreißer entfernt, indem die Ergebnisse nach oben hin um 15 % gekürzt wurden. Mit Hilfe des getrimmten Ergebnisses konnten genaue statistische Analysen der Daten durchgeführt werden. Tab. 13 zeigt in diesem Zusammenhang die statistischen Parameter „Maximum“, „Mittelwert“, „Standardabweichung“ sowie „Median“ der getrimmten Ergebnisse.

Parameter	Bing	CloudMade	Google	MapQuest	OSM
Max	84,4 m	3.891,0 m	46,5 m	1.679,1 m	1.360,7 m
Mittelwert	7,7 m	457,9 m	7,7 m	295,4 m	269,6 m
Median	1,1 m	164,4 m	3,2 m	171,1 m	457,9 m

Quelle: ILLMEYER, 2013

Tab. 13: Statistische Distanz-Parameter der getrimmten Geokodierungsergebnisse

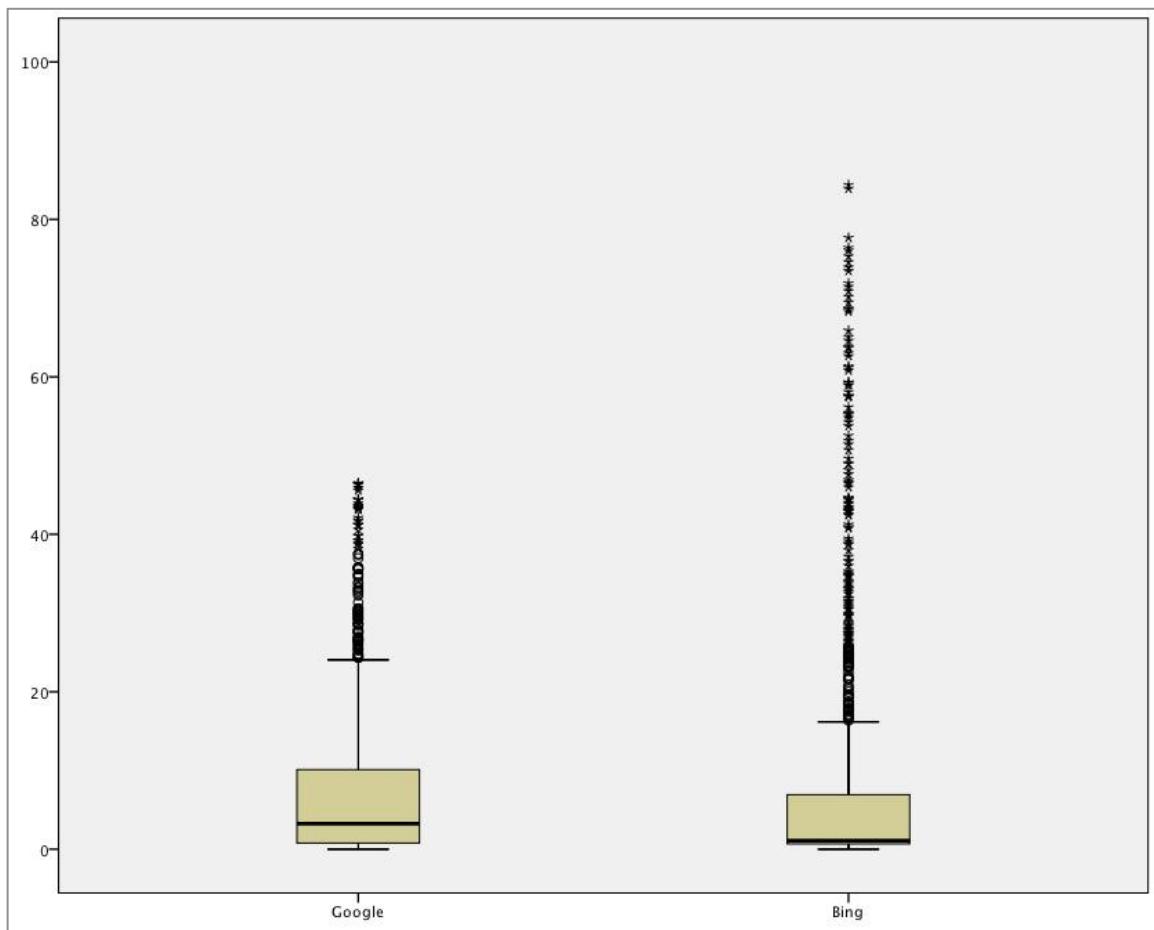
Um die Verteilung der Daten noch genauer interpretieren zu können sowie zur übersichtlichen Präsentation der Resultate, wurden *Boxplot-Diagramme* für alle Distanzen der Geokodierungsdienste gegenübergestellt (siehe Abb. 12). Ein Boxplot umfasst die graphische Darstellung des Medians, die zwei Quartile (25%- und 75%-Quartile), Extremwerte sowie eventuelle Ausreißer.



Quelle: ILLMEYER, 2013

Abb. 12: Boxplot-Diagramme für Distanzen aller Geokodierungsergebnisse

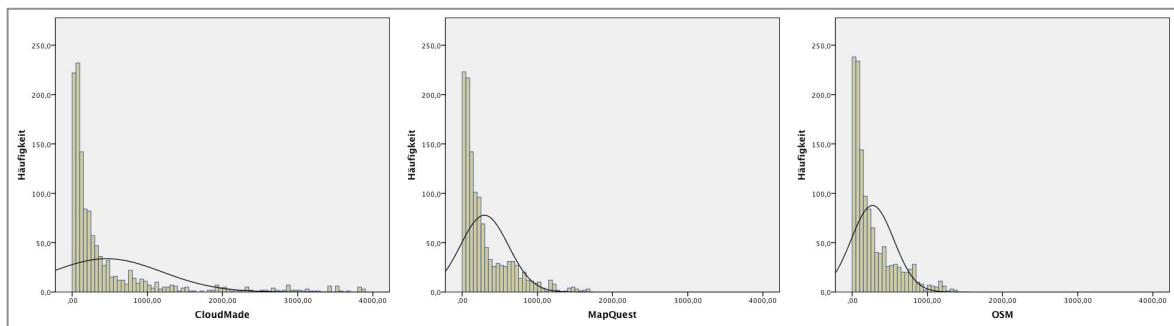
Um die Boxplot-Diagramme von *Bing* und *Google* genauer erfassen zu können, wurde eine weitere Gegenüberstellung auf Basis eines neuen Skalenniveaus erstellt (siehe Abb. 13).



Quelle: ILLMEYER, 2013

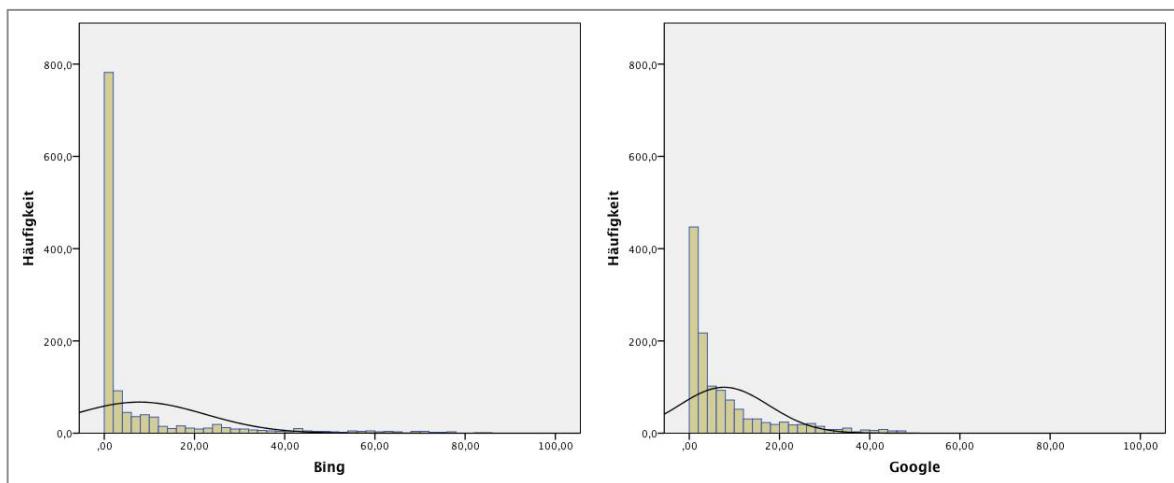
Abb. 13: Boxplot-Diagramme für Distanzen der Geokodierungsergebnisse von Bing und Google

Wie auf den Abbildungen zu erkennen ist, können die Dienste von *OpenStreetMap*, *CloudMade* und *MapQuest* gut miteinander verglichen werden, wobei *CloudMade* die größte Streuung bzw. die meisten Ausreißer aufweist. Im Vergleich zwischen *Bing* und *Google* weist ersterer Dienst sowohl niedrigere Quartile als auch einen kleineren Median auf. Der Dienst von *Google* streut hingegen mehr. Um diese Ergebnisse zu bestätigen bzw. die Verteilung noch besser interpretieren zu können, wurden Histogramme der Distanzen für die fünf Geokodierungsdienste erstellt. Hierbei wurden ebenfalls Vergleiche auf zwei verschiedenen Skalenniveaus zwischen *CloudMade*, *MapQuest* und *OpenStreetMap* (siehe Abb. 14) bzw. zwischen *Bing* und *Google* (siehe Abb. 15) durchgeführt.



Quelle: ILLMEYER, 2013

Abb. 14: Histogramm der Distanzen für CloudMade, MapQuest und OpenStreetMap

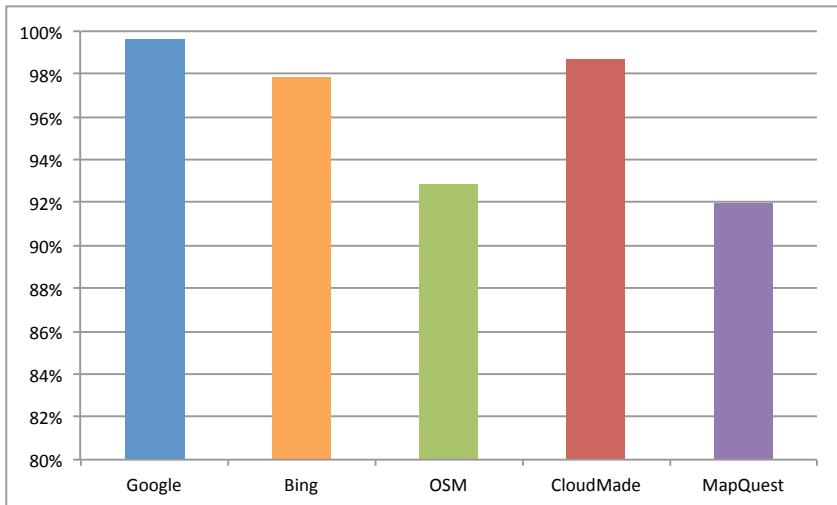


Quelle: ILLMEYER, 2013

Abb. 15: Histogramm der Distanzen für Bing und Google

Anhand der Histogramme kann bestätigt werden, dass *Google* bessere Ergebnisse hinsichtlich der Entfernung der geokodierten Ergebnisse über den gesamten Adressdatensatz liefert. Des Weiteren kann aufgrund der flachen Kurve von *CloudMade* entnommen werden, dass dieser Dienst die meisten Ausreißer und somit i.d.R. schlechtere Geokodierungsergebnisse erzielt.

Die Trefferraten der jeweiligen Geokodierungsdienste werden bestimmt, indem ermittelt wird, wie viele der Geokodierungsanfragen ein Ergebnis zurück liefern. Da die Lagegenauigkeit mit Hilfe eines anderen Parameters (siehe oben) bestimmt wird, ist es für diesen Parameter bedeutungslos, ob das Geokodierungsergebnis einigermaßen lagerichtig ist oder nicht. Es wird lediglich überprüft, ob ein korrektes Koordinatenpaar vom Dienst übermittelt wird. Da die Analysen an fünf aufeinanderfolgenden Tagen durchgeführt wurden, ist anzunehmen, dass die Dienste immer gleiche Trefferraten erzielen, da sich die Testdaten und der Geokodierungsprozess unverändert blieben. Allerdings wurden dennoch voneinander abweichende Trefferraten beobachtet, die sich jedoch lediglich um maximal ein Prozent unterscheiden. Abb. 16 stellt die durchschnittlichen Trefferraten der fünf Geokodierungsdienste dar. *Google* führt das Ranking mit einer Treffrate von 99,6 % an, gefolgt von *CloudMade* (98,7 %), *Bing* (97,8 %), *OpenStreetMap* (92,8 %) sowie an letzter Stelle *MapQuest* mit durchschnittlich 92,0 %.



Quelle: ILLMEYER, 2013

Abb. 16: Trefferraten aller Geokodierungsdienste

Das Kriterium der Limitierung wird auf Basis der vorhandenen Information über die jeweiligen Dienste bewertet. Die Skala umfasst die Bewertungen „wenig limitiert (9)“ bis „stark limitiert (1)“. Der *OpenStreetMap* Geokodierungsdienst kann bspw. auf Grund der offenen Lizenzierungsrichtlinien als gut bewertet werden (8). Im Gegenzug dazu erhält der *Google Maps* Geokodierungsdienst aufgrund der geringen Zugriffszahlen pro Tag und der Tatsache, dass die Ergebnisse nur in Verbindung mit einer *Google* Karte verwendet werden dürfen, die schlechteste Benotung (3) der untersuchten Services (siehe Tab. 15).

Der letzte Parameter zur Evaluierung der Online-Geokodierungsdienste befasst sich mit dem Umgang der Services hinsichtlich fehlerhafter Primärdaten. Da diese Fehler am ehesten vom Benutzer vermieden werden können, spielen sie keine allzu große Rolle bei der Auswahl eines Dienstes. Um die Vollständigkeit der Evaluierung zu gewährleisten, werden diese dennoch genauer behandelt sowie deren Ergebnisse in den Entscheidungsprozess miteinbezogen. Potentielle Fehler, die ein Adressdatensatz aufweisen kann, wurden bereits in Kapitel 3.5.1 aufgezeigt. Die Evaluierung auf Basis der Datenfehler wurde mit dem in Kapitel 3.5.1 verwendeten Datensatz durchgeführt. Zur Überprüfung der Korrektheit der Ergebnisse, muss zuerst festgelegt werden, was von den Diensten in welchen Fällen zu erwarten ist. Mit Ausnahme einer falschen Hausnummer, sollten alle Dienste ein Ergebnis liefern, dass jenem eines vollständigen Adresseintrages übereinstimmt. Im Falle einer fehlenden Hausnummer sollte idealerweise der Straßenmittelpunkt geokodiert werden, sowie bei einer falschen Hausnummer das nächstliegende Ergebnis. Der *Google*-Geokodierungsdienst liefert, für alle Adresseinträge das bestmögliche Ergebnis. Die weiteren Ergebnisse der Untersuchung sind in Tab. 14 ersichtlich. Die Bewertung erfolgt nach Punkten, wie viele der acht Datenfehler korrekt behandelt wurden.

Datenfehler	Bing	CloudMade	Google	MapQuest	OSM
Kein Ort	✓	✓	✓	✓	✓
Keine Hausnummer	Straßenende	✓	✓	✓	✓
Kein Land	✓	✓	✓	✓	✓
Keine PLZ	✓	✓	✓	✓	✓
Abkürzung	✓	Ortsmitte	✓	✓	✓
Keine PLZ & Ort	Kein Ergebnis	✓	✓	✓	✓
Falsche Hausnummer	✓	Straßenmitte	✓	Straßenmitte	Straßenmitte
Schreibfehler	✓	Ortsmitte	✓	Kein Ergebnis	Kein Ergebnis

Quelle: ILLMEYER, 2013

Tab. 14: Umgang der Geokodierungsdienste mit Fehlern in Quelldaten

Auf Grund der vorliegenden Ergebnisse und Analysen kann die Ergebnismatrix der Bewertungen gebildet werden (siehe Tab. 15). Für den Parameter der „Schnelligkeit“ wird die durchschnittliche Geokodierungszeit pro Adresse angegeben. Für die Bewertung des „Lagefehlers“ wird das getrimmte Mittel der Distanzen herangezogen. Die „Trefferraten“ werden in Prozent angegeben. Hinsichtlich der Datenfehler wird die Anzahl dargestellt, wie viele der Fehler korrekt behandelt wurden, wobei höhere Werte besseren Ergebnissen entsprechen.

	Bing	CloudMade	Google	MapQuest	OSM
Schnelligkeit	0,37 s	0,59 s	0,20 s	0,64 s	0,48 s
Genauigkeit					
Lagefehler	7,71 m	457,87 m	7,69 m	295,44 m	269,62 m
Trefferraten	97,83%	98,72%	99,61%	91,96%	92,84%
Zuverlässigkeit					
Limitierungen	6	5	3	6	8
Datenfehler	6	5	8	6	6

Quelle: ILLMEYER, 2013

Tab. 15: Ergebnismatrix der Bewertung

Im nächster Schritt erfolgt eine Normalisierung der Bewertung (siehe Tab. 16). Für Parameter, bei denen hohe Werte negativ zu beurteilen sind („Schnelligkeit“ und „Lagefehler“), werden die Kehrwerte normalisiert. Als Beispiel wird hierbei die Geokodierungsschnelligkeit von *Bing* mit deren Kehrwerten normalisiert.

$$\frac{\frac{1}{0,37}}{\frac{1}{0,37} + \frac{1}{0,59} + \frac{1}{0,20} + \frac{1}{0,64} + \frac{1}{0,48}} = 0,21$$

	Bing	CloudMade	Google	MapQuest	OSM	Globale Gewichtung
Schnelligkeit	0,21	0,13	0,38	0,12	0,16	0,28
Genauigkeit						
Lagefehler	0,48	0,01	0,48	0,01	0,01	0,155
Trefferraten	0,20	0,21	0,21	0,19	0,19	0,465
Zuverlässigkeit						
Limitierungen	0,21	0,18	0,11	0,21	0,29	0,08
Datenfehler	0,19	0,16	0,26	0,19	0,19	0,02

Quelle: ILLMEYER, 2013

Tab. 16: Normalisierte Ergebnisse der Alternativenbewertung

Abschließend werden die normalisierten Ergebnisse mit den globalen Gewichten multipliziert. Die gewichteten Werte können im Anschluss für die jeweilige Alternative aufsummiert werden, um dadurch die Ergebnisse des AHP zu erhalten (siehe Tab. 17).

	Bing	CloudMade	Google	MapQuest	OSM
Schnelligkeit	0,06	0,04	0,11	0,03	0,05
Genauigkeit					
Lagefehler	0,07	0,00	0,07	0,00	0,00
Trefferraten	0,09	0,10	0,10	0,09	0,09
Zuverlässigkeit					
Limitierungen	0,02	0,01	0,01	0,02	0,02
Datenfehler	0,00	0,00	0,00	0,00	0,00
Summen	0,25	0,15	0,29	0,15	0,16

Quelle: ILLMEYER, 2013

Tab. 17: Gewichtete Ergebnisse der Alternativenbewertung

Das Beispiel demonstriert, dass für die Bewertung von Alternativen sowohl quantitative als auch qualitative Daten einfließen können. Die erzielten Ergebnisse sowie die darauf gegründete Auswahl eines Geokodierungsdienstes werden in Kapitel 3.5.6 thematisiert.

3.5.5 Vergleich mit anderen Studien

Zur Überprüfung der ermittelten Ergebnisse des Evaluierungsprozesses bzw. zur Bestätigung deren Aussagekraft, werden in weiterer Folge Ergebnisse gleichartiger Studien diskutiert. Aufgrund der Tatsache, dass sich diese Studien teilweise auf eine andere Auswahl an Geokodierungsdiensten beziehen bzw. die Qualität der Datengrundlage unterschiedlich ist, können die Aussagen nicht unverändert auf die hier vorgestellte Studie umgelegt werden.

ROONGPIOONSOPIT und KARIMI (2010) führten eine Studie für das Gebiet der USA durch. Sie verglichen dabei die Dienste *Geocoder.us*, *Google*, *Bing* (früher *MapPoint*), *MapQuest* und *Yahoo* hinsichtlich „Trefferraten“ und „Lagefehler“. Da *Geocoder.us* ausschließlich das Gebiet der USA abdeckt und *Yahoo* keinen kostenlosen Geokodierungsdienst mehr anbietet, werden diese Ergebnisse in diesem Rahmen nicht berücksichtigt. Die Trefferraten dieser Studie für die Dienste *Google*, *Bing* und *MapQuest* betrugen jeweils 96,56 %, 92,39 % und 100 %. Für weitere Analysezwecke wurden die Ergebnisse in die drei Kategorien „gut“, „mehrdeutig“ und „keine Übereinstimmung“ klassifiziert. Diese Einordnung erfolgte auf Basis der von den Diensten zurückgelieferten Qualitätscores. Gute Ergebnisse beinhalten geokodierte Punkte auf Straßengenauigkeit. Mehrdeutige Ergebnisse hingegen repräsentieren Koordinaten mit der Genauigkeit von Städten, Postleitzahlen oder Staaten. Tab. 18 zeigt, dass *Google* und *Bing* qualitativ bessere Geokodierungsergebnisse liefern, da diese mehr Ergebnisse mit guten Trefferraten zurückgeben.

Übereinstimmungsqualität	Google	Bing	MapQuest
Gut	88,04%	86,05%	69,20%
Mehrdeutig	8,52%	6,34%	30,80%
Keine Übereinstimmung	3,44%	7,61%	-

Quelle: nach ROONGPIBOONSOPIT und KARIMI (2010): 1088

Tab. 18: Trefferraten der Geokodierungsergebnisse

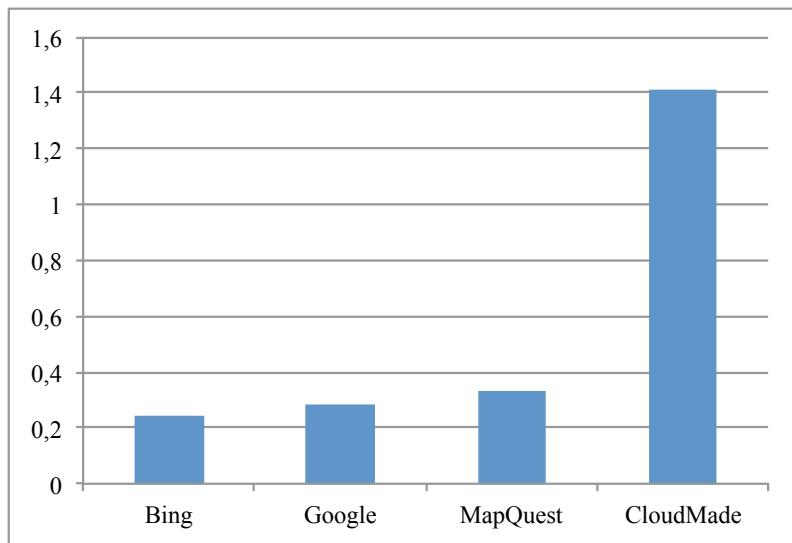
Des Weiteren unterteilten ROONGPIBOONSOPIT und KARIMI (2010) die Adressen nach Gesellschaftstypen (städtisch, vorstädtisch, ländlich) sowie der Flächennutzung (Wohngegend, Gewerbe, Landwirtschaft, Industrie). Auf Grund der Ausgangsdaten dieser Studie konnte keine Unterscheidung bezüglich dieser Klassifizierungen vorgenommen werden. Deshalb werden diese konkreten Ergebnisse hier nicht weiter diskutiert. Der Lagefehler wurde durch die Differenz der geokodierten Ergebnisse mit den tatsächlichen Koordinaten der Ausgangsdaten ermittelt. Tab. 19 gibt diesbezüglich die minimale und maximale Abweichung, den Median, den Mittelwert und die Standardabweichung (jeweils in Meter) für die untersuchten Dienste an, wobei Ausreißer nicht berücksichtigt wurden. Dabei ist zu erkennen, dass *Bing* die besseren Ergebnisse, bezogen auf die Lagegenauigkeit der Geokodierungsergebnisse, liefert, da sowohl Mittelwert, Standardabweichung, Median und maximale Abweichung die geringsten Werte einnehmen.

Dienst	Min. Abw.	Max. Abw.	Median	Mittelwert	Standardabw.
Google	5,5	44.537,4	178,5	1.000,1	3.603,6
Bing	8,9	12.966,0	171,0	436,4	1.070,8
MapQuest	6,8	59.926,7	238,4	1.474,0	4.646,7

Quelle: nach Roongpiboonsoopit und Karimi 2010: 1091

Tab. 19: Abweichungen der Geokodierungsdienste in Meter

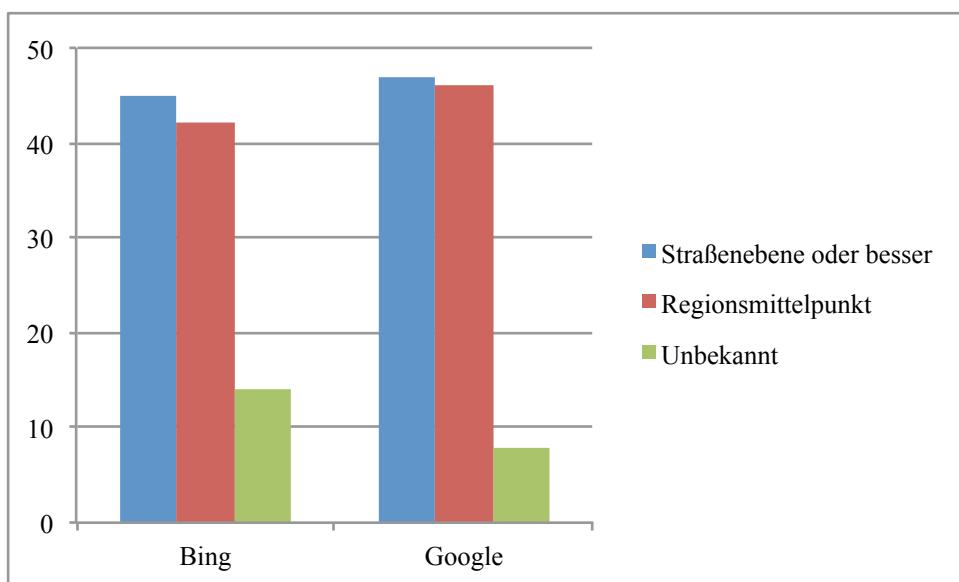
Eine Studie hinsichtlich der Schnelligkeit von Online-Geokodierungsdienste lieferte DUVANDER (2012). Zur Ermittlung der Geschwindigkeit wurde eine Woche lang jeweils eine Adresse alle zehn Minuten an die Dienste gesendet (siehe Abb. 17). Des Weiteren wird erwähnt, dass nur die Services von *Google* und *CloudMade* die ganze Zeit über erreichbar waren, wobei die meisten anderen lediglich eine einzelne Abfrage nicht beantworten konnten.



Quelle: nach DUVANDER 2012

Abb. 17: Durchschnittliche Geokodierungszeit pro Anfrage in Sekunden

XU et al. (2012) untersuchten die Geokodierungs-APIs von *Bing* und *Google* auf die Qualität der Ergebnisse. Dafür haben sie einen Testdatensatz der USA mit 150.000 Adressen jeweils über beide Services geokodiert, wobei die zurückgelieferten Ergebnisse auf die vom Anbieter mitgelieferte Präzision hin überprüft wurden. Diese Parameter wurden in die drei Genauigkeitskategorien „Straßenebene oder besser“, „Regionsmittelpunkt“ und „unbekannt“ unterteilt. Die etwas besseren Ergebnisse lieferte hierbei der Dienst von *Google* (siehe Abb. 18).



Quelle: nach XU et al. 2012: 7

Abb. 18: Geokodierungsqualität von Bing und Google nach Genauigkeitsstufe in Prozent

Die letzte in dieser Arbeit präsentierte Vergleichsstudie zur Evaluierung von Online-Geokodierungsdiensten stammt von DAVIS u. ALENCAR (2011), die den *Google* Geokodierungsdienst für eine brasilianische Großstadt in ihrer Arbeit anwendeten. Der

verwendete Datensatz hat eine Größe von ca. 540.000 Einträgen, wobei in 99,9 % der Fälle *Google* ein Ergebnis liefern konnte. Zur Untersuchung der geokodierten Punkte wurde der Genauigkeitswert, der im Ergebnis mitgeliefert wird, herangezogen. Diese Werte werden in neun Stufen unterteilt, wobei die höchste Genauigkeit bei einem Rückgabewert von neun gegeben ist. Tab. 20 lässt erkennen wieviel Prozent der Adressen mit einem bestimmten Genauigkeitswert geokodiert werden konnten.

Genauigkeit	Anzahl	Prozent
Nicht definiert	662	0,1%
3	7	0,0%
4	123.337	22,7%
6	13.623	2,5%
8	405.016	74,6%
9	435	0,1%
Summe	543.080	100%

Quelle: nach DAVIS und ALENCAR 2011: 859

Tab. 20: Ergebnisse gruppiert nach Genauigkeitsstufe

Nach eingehender Betrachtung der vorgestellten Studien kann festgehalten werden, dass kein Online-Geokodierungsdienst in der Lage ist für alle Daten ein zuverlässiges Ergebnis zu liefern. Hinsichtlich „Genauigkeit“ und „Schnelligkeit“ sind die Dienste von *Google* und *Bing* gegenüber den anderen untersuchten Diensten überlegen. Dies konnte auch im Rahmen dieser Arbeit bestätigt werden. Da allerdings alle für die Evaluierung verwendeten Datensätze aus Adressen außerhalb Europas bestehen, können diese Studien nicht spiegelbildlich auf die in dieser Arbeit vorhandene Problemstellung umgelegt werden.

3.5.6 Diskussion der Ergebnisse und Auswahl eines Geokodierungsdienstes

Die Evaluierung der Geokodierungsdienste erfolgte hinsichtlich der fünf Parameter „Schnelligkeit“, „Lagefehler“, „Trefferraten“, „Limitierungen“ und „Datenfehler“. Die Dienste von *Bing*, *CloudMade*, *Google*, *MapQuest* und *OpenStreetMap* wurden dabei untersucht, wobei die Ergebnisse mittels eines Analytischen Hierarchieprozesses bewertet wurden.

Bezogen auf die „Schnelligkeit“ schnitt der *Google* Geokodierungsdienst am besten ab, wobei die Geokodierung einer Adresse im Durchschnitt 0,2 s dauerte, gefolgt von *Bing* (0,37 s) und *OpenStreetMap* (0,59 s). Am schlechtesten schnitten die Dienste von *CloudMade* (0,59 s) und *MapQuest* (0,64 s) ab.

Bei Betrachtung der Lagefehler der resultierenden Adressen bzw. hinsichtlich der Verteilung der Distanzen einzelner Koordinatenpaare sind die Ergebnisse für *Google* und *Bing* annähernd gleich. *Bing* lieferte für die Hälfte der Daten genauere Geokodierungsergebnisse als *Google*, verzeichnet jedoch in der zweiten Hälfte des Testdatensatzes größere Distanzen und mehr Ausreißer. Bei den drei anderen Diensten können *OSM* und *MapQuest* ebenfalls annähernd gleichgut bewertet werden. *CloudMade* liegt mit einem großen Rückstand an letzter Stelle. An dieser Stelle muss jedoch festgehalten werden, dass auch die Dienste von *OSM* und *MapQuest* nur bedingt zufriedenstellende Ergebnisse

lieferten, da mehr als 10 % der Adressen eine Abweichung von über einem Kilometer gegenüber der tatsächlichen Position aufweisen. Bei *CloudMade* betrifft dies sogar fast 25% der Daten. Bei einer Rasterweite unter einem Kilometer kann dieses Faktum bereits zu stark verfälschten Interpretationsergebnissen führen.

Bezogen auf die „Trefferraten“ weisen alle Dienste über 90 % auf. Auch hier ragt *Google* mit 99,61 % an resultierenden Koordinaten heraus. Obwohl *CloudMade* ebenfalls sehr hohe Trefferraten (98,72 %) verbucht, muss das Ergebnis in Kombination mit der Lagegenauigkeit interpretiert werden, da viele Ergebnisse hohe Abweichungen von der tatsächlichen Position aufweisen. *Bing* weist mit knapp 98 % auch gute Trefferraten mit qualitativ guten Ergebnissen auf. Der Prozentsatz an zurückgegebenen Adressen liegt bei *MapQuest* und *OpenStreetMap* innerhalb von 92 % - 93 %.

Die „Zuverlässigkeit“ von Geokodierungsdiensten spiegelt sich in deren Limitierungen und dem Umgang mit Fehlern in Adressdatensätzen wider. Die „Limitierungen“ wurden bereits genauer erläutert und bewertet. Da der *OpenStreetMap*-Dienst annähernd frei von Limitierungen ist, ist dieser, bezogen auf diese Kategorie, die am besten geeignete Alternative. Allerdings können die Limitierungen bei anderen Diensten z.T. durch Zahlungsaufwendungen reduziert werden, wodurch bspw. mehrere Anfragen pro Tag oder Monat gestellt werden können. Fehlern in den Quelldaten werden bei allen Diensten annähernd gleich gut erkannt und behandelt. Lediglich die falsch geschriebene Adresse führt bei den meisten Diensten zu fehlerhaften oder im schlimmsten Fall zu gar keinen Ergebnissen.

Die Bewertung mittels Analytischen Hierarchieprozesses ergibt folgende Reihung der Geokodierungsdienste:

1. *Google* (29,32 %)
2. *Bing* (24,83 %)
3. *OpenStreetMap* (16,30 %)
4. *CloudMade* (15,04 %)
5. *MapQuest* (14,51 %)

Das Ergebnis ist plausibel, da *Google* und *Bing* in fast allen Kategorien, mit Ausnahme bei den Limitierungen, besser als die Konkurrenzdiensste abschnitten. Des Weiteren erscheint der Dienst von *Google* im Rahmen dieser Arbeit als der am besten Geeignete. Da jedoch für die Erstellung von Karten auf Basis von Rastereinheiten teilweise sehr große Mengen an Adressen verarbeitet werden müssen, ist die Limitierung von 2.500 Adressen pro Tag nicht ausreichend. Zur Lösung dieses Problem gibt es zwei Alternativen: Einerseits könnte der Geokodierungsdienst von *OpenStreetMap* verwendet werden der hinsichtlich Geokodierungsqualität ebenfalls gut abgeschnitten hat, wobei allerdings etwas ungenauere Ergebnisse in Kauf genommen werden müssen. Andererseits besteht die Möglichkeit den kommerziellen Dienst von *Google* zu benutzen, wobei jedoch für das zusätzliche Kontingent an Anfragen bezahlt werden muss. Da dies den finanziellen Rahmen dieser Arbeit jedoch sprengen würde, wird in weiterer Folge zur Erstellung von Karten auf Rasterbasis der freie Dienst von *OpenStreetMap* für große Datenmengen verwendet. Im

Falle der Verwendung dieses Online-Tools im Bereich des Geomarketings o.dgl. ist allerdings die kostenpflichtige, jedoch genauere Variante mit Hilfe des *Google-Geokodierungsdienstes* zu bevorzugen.

4 Daten auf Basis regionalstatistischer Raster in Österreich und Europa

In diesem Abschnitt der Arbeit wird aufgezeigt, wo Daten auf Basis von regionalstatistischen Rastern bzw. adressgenaue Daten für Österreich und Europa zur Verfügung stehen. Auf europäischer Ebene wird dabei vor allem auf das *GEOSTAT*-Projekt Bezug genommen, welches vom *European Forum for Geostatistics (EFGS)* ins Leben gerufen wurde. Auf österreichischer Ebene werden Datenanbieter aufgezählt, die rasterbasierte Daten anbieten. In Hinblick auf adressgenaue Daten, wird auf die jeweiligen Ausführungen der Statistik- oder Kartographie- bzw. Geodäsie- oder Geoinformations-Institute der Länder verwiesen. Dies soll die Möglichkeiten aufzeigen, welche Daten als Input für die im Rahmen dieser Arbeit erstellte Webapplikation verwendet werden können. Dabei werden vor allem bereits existierende Daten aufgezeigt und erläutert.

Neben den bestehenden Daten der verschiedenen Länder und Organisation können ebenfalls beispielsweise unternehmensinterne Daten verwendet werden. Sofern diese auf adressgenauer oder koordinatengenauer Basis vorliegen, können die Daten mit Hilfe der Online-Anwendung auf einer Karte dargestellt werden. Jedoch können dafür auch die in diesem Teil der Arbeit beschriebenen Daten als Grundlage verwendet werden.

In weiterer Folge werden die *INSPIRE*-Richtlinien für regionalstatistische Rasterdaten erklärt. Zum Ende des Kapitels werden wichtige Datenschutzaspekte diskutiert und aktuelle und zukünftige Entwicklungen sowie Forschungsstände in den teilnehmenden Ländern dargestellt.

4.1 Projektionen für Daten auf Rasterbasis

Die Gebietsgliederung in Raster kann in koordinatengebundene und koordinatenunabhängige Netzkonfigurationen unterschieden werden. Um die zeitliche Vergleichbarkeit von Daten zu gewährleisten, wird allerdings ein koordinatenabhängiges Netz benötigt. (vgl. WONKA 2010: 28)

In Österreich kommen in diesem Kontext vor allem jene Projektionssystemen in Frage, die in der amtlichen Kartographie verwendet werden, wie bspw. für großmaßstäbige Karten das Bundesmeldenetz (BMN, Anm.: = eine modifizierte Gauß-Krüger Abbildung) bzw. das für neuere Karten gültige UTM-Netz (Universale transversale Mercator-Abbildung). Für österreichweite Darstellungen, wie z.B. die ÖK500 wird eine konforme Lambert'sche Kegelabbildung verwendet (vgl. ebd.: 28). Auf Grund der Tatsache, dass die meisten rasterbasierte Daten für das gesamt Bundesgebiet vorliegen, wurde im Rahmen dieser Arbeit für das Gebiet von Österreich ebenfalls die Lambert-Projektion als Bezugsbasis gewählt. Es handelt sich dabei um einen Schnittkegel, der den räumlichen Bezugskörper auf 46° sowie 49° n.B. schneidet. Außerdem werden Additionskonstanten für den Rechts- und Hochwert von jeweils 400.000 m verwendet. Diese Anpassungen orientieren sich an die Anforderungen der ÖK500, weshalb länderübergreifende Vergleiche basierend auf dieser Darstellung schwierig sind.

Auf europäischer Ebene kam es zu einer Einigung auf einen Raster basierend auf dem ETRS-LAEA Koordinatensystem, wobei Rastergrößen von 100 m und 1.000 m zur Verfügung stehen. Eine wesentliche Rolle bei der Homogenisierung der Geodaten auf europäischer Ebene spielt die *INSPIRE*-Initiative. Diese verfolgt das Ziel eine europäische Geodatenbasis mit integrierten raumbezogenen Informationsdiensten zu schaffen. Durch die Verwendung eines einheitlichen Projektionssystems sollen die Mitgliedsstaaten dazu angeregt werden, auf regionale Projektionssysteme zu Gunsten des ETRS-LAEA Systems zu verzichten. Vorteile eines einheitlichen Systems ergeben sich sowohl für die Kartennutzer als auch für Kartenerzeuger. Bei grenzüberschreitenden Darstellungen müssen dadurch keine arbeits- und zeitintensiven Transformationen mehr durchgeführt werden. (vgl. ebd.: 29ff)

Die Vorteile eines standardisierten europäischen Rasters sind laut WONKA (ebd.: 34) folgende:

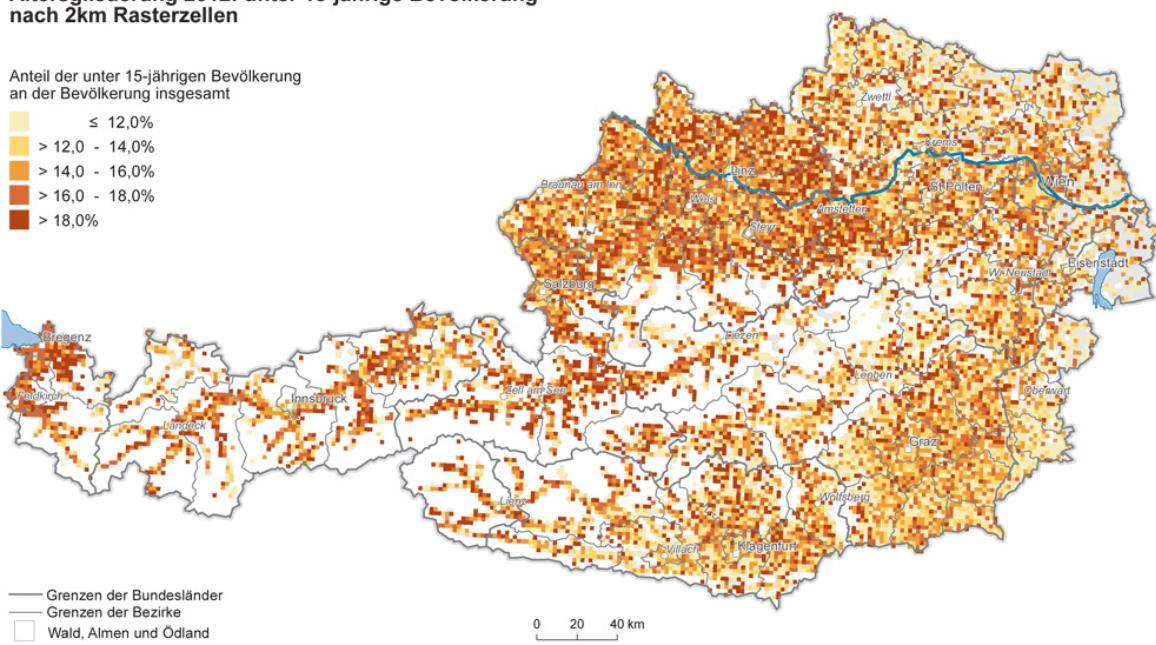
- *Austausch von Daten unterschiedlicher Behörden und Institutionen aus verschiedenen Ländern.*
- *Erleichterung der technischen Zusammenführung von Geometriedaten aus unterschiedlichen Quellen.*
- *Grenzüberschreitende Harmonisierung von digitalen Geodatenbeständen.*
- *Erleichterte Integrationsmöglichkeit internationaler Daten.*
- *Arbeitserleichterung in internationalen Arbeitsgruppen.*

4.2 Daten in Österreich

Koordinatengebundene Daten bzw. Datenbanken können als Voraussetzung für die Zuweisung von statistischen Daten zu Rasterzellen angesehen werden (vgl. ebd.: 34). In weiterer Folge wird erläutert, welche Datenbanken in Österreich zur Verfügung stehen, die diese Eigenschaften aufweisen.

Die Bundesanstalt *Statistik Austria* bietet ausgewählte Inhalte der Großzählung 2011 (Volks-, Gebäude- und Wohnungszählung sowie Arbeitsstättenzählung) auf der Basis von regionalstatistischen Rastereinheiten an, in den Rastergrößen von 125 m, 250 m, 500 m, 1 km, 2,5 km, 5 km sowie 10 km. Auf kleinster Rasterebene (125 m) werden allerdings nur Fallzahlen aus der Großzählung (z.B. Zahl der Wohnbevölkerung, Gebäudeanzahl, Beschäftigtenzahl, etc.) bereitgestellt. Genauere Merkmale zu den Fallzahlen (z.B. Familienstand, Geschlecht, Haushaltsgröße, Gebäudealter, etc.) sind erst ab einer Rastergröße von 250 m erhältlich, wobei jedoch bestimmte Mindestausprägungen pro Rasterzelle erfüllt werden müssen (siehe Kapitel 4.4). In Abb. 19 wird beispielhaft die Altersgliederung in Österreich 2012 auf Basis von 2km-Rasterzellen vorgestellt. (vgl. STATISTIK AUSTRIA 2013b)

**Altersgliederung 2012: unter 15-jährige Bevölkerung
nach 2km Rasterzellen**



Quelle: bearbeitet nach STATISTIK AUSTRIA 2013b

Abb. 19: Altersgliederung der unter 15-jährigen Bevölkerung 2012 auf Basis 2km-Rasterzellen

Mit der Inbetriebnahme des Adress-GWR-Online im November 2004 wurde eine Basis zur Führung eines Adress-, Gebäude- und Wohnungsregisters geschaffen. Das Unternehmensregister enthält alle Unternehmen und Arbeitsstätten, ausgenommen land- und forstwirtschaftliche Betriebe, die in einem eigenen Betriebsregister gespeichert werden. Des Weiteren werden Daten aus dem Zentralen Melderegister (ZMR), in dem alle gemeldeten Personen erfasst sind, laufend in eine bevölkerungsstatistische Datenbank (POPREG) integriert, die ebenfalls bei der *Statistik Austria* aufliegt. (vgl. WONKA 2010: 36ff)

Das Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft stellt eine eigene zentrale und standardisierte Geodatensammlung zur Verfügung, wobei ein Großteil dieser Daten auch auf Basis von Rastereinheiten vorhanden ist. Im Folgenden werden die derzeit auf Rasterbasis verfügbaren Datensätze des Lebensministeriums aufgelistet. (vgl. FAHRNER 2012: 22ff)

- *Verwaltungsgebietsinformationen*
- *Rasterquerverweistabelle: Zuordnungstabellen zwischen verschiedenen Rasterebenen des INSPIRE-Rasters.*
- *Postadressinformation: Anzahl der Adresspunkte je Rasterzelle.*
- *Grundbuchinformation: Umlage der Grundbuchhauptnutzungsarten je Rasterzelle.*
- *Benachteiligte Gebiete*
- *Landwirtschaftliche Bodenkartierung*
- *Geländehöhenmodell*
- *Landwirtschaftliche Förderungsabwicklung der AMA: Auf Raster umgelegte aus dem INVEKOS-GIS stammende Feldstücknutzungsart (z.B.: Ackerland, Grünland usw.),*

Schlagnutzungsartengruppen (z.B.: Getreide, Hackfrüchte usw.) und GVE-Großviecheinheiten.

- *Klimadaten der ZAMG*
- *Ertragspotential: Auf Raster umgelegtes landwirtschaftliches Ertragspotential abgeleitet aus dem Geländehöhenmodell und von Klimadaten nach der Methode Schaumberger.*
- *Hagelversicherung: Auf Raster umgelegte Informationen der österreichischen Hagelversicherung aus dem Teilprojekt Schäden (Frost, Hagel, Dürre usw.), Teilprojekt Entwicklung (Pflanzenentwicklung im täglichen Zeitraffer) und Wasserbilanz.*
- *Wald-Layer aus ÖK*

Wie bereits erwähnt können diese Daten zur Erstellung von Karten mit der im Rahmen dieser Arbeit entworfenen Webapplikation herangezogen werden. Gut dafür geeignet sind Adressdaten aus dem Adress-GWR-Online, da diese durch eine vorhergehende Geokodierung direkt auf Rasterbasis umgelegt werden können. So könne beispielsweise einfache Bevölkerungsverteilungskarten für Österreich auf Basis regionalstatistischer Raster erstellt werden. Daten, welche bereits auf Rasterbasis vorliegen, müssen eventuell vor der Verarbeitung mit dem Online-Tool noch bearbeitet werden, falls diese nicht in dem benötigten Raster vorliegen.

4.3 Daten in Europa

Harmonisierte und frei verfügbare Daten werden auf europäischer Ebene vorrangig im Rahmen des *GEOSTAT*-Projektes angeboten. Im Zuge dieses Projektes wurden Bevölkerungsdaten auf Basis eines 1 km-Rasters erstellt und öffentlich zugänglich gemacht (vgl. KAMINGER 2010c).

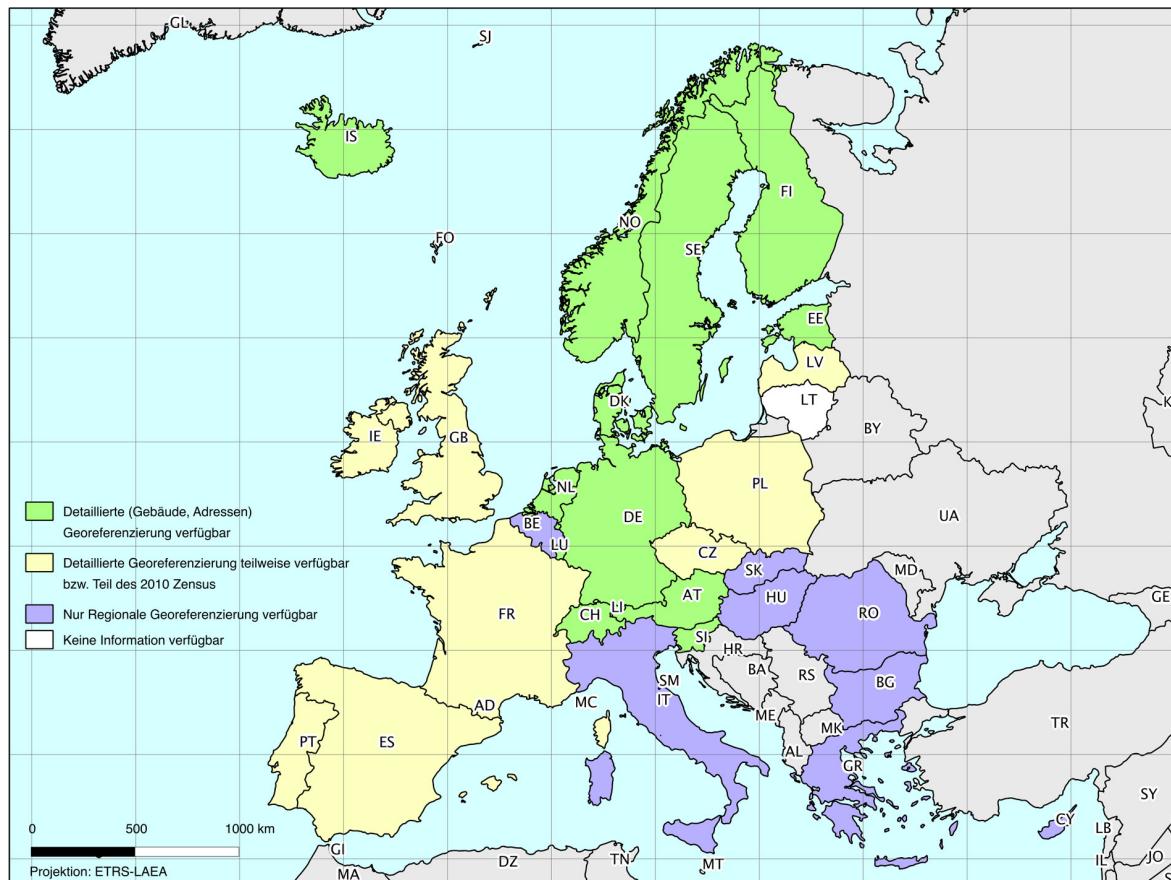
Der erste Teil des *GEOSTAT* Projekts beschäftigte sich mit der Entwicklung von Richtlinien für die Darstellung der Ergebnisse aus der Bevölkerungszählung 2010/2011 auf Basis eines harmonisierten 1 km-Rasters (vgl. BLOCH 2010). Derzeit gibt es elf Länder, die detaillierte koordinatengebundene Bevölkerungsdaten bereitstellen, wodurch eine Darstellung auf Basis aggregierter Daten möglich ist (vgl. KAMINGER 2010a). Das *GEOSTAT* Projekt wird durch das Statistische Amt der Europäischen Union (*EUROSTAT*) sowie den teilnehmenden nationalen statistischen Ämtern und im Rahmen von ESSnet (European Statistical Systems Network) geführt (vgl. KAMINGER 2010c).

Da *GEOSTAT* als Pionier-Projekt für die länderübergreifende Darstellung von rasterbasierten Daten gilt, wird es im Rahmen dieser Arbeit näher beschrieben und die Vorgehensweise erläutert. Die Bearbeitung dieser Daten kann als Grundlage für den Entwurf der im Rahmen dieser erstellten Webapplikation angesehen werden.

4.3.1 Das *GEOSTAT 1A* Projekt

Das *GEOSTAT*-Projekt mit dem Titel „*Darstellung von Zensus Daten in einem europäischen Bevölkerungs-Rasterdatensatz*“ wurde im Jahr 2010 ins Leben gerufen. Der

Abschnitt *GEOSTAT 1A* verfolgt das Ziel die Methodologie für einen rasterbasierten Bevölkerungsdatensatz zu schaffen. Vielen nationalen Statistikinstituten ist es mittlerweile möglich detaillierte Statistiken für kleine Flächeneinheiten auf Grund genauer Georeferenzierung durchzuführen (siehe Abb. 20). Im Zuge dessen wurden Voraussetzungen und Methoden für die Disaggregation von Daten in Ländern ohne koordinatengebundene Erhebungen entwickelt. Basierend auf den Nutzeranforderungen wurden Datenspezifikationen, Projektionssysteme, Methoden für räumliche Analysen als auch für Kartenentwürfe, sowie Lösungsansätze zu Datenschutzproblemen entworfen. (EFGS 2012: 7ff)



Im Zuge von *GEOSTAT* wurde besonderes Augenmerk auf die Qualität der Datensätze gelegt, da es in dieser Hinsicht viele verschiedene Ansätze zur Ermittlung dieser Daten gibt (geographisch, statistisch, Erzeugung, Standards, etc.). Die Erstellung von rasterbasierten Daten wird in den einzelnen Ländern mit Hilfe unterschiedlichster Methoden und Datenquellen durchgeführt, wobei die Verwendung diverser Methoden oft nicht ausreichend dokumentiert ist. Die Qualität wird meist von zwei Seiten betrachtet: Die Qualität der georeferenzierten Basisdaten sowie die Qualität des Produktionsprozesses. Die *INSPIRE*-Richtlinie beinhaltet hauptsächlich Vorgaben hinsichtlich der Genauigkeit der Georeferenzierung bzw. räumlichen Lage der Daten. In der Statistik gibt es eigene Standards bezüglich der Produktion von Daten. (vgl. ebd.: 7)

Da Datenschutzrichtlinien in Abhängigkeit vom jeweiligen Mitgliedstaat variieren, ist es schwierig einen europaweiten Datensatz mit verschiedenen Attributausprägungen zur Verfügung zu stellen. Dennoch ist das *EFGS* bestrebt, einem Ansatz nachzukommen, der eine zentrale und kostenlose Verteilung der Daten vorsieht. Derzeit ist ausschließlich die Gesamtbevölkerung auf Basis eines 1 km-Rasters zentral für alle Länder verfügbar. Probleme treten dann auf, wenn zusätzliche spezielle Merkmalsausprägungen (z.B. Bevölkerung pro Altersgruppe, Bevölkerung pro Geschlecht) benötigt, oder die Daten für Raster unter 1 km angefordert werden. (vgl. ebd.: 8)

4.3.2 Der *GEOSTAT 1A* Datensatz

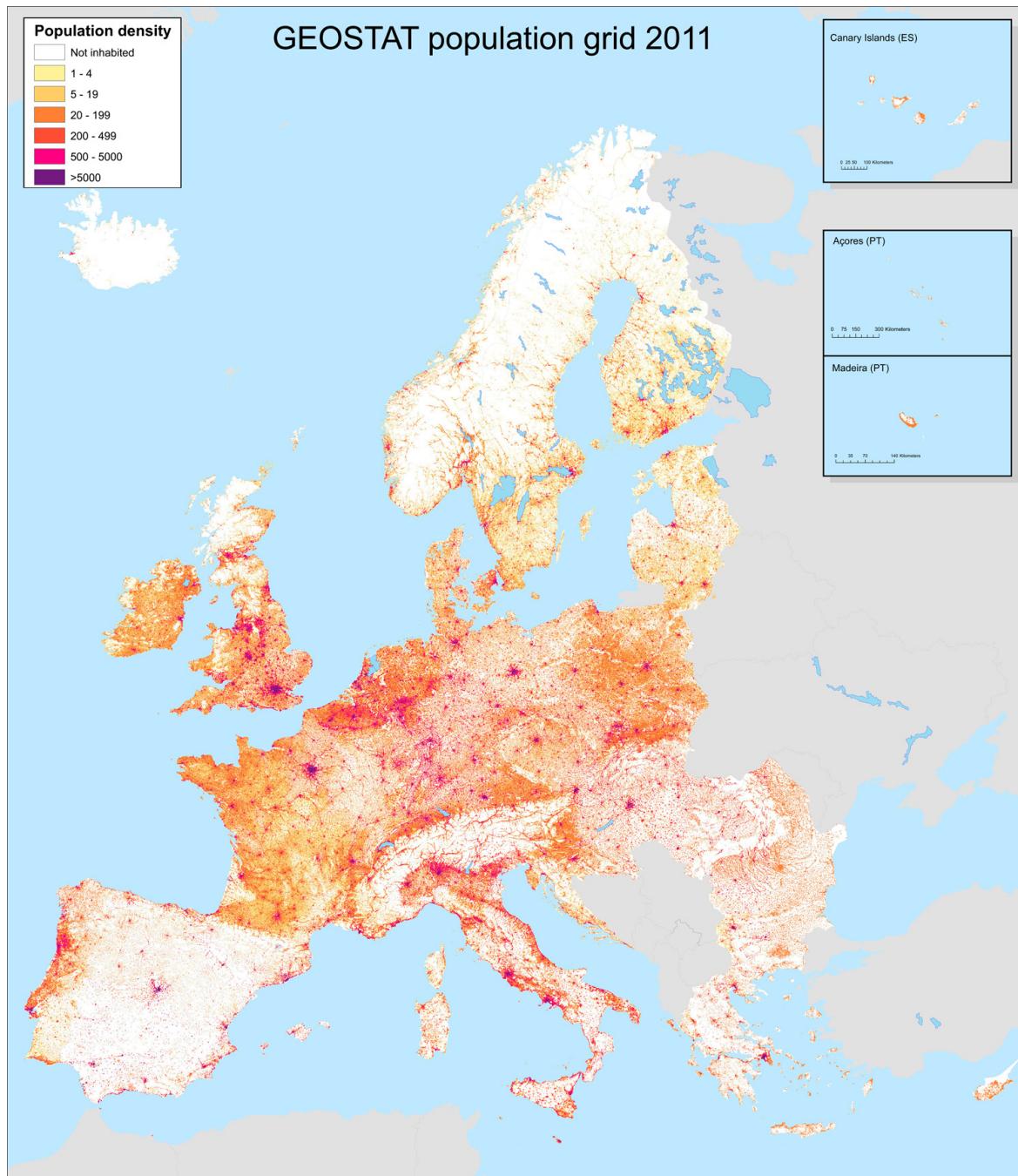
Wie bereits im vorherigen Kapitel erwähnt wurde, ist eines der Ziele des *GEOSTAT*-Projekts die Bereitstellung von Richtlinien zur Erstellung von Rasterstatistiken. Um den Prozess der Herstellung dieser Daten zu vereinfachen, wurde ein Beispieldatensatz im Zuge des Projektes erstellt, der die wichtigsten Inhalte aufzeigen soll. Dabei handelt es sich um einen Bevölkerungsdatensatz auf Basis eines 1 km-Rasters (siehe Abb. 21). Die Daten entsprechen einer Zusammenführung unterschiedlicher nationaler Datensätze, denen unterschiedliche Herstellungsmethoden zugrunde liegen. Das Ergebnis ist ein europaweiter Bevölkerungsraster für das Jahr 2011. (vgl. *EFGS* 2012: 22)

Die verschiedenen Methoden und Datenquellen, auf denen die gerasterten Daten basieren nehmen hinsichtlich der Datenqualität eine zentrale Rolle ein. Im Zuge des *GEOSTAT 1A* Projektes haben sechs Staaten ihre Erfahrungen beim Herstellungsprozess genauer erläutert und mit der Öffentlichkeit geteilt. In Österreich und Finnland bspw. besteht ein nationales Register für Gebäude und Wohnhäuser, die eine bedeutende Rolle in der Bereitstellung der benötigten Georeferenzierungsgenauigkeit spielen (vgl. KAMINGER 2010c und TAMMILEHTO-LUODE 2010). In Norwegen liegt die Verantwortung für diese Daten beim öffentlichen Register für Grundeigentum, Adressen und Gebäude (vgl. BLOCH et al. 2010).

Der erste Ansatz für einen europaweiten rasterbasierten Bevölkerungsdatensatz basiert auf einem *INSPIRE*-konformen 1 km Rasternetz, welches die EU27 und EFTA Staaten umfasst. Um den Qualitäts- und Genauigkeitsansprüchen Genüge zu leisten, wurden, wo es möglich war, Punktdatenquellen direkt in das ETRS-LAEA Netz projiziert und anschließend aggregiert. Länder, in denen Daten nicht in Form geokodierter Punkte zur Verfügung stehen, haben die Informationen auf Basis der kleinsten verfügbaren Rasterzellen auf den *INSPIRE*-Raster umgeformt. (vgl. ebd.: 28ff)

Der *GEOSTAT 1A* Datensatz, der die Fläche der EU27 und EFTA Staaten umfasst, beinhaltet 4.884.516 Rasterzellen, wovon 1.945.315 (ca. 40 %) Zellen mit mindestens einer Person bevölkert sind. Der aggregierte Teil des Datensatzes umfasst 43 % der Bevölkerung sowie der disaggregierte Teil respektive 57 %. Bezogen auf das gesamte Gebiet beinhaltet eine Rasterzelle durchschnittlich 86 Einwohner bzw. eine bewohnte Zelle 170 Personen. Die Zelle mit der höchsten Bevölkerungsdichte (52.898 Einwohner pro km²) befindet sich im Zentrum von Barcelona. In etwa 70.000 Zellen lebt nur eine

Person, wovon 20.774 Zellen aus datenschutzrechtlichen Gründen nicht ins Bevölkerungsraster aufgenommen werden konnten. (vgl. ebd.: 30f)



Quelle: bearbeitet nach EFGS 2012

Abb. 21: Europaweiter Bevölkerungsraster 2011

Bereits in der Anfangsphase von *GEOSTAT* wurden einige Verbesserungspotentiale offensichtlich. Die Themen reichen von umfassenderen und detaillierteren Metadaten sowie standardisierten Qualitätskontrollen bis hin zu einheitlichen Aktualisierungsprozeduren für die Datengrundlagen. Eine erste Maßnahme zur Steigerung der Qualität ist die Eingliederung der Zensusdaten von 2011 in den *GEOSTAT*-Datensatz. Bis zum aktuellen

Zeitpunkt beinhalteten die Daten lediglich die gesamte Bevölkerung. Allerdings sind in manchen Ländern, die ihre Daten in Registern führen (Österreich, Norwegen, Schweden, Finnland, Niederlande, Dänemark, Slowenien) bereits genauere Merkmalsausprägungen (z.B. Bevölkerung pro Alterskategorie) verfügbar. Derzeitige Entwicklungen, bei denen die Adressdaten und die Geokodierung von Gebäuden in Kombination mit den Zensusdaten von 2011 stehen, ermöglichen die Ausweitung von Variablen auf Basis von Rasterzellen. (vgl. ebd.: 34f)

Das Nachfolgeprojekt von *GEOSTAT 1A*, *GEOSTAT 1B* (Laufzeit: 2012-2014), verfolgt das Ziel Richtlinien für Datensätze und Methoden zu entwickeln, um damit die Bevölkerungsstatistik der Zensusdaten von 2011 in einem harmonisierten Raster darzustellen. Dieses Projekt wurde von den Partnern des *EFGS* ins Leben gerufen, wobei der Großteil der Teilnehmer bereits am Vorgängerprojekt mitgearbeitet hat. Ziel ist es Richtlinien für andere nationale Statistikinstitute bereitzustellen, die für zwei Ansätze produziert wurden. (vgl. NORDBECK 2013: 3)

- „*Bottom-Up*“ Länder bei denen die georeferenzierte Bevölkerungsregister bzw. georeferenzierte Bevölkerungsdaten aus dem Zensus vorliegen. Diese müssen die Daten auf Basis von 1km-Rastern aggregieren.
- „*Hybride*“ Länder welche georeferenziert Bevölkerungsdaten mit berechneten Daten auf Basis von Satellitenbildern verschneiden.

Das *GEOSTAT 1B* Projekt legt das Hauptaugenmerk dabei auf die folgende wichtigen Themenbereiche (vgl. ebd.: 4ff):

- Qualitätsparameter der Datensätze sowie Aggregations- und Disaggregationsmethoden und die Integration von Bevölkerungsrasterdaten
- Produktion von Richtlinien für Variablen und statistische Einheiten, mit Fokus auf Open Data-Standards
- Bereitstellung von Diensten zur Darstellung, zum Download sowie zum Kauf von bi- und multilateralen Datensätzen
- Aktualisierung, Wartung sowie Entwicklung der *EFGS* Webseite
- Verwertung und Verteilung der gewonnenen Daten.

4.4 Datenschutzbestimmungen und Vertraulichkeitsaspekte

Datenschutz spielt bei der Erstellung bzw. Veröffentlichung von rasterbasierten Statistiken und Darstellungen eine bedeutende Rolle. Obwohl europaweit unterschiedliche Gesetzgebungen vorliegen, besteht über das Recht einzelner Personen, das wichtigste Element bei Vertraulichkeitsfragen, weitestgehend Konsens. Daher sind Merkmalsausprägungen einzelner Rasterzellen, die auf einzelne Individuen (Gebäude, Personen) zurückgeführt werden könnten, unzulässig. Dies erfordert wiederum die Entwicklung von Strategien, die den Umgang mit Rasterzellen, die nur wenige Punktinformationen enthalten, reglementieren.

Bei klein gewählten Rasterzellen, besteht das direkte sowie indirekte Risiko einzelne statistische Einheiten zu enthüllen. Obwohl die nationalen Gesetzgebungen und jene der EU bezüglich Datenschutzes annähernd konform sind, werden diese länderspezifisch interpretiert und ausgelegt. *GEOSTAT* propagiert eine kostenlose und freie Verfügbarkeit von absoluten statistischen Einheiten (Anzahl Einwohner, Gebäude, Arbeitsplätze, etc.). Für spezifischere Anfragen (z.B. Geschlecht, Alter) sollten Datenschutzbestimmungen das Recht des Einzelnen wahren. Des Weiteren wurde die Empfehlung ausgesprochen, dass Charakteristiken zu Personen größeren Schutzbedarf haben, als Merkmale von Gebäuden. Zudem ist Vorsicht geboten, wenn Daten in mehr als einem Koordinatensystem erzeugt werden, da durch eine Umprojizierung detaillierte Merkmalsausprägungen aufgedeckt werden können. (vgl. EFGS 2012: 38f)

Die am häufigsten verwendete Methode für den Umgang mit Zellen, die zu geringe Merkmalsausprägungen aufweisen, ist diese einfach auszulassen. Das führt jedoch dazu, dass die Daten für Analysen nicht mehr im Detail verwendet werden können. Weitere Methoden basieren bspw. auf Rundungen der Werte (s. MASIK 2011), Verschlüsselungen (s. LIPATZ 2011), Gruppierungen (s. SOMMER 2007) oder auf die Verwendung unterschiedlicher Zellgrößen (s. SEHLIN 2011). Für die Verteilung eines europaweiten Datensatzes muss in diesem Zusammenhang eine geeignete Methode für vertrauliche Daten gefunden werden (vgl. EFGS 2012: 39)

In Österreich werden Rasterzellen mit zu geringen Fallzahlen bspw. nicht weitergegeben. Merkmale über Personen oder Haushalte sind für Rasterzellen erst ab mindestens 31 inkludierten Personen, die in dieser Zelle ihren Hauptwohnsitz haben, verfügbar. Im Falle von Gebäude- oder Wohnungsmerkmalen müssen sich mindestens vier Einheiten in einer Rasterzelle befinden. Somit müssen bei Datenanfragen die Einschränkungen für kleinräumige Auswertungen berücksichtigt werden. Alternativ kann eine Auswertung auf Basis größerer Maschenweiten (z.B. 500 m statt 250 m) vorgenommen werden. Tab. 21 gibt an, wieviel Prozent der Merkmale in Österreich in den verschiedenen Rastergrößen auf Grund von Datenschutzbestimmungen nicht weitergegeben werden dürfen. Bezogen auf Einwohnermerkmalen sind verhältnismäßig viele Merkmalsausprägungen auf Basis des 250 m-Rasters nicht freigegeben. (vgl. WONKA 2010: 35f)

Rastergröße	Einwohnermerkmale	Gebäudemerkmale	Wohnungsmerkmale
250m	20,7%	12,1%	5,9%
500m	10,2%	4,9%	2,3%
1km	2,9%	0,1%	0,4%
2,5km	0,2%	0,0%	0,1%

Quelle: nach Wonka 2010: 36

Tab. 21: Anteil verschiedener Bevölkerungsmerkmale die aus Datenschutzgründen in Österreich nicht weitergegeben werden

Für den Entwurf der Webapplikation bedeutet dies, dass eine Möglichkeit bedacht werden muss, um mit diesen Datenschutzproblemen umgehen zu können. So wäre ein Auswahlfeld denkbar, welches angibt, wie viele Ausprägungen eine Rasterzelle mindestens haben muss, damit sie auf der Karte angezeigt wird. Dies bedeutet, dass diejenigen Zellen, die zu

wenige Ausprägungen beinhalten, nicht farbig dargestellt werden. Neben den Vertraulichkeitsaspekten muss jedoch noch beachtet werden, dass Auswertungen auf Basis zu großer Rasterzellen eventuell nicht mehr aussagekräftig genug sind.

4.5 Entwicklungen

Aus den präsentierten Ergebnissen geht hervor, dass es noch einige Aspekte bezüglich der Erstellung und Verteilung von rasterbasierten Bevölkerungsstatistiken gibt, die noch genauer behandelt werden müssen. Vor allem in Hinblick auf die Datenschutzbestimmungen sollte ein einheitliches Konzept für ganz Europa erstellt werden, auf dessen Grundlage die Daten zugänglich gemacht werden sollten. Mit der verpflichtenden schrittweisen Umsetzung der *INSPIRE*-Initiative müssen die Länder ihre Daten und Metadaten in harmonisierter Art und Weise zur Verfügung stellen. Dadurch können auch rasterbasierten Daten in einem für Europa einheitlichen Rasternetz dargestellt werden.

In Österreich erfolgte bereits der Umstieg vom nationalen Rastersystem MGI-Lambert auf das ETRS-LAEA System, wobei nur noch das europaweite System unterstützt wird. Eine Umlegung mehrerer Daten auf Raster, wie bspw. ein jährliches Gebäudepaket, wird angestrebt (vgl. KATZLBERGER 2012). Andere Länder, wie etwa Italien, bieten noch keine Daten auf Rasterbasis an. Darüber hinaus gibt es keine Bestrebungen dies in naher Zukunft zu ändern (vgl. TICCA 2013). Anders verhält sich die Situation in Kroatien, Luxemburg und der Tschechischen Republik, wo zurzeit auch noch keine rasterbasierten Daten zur Verfügung stehen, jedoch die Veröffentlichung des Zensus 2011 in Rasterform in Planung ist (vgl. CRKVENČIĆ 2013, LARUE 2013 und MORAVEC 2013). Dieses Vorhaben wurde in Slowenien bereits realisiert, wobei Rasterdaten des Zensus 2011 kostenlos für Rastergrößen von 100 m angeboten werden. Zusätzlich ist die Erstellung der erhobenen Daten aus den letzten 15 - 20 Jahren in Planung (vgl. KUZMA 2013). In Frankreich, Irland, den Niederlanden und der Schweiz sind rasterbasierte Daten des Zensus 2011 ebenfalls kostenlos abrufbar (vgl. LIPATZ 2013, CORCORAN 2013, VAN LEEUWEN 2013 und HENRIOD 2013). Laut SCHNORR-BÄCKER (2013) sind in Deutschland hingegen aus rechtlichen Gründen keine rasterbasierten Daten erhältlich. Im Entwurf des E-Gouvernement-Gesetzes ist jedoch eine allgemeine Aufbereitung nach Rastern vorgesehen. Geplant ist, dass dieses Gesetz noch in der aktuellen Legislaturperiode verabschiedet werden soll. In Dänemark sind rasterbasierte Daten zwar vorhanden, allerdings kostenpflichtig (vgl. SOMMER 2013). Estlands Daten des Zensus von 2000 und 2011 sind ebenfalls kostenlos verfügbar. Andere Indikatoren, unabhängig vom Untersuchungsgebiet oder Auflösung, sind jedoch kostenpflichtig. Darüber hinaus ist die Integration der rasterbasierten Daten in eine Webmapping Applikation bereits teilweise vorgesehen (vgl. VALGMA 2013). Schweden bspw. speichert Daten wie etwa zur Bevölkerungsstatistik, Gebäudestatistik, Arbeitsplätze, etc. in punktbasierten Datenbanken ab, sodass Daten relativ einfach auf Raster aggregiert werden können (vgl. MOSTRÖM 2013). Für die übrigen, an *GEOSTAT* teilnehmenden, Ländern wurden keine genaueren Informationen zur Verfügung gestellt.

Im Rahmen von *GEOSTAT* wurde festgestellt, dass viele Länder die Einführung von neuen Variablen in den Bevölkerungsstatistiken befürworten. Dafür sollten die benötigten

Variablen im Vorhinein definiert werden, um individuelle Standards einzelner Länder zu vermeiden. Um disaggregierte und hybride Datenbestände zu vereinheitlichen, wurden bzw. werden weitere Richtlinien aufgestellt. (vgl. NORDBECK 2013: 13)

Dieses europäische GEOSTAT-Projekt soll einen fortwährenden Dialog zwischen privaten und öffentlichen Interessen ermöglichen. Während bei *GEOSTAT 1A* noch die Vision einer rasterbasierten Statistik in Europa im Vordergrund stand, wird im Zuge von *GEOSTAT 1B* die Produktion der Daten sichergestellt. In *GEOSTAT 1C* soll die vollständige Integration in den Ländern umgesetzt werden. Dadurch sollen sich die einzelnen Mitglieder als Teil eines globalen Systems verstehen und nicht mehr ausschließlich nationale Interessen verfolgen. Um diese Ziele zu erreichen, sowie zur Unterstützung der Länder, wurde das *EFGS* als Hauptansprechpartner für nationale Statistikinstitute ausgewählt. (vgl. ebd.: 3ff)

Anhand dieser Ausführungen kann beobachtet werden, dass rasterbasierte Daten nicht nur in den Statistikinstituten von großem Interesse sind, sondern auch bereits teilweise in privaten Geodatenstellen. Dies wird vor allem durch die EU-weite Regelung mit Hilfe der *INSPIRE*-Initiative erleichtert bzw. beschleunigt, da es für private Stellen schwierig ist, Geodaten in mehreren unterschiedlichen Formen vorzuhalten.

5 Regionalstatistische Raster

Dieses Kapitel der Arbeit befasst sich mit dem umfassenden Begriff der „regionalstatistischen Raster“. Einleitend wird eine kurze Einführung über diese Darstellungsart gegeben. Nach Skizzierung deren geschichtlichen Entwicklung, wird ein Vergleich mit anderen Darstellungsweisen von Gebietsgliederungen gezogen, wobei aufgezeigt wird, mit welchen Methoden Daten auf Rasterbasis abgebildet werden können. Eine Aufstellung verschiedener Abbildungsarten bzw. das Aufzeigen der Einsatzbereiche von Rasterdarstellungen soll darlegen, wie vielseitig anwendbar diese Kartenart ist. Zum Abschluss dieses Kapitels werden einige Anwendungsbeispiele aus den Bereichen der klassischen Regionalstatistik, Geomarketing, Erreichbarkeitsanalysen, Infrastrukturplanung, etc. präsentiert.

Gebietsgliederungen auf Basis von Verwaltungseinheiten (Bundesländer, Politische Bezirke, Gemeinden) können als Standard für die Abbildung statistischer Daten angesehen werden. Allerdings ist ein Trend zur Speicherung von raumbezogenen Individualdaten zu beobachten, weswegen die Bereitstellung statistischer Daten den Anforderungen der Raumwissenschaften entspricht. Lagerichtige Punktdaten können auf Basis verschieden großer Raster aggregiert und dargestellt werden (vgl. WONKA 2010: 1). Gemäß WONKA (ebd.: 1) sind regionalstatistische Raster „flächendeckend, regelmäßig, hierarchisch, unterteilbar bzw. zusammenfügbar. Dadurch sind kleinräumliche, regionale, überregionale und gesamtstaatliche Untersuchungen gleich gut verwendbar“.

Regionalstatistische Rastereinheiten bilden ein flächendeckendes, räumliches Bezugssystem für Geodaten aller Art. Auf Grund der Unabhängigkeit von Verwaltungsgrenzen und eventuellen Grenzänderungen sind reine sachbezogene Vergleiche möglich. Da, anders als bei Verwaltungseinheiten, Kleinräumigkeit bei regionalstatistischen Rastern gewährleistet werden kann, können bestimmte Umweltpheomene und räumliche Verteilungen besser erkannt werden. (vgl. KAMINGER 2010a: 238f)

Da regionalstatistische Informationen aus Demographie, Wirtschaft und dem Sozialbereich als Grundlage für planerische Arbeiten, Geomarketing, etc. eingesetzt werden, werden kleinräumige Gebietsgliederungen benötigt. Allerdings können adressbezogene Punktdaten aus datenschutzrechtlichen Gründen nicht eingesetzt werden, sodass Aggregate gebildet werden müssen. Diese können bspw. auf Basis von Baublöcken, administrativen oder geometrischen Bezugsflächen gebildet werden. Bei der Analyse eines Gebietes muss der Endmaßstab bekannt sein, da dies die Statistik einzelner Zellen beeinflussen kann. Als Einsatzbereiche administrativ unabhängiger Raumseinheiten kann nicht nur die Raumplanung, sondern u.a. die Marktforschung oder Infrastrukturplanung angesehen werden. Diese Darstellungsart wird zunehmend populärer, da die topographisch unabhängige Betrachtung von statistischen Informationen und Raumstrukturen wertvolle, räumlich fein gegliederte Informationen liefert. (vgl. PRINZ et al. 2004: 556f)

Da regionalstatistische Untersuchungen in der Vergangenheit maßgeblich auf Basis von administrativen Einheiten durchgeführt wurden, konnte nicht festgestellt werden, wie die

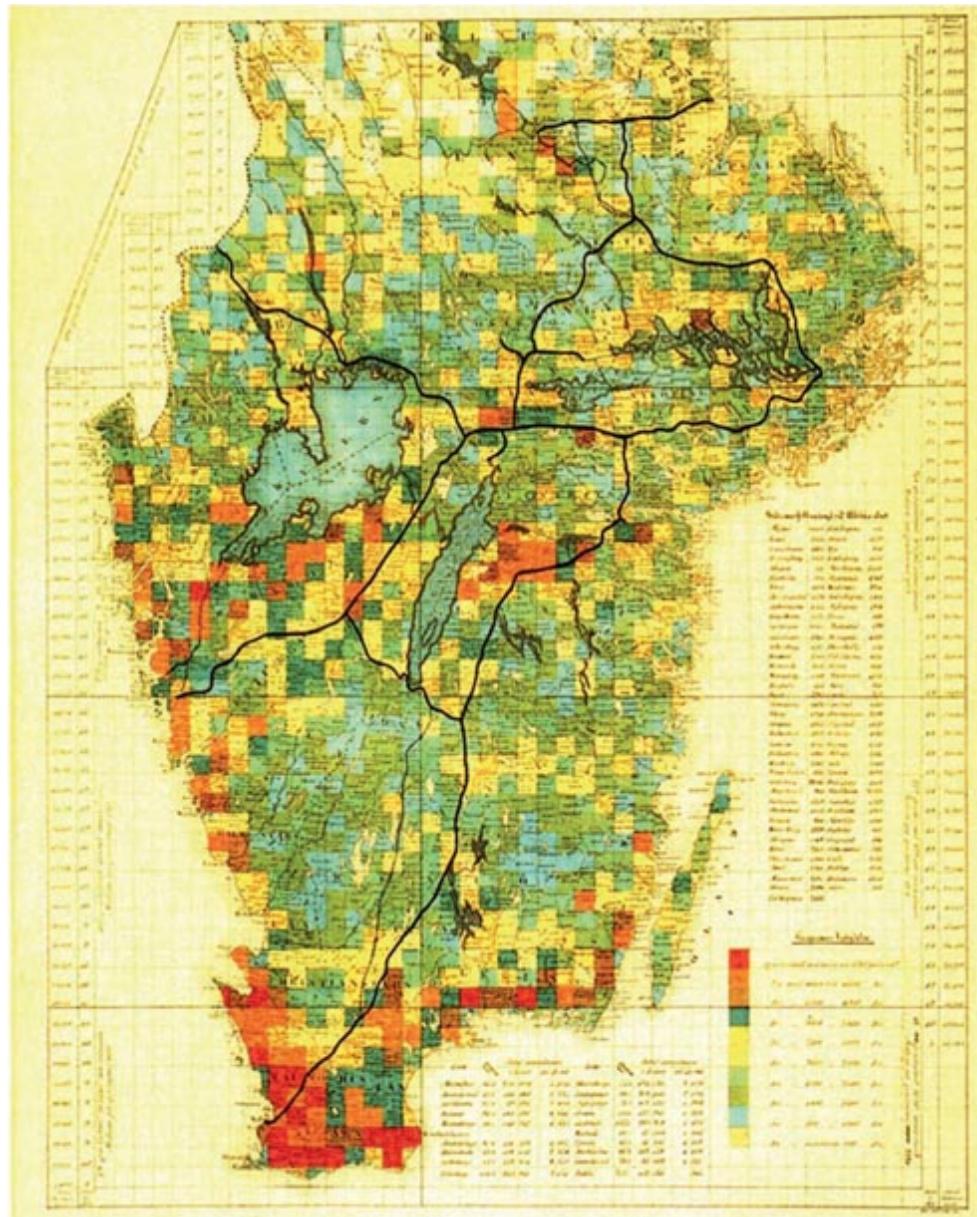
zu untersuchenden Werte durch die Grenzziehung beeinflusst wurden. Um fehlerhafte Interpretationen zu vermeiden, werden gleich große Raumeinheiten, die dem Kartenmaßstab anpasst sind benötigt. Bspw. stellten die Größenunterschiede und die fehlende Homogenität der NUTS- und LAU-Gliederungen Probleme im Zuge der Analyse statistischer Daten dar. Aus diesem Grund wurde in den vergangenen Jahren die Benutzung von regionalstatistischen Rastern für unabhängige Gebietsgliederungen forciert. (vgl. WONKA et al. 2007: 72f)

Diese Sachverhalte sollen im folgenden Abschnitt nochmals aufgegriffen und näher erläutert werden. Die erworbenen Erkenntnisse sollen im Zuge des Entwurfes einer Webapplikation weiterverwendet werden. So ist es beispielsweise wichtig, welche Arten von rasterbasierten Darstellungen es gibt, und ebenfalls wie statistische Daten auf Rasterbasis umgerechnet werden können.

5.1 Geschichte

Obwohl Darstellungen auf Basis regionalstatistischer Rastereinheiten erst im Zuge der letzten Dekade einen regelrechten Aufschwung erlebten, datiert die älteste bekannte Rasterkarte bereits aus dem 19. Jahrhundert. Die „*Bevölkerungsdichtheitkarte Südschwedens von 1856*“, die im Meilenraster abgebildet ist (siehe Abb. 22 Abb. 22:), wurde als Grundlage für die Planung wichtiger Bahnlinien konstruiert. Im Allgemeinen können skandinavische Länder in diesem Kontext als Pioniere im Bereich der Rasterabbildungen angesehen werden. (vgl. KAMINGER 2010a: 237; KAMINGER 2010b)

In weiterer Folge wurden diverse exemplarische Arbeiten in Schweden, Finnland und Schweiz durchgeführt. Während in der Schweiz früher die Gemeinden die kleinste räumliche Bezugseinheit darstellten, werden seit 1970 auch Daten auf Basis von 1 ha-Rastern angeboten (vgl. PRINZ et al. 2004: 557). 1970 wurden ebenfalls die Ergebnisse des Zensus in Finnland erstmals auf einem 1 km-Raster dargestellt. 1995 wurde eine Kooperation zwischen Schweden und Finnland beschlossen, welche die weitere Forschung bezüglich Karten auf Rasterbasis beinhaltete. Aus dieser Zusammenarbeit entstand 1998 das *Nordic Forum of Geostatistics*. Der im Jahr 2003 abgehaltene *Workshop on European Reference Grids* in Ispra kann als Meilenstein betrachtet werden, da ein europäisches Referenz-Rastersystem definiert wurde. 2006 wurde der *European Grid Club* gegründet, der später in *European Forum for Geostatistics (EFGS)* umbenannt wurde. Mit der *INSPIRE-Richtlinie* aus dem Jahr 2007 wurde festgelegt, dass in Zukunft umweltrelevante Daten sowie demographische und andere statistische Daten auf Basis von Rastersystemen zur Verfügung gestellt werden müssen. Damit wurde die große Bedeutung von rasterbasierten Statistiken für die Kartographie endgültig fixiert. (vgl. KAMINGER 2010a: 237; KAMINGER 2010b)



Quelle: ÖBERG und SPRINGFELDT 1991

Abb. 22: Bevölkerungsdichtekarte Südschwedens 1856

5.2 Administrative vs. geometrische Gebietsgliederungen

In diesem Abschnitt der Arbeit soll aufgezeigt werden, warum rasterbasierte Darstellungen in Zukunft eine wichtige Rolle in der Kartographie spielen könnten. Es werden der Unterschied bzw. die Vor- und Nachteile rasterbasierter Karten gegenüber konventionellen Karten basierend auf administrativen Gebietsgliederungen ausgewiesen. Vor allem bei länderübergreifenden kartographischen Darstellungen eines Themas gab es bisher oft Probleme in Bezug auf die Vergleichbarkeit, da sich die administrativen Verwaltungseinheiten der Länder teilweise stark unterscheiden.

Administrative Gebietsgliederungen werden in Europa vorwiegend durch NUTS- bzw. LAU-Einheiten beschrieben. Diese Bezeichnungen beschreiben „*eine hierarchisch gegliederte Systematik der Gebietseinheiten für die Statistik, die schon vor Jahren von EUROSTAT in Zusammenarbeit mit den Mitgliedsstaaten etabliert wurde [...]*“ (STATISTIK AUSTRIA 2013a). Auf österreichischer Ebene zählen dazu etwa Bundesländer, politische Bezirke, Gemeinden und Zählsprenge. Geometrische Gebietsgliederungen dagegen sind gleichbleibend große, meist quadratische Bezugsflächen, die über ein Gebiet gelegt werden. Darunter können i.d.R. Rasternetze verstanden werden, deren einzelne Maschen einheitliche Kantenlängen aufweisen.

Traditionell wurden offizielle statistische Auswertungen auf Basis administrativer Einheiten durchgeführt (vgl. KAMINGER et al. 2010: 5). Politische Verwaltungseinheiten, wie Gemeinden sind wichtige Bezugsflächen für die örtliche und überörtliche Raumplanung. Um regionalstatistische Unterschiede in einer Karte aufzuzeigen, sind Gebietsgliederungen auf Basis von Gemeindegrenzen jedoch ungeeignet. Die Unterteilung der Gemeinden in Zählsprenge kann dieses Problem nicht entschärfen, da Flächenunterschiede noch schwerwiegender ins Gewicht fallen und somit räumliche Zusammenhänge teilweise zerstört werden würden. (vgl. WONKA 2004: 346ff)

Da seit geraumer Zeit in vielen Ländern Europas statistische Daten in koordinatengebundenen Datenbanken gespeichert werden, ist es möglich Statistiken auf Basis von einzelnen Gebäuden zu erstellen. Auf Grund von Datenschutzbestimmungen und regionalstatistischen Überlegungen müssen diese jedoch zu Gebietseinheiten zusammengefasst werden. Dadurch können Daten auf Basis geometrischer Gebietsgliederungen dargestellt werden. (vgl. ebd.: 346)

Administrative Grenzziehungen spiegeln die Verwaltungszuständigkeiten wider, repräsentieren jedoch nicht direkt einen statistischen Sachverhalt. Da diese Grenzen festgeschrieben sind, kann deren Einfluss auf die Verteilung der Werte nicht ermittelt werden. Durch gleich große und regelmäßige Gebietseinheiten, wie etwa dem Raster, können diesbezüglich Fehlinterpretationen vermieden werden. Allerdings besteht vor allem bei großmaschigen Rasterzellen der Nachteil, dass gewisse zusammenhängende Gebietseinheiten, wie Siedlungen, zerschnitten werden. (vgl. WONKA 2010: 7)

Demgegenüber weisen geometrischer Raster im Vergleich zur Darstellung auf Basis administrativer Einheiten folgende Vorteile auf (vgl. KAMINGER u. WONKA, o.J.: 1):

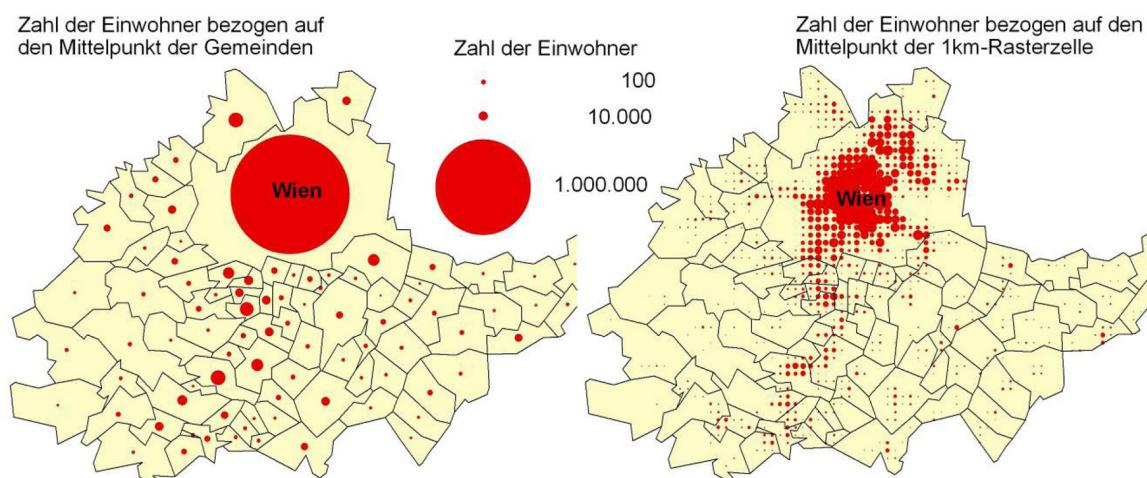
- *Rastereinheiten ermöglichen durch die einheitliche Flächengröße und Form neutrale Raumgliederungen und bessere flächenmäßige Vergleiche*
- *Da sich die statistischen Daten auf gleiche Flächen beziehen sind Absolutwerte gleichzeitig auch Dichtewerte.*
- *Bei Rasterzellen gleicher Dichteklasse kann die trennende Zwischenlinie weggelassen werden um einheitliche Dichtengebiete zu erhalten.*
- *Da bei Rastern keine Grenzänderungen auftreten können zeitliche Veränderungen besser festgestellt werden.*
- *Es können räumliche Analysen durchgeführt werden die auf Basis von Gebietsgliederungen gleicher Flächengröße und Form aufgebaut sind.*

5.2.1 Bezugflächen

Da bereits viele statistische Daten in koordinatengebundener Form vorliegen bzw. Daten auf Basis statistischer Raster in verschiedenen großen Rasterzellen angeboten werden, kann die Rastergröße dem Kartenmaßstab verhältnismäßig leicht angepasst werden (vgl. WONKA 2010: 7)

Laut KELNHOFER (1971: 71) eignen sich Raster besonders gut als statistische Bezugsfäche, da der bei derartigen geometrischen Netzen auftretende Lagefehler im Gegensatz zu den weit größeren statistischen Bezugsfächen bekannt ist. Dieser entspricht der Maschengröße, um die sich die Werte bei willkürlicher Verschiebung des Netzes ändern können.

Neben kartographischen Relativwertdarstellungen können statistische Verteilungen auch anhand deren Absolutwerten abgebildet werden. Relativwerte werden üblicherweise durch Farbabstufungen dargestellt, hingegen Absolutwerte durch gestufte oder flächenproportionale Kreise oder Quadrate (siehe Abb. 23). Dies kann allerdings zu Problemen führen, wenn sich die absoluten Werte auf Flächen unterschiedlicher Größe und Form beziehen (vgl. WONKA 2010: 9f). In Abb. 23 ist die Zahl der Bevölkerung auf Basis von Gemeinden sowie auf Basis von 1 km-Rasterzellen dargestellt. Auch dieses Beispiel verdeutlicht die Unterschiede der beiden Abbildungen. Während in der linken Karte keine signifikante Verteilung zu erkennen ist, fällt bei Betrachtung der rechten Karte auf, dass die Zahl der Einwohner sowohl entlang der Südbahn als auch des südlichen Wiener Speckgürtels höher ist als in anderen Gebieten des Wiener Umlandes. Dieser Unterschied entsteht trotz gleicher statistischer Ausgangsdaten.



Quelle: WONKA 2010: 10

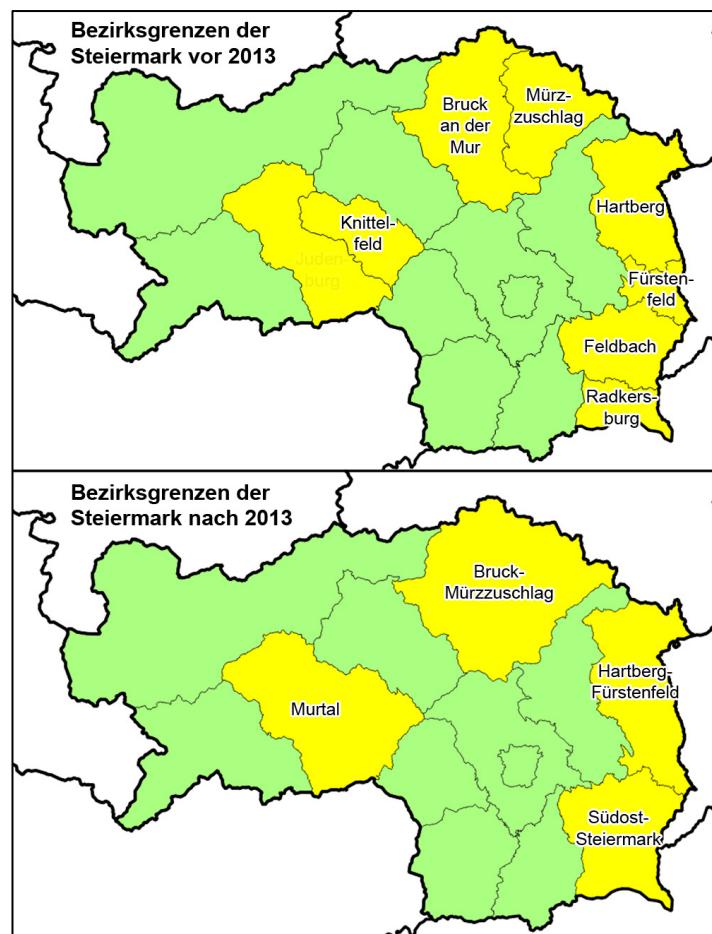
Abb. 23: Absolute Einwohnerzahlen für das Wiener Umland

5.2.2 Zeitliche Vergleichbarkeit

Administrative Grenzen auf unterschiedlichen Verwaltungsebenen sind nicht in Stein gemeißelt. Neben Gemeinden und Zählsprenge können auch Staatsgrenzen neu festgelegt

werden. Dadurch werden Vergleiche von statistischen Verteilungen bei Zeitreihenanalysen erschwert. Bei Änderungen der Grenzziehungen müssen die statistischen Werte auf die neue Bezugsfläche umgelegt werden, wobei deren Aussagekraft schwinden kann. Diese Problematik ist hingegen bei koordinatengebundenen Rasterdarstellungen nicht gegeben, weswegen exakte temporale Vergleiche für jede Rasterzelle durchgeführt werden können. (vgl. WONKA 2010: 13)

Abb. 24 zeigt ein Problem, welches sich in Österreich durch die Gemeindestrukturreform in der Steiermark ergibt, im Zuge dessen Gemeinden und Bezirke zusammengelegt werden. Durch diese Vergrößerung bzw. Veränderung der administrativen Einheiten ist einerseits eine detaillierte Betrachtung des Gebietes nicht mehr möglich, andererseits ändern sich dadurch die Grenzen, wodurch eine zeitliche Vergleichbarkeit erschwert wird. Jedoch nicht nur größere Reformen erschweren die Darstellung statistischer Daten. Auch kleinere Grundstückszusammenlegungen, Kommassierungen oder geodätische Neuvermessungen können administrative Grenzen, wie Zählsprenge oder Gemeinden, im Laufe der Zeit verändern.



Quelle: ILLMEYER 2014

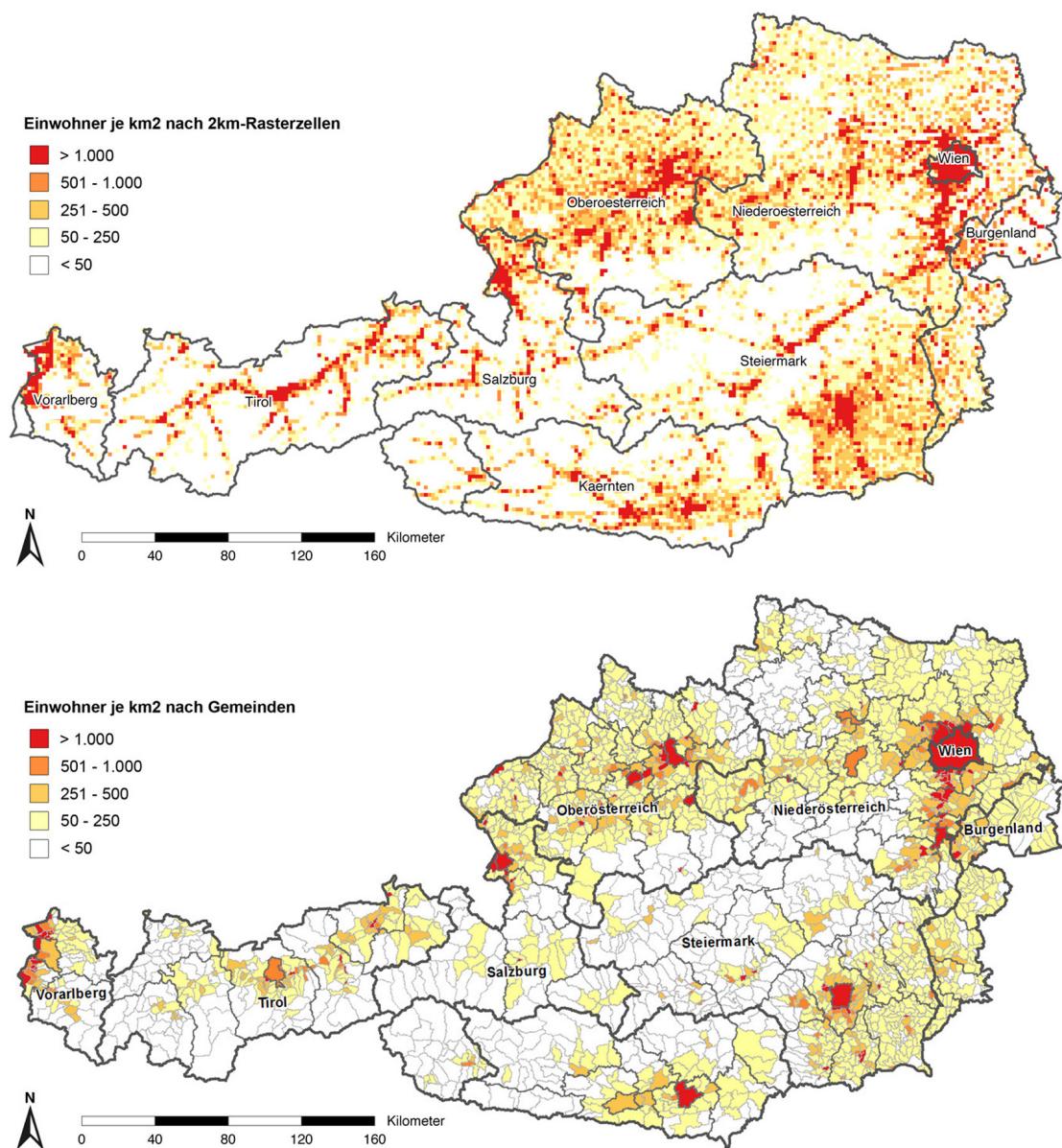
Abb. 24: Stand der Bezirksgrenzen Steiermark vor und nach der Zusammenlegung 2013

5.2.3 Darstellungen auf Länderebene

Die Wahl räumlicher Gebietsgliederungen wirkt sich erheblich auf die Vergleichbarkeit statistischer Daten aus. Dies kann am Beispiel von Österreich gut gezeigt werden (siehe Abb. 25). Der zugrundeliegende Datensatz stammt aus den Daten des EFGS, mit denen die 1km-Rasterabbildung der Bevölkerung Europas hergestellt wurde. Diese Daten wurden mit einer gewissen Streuung interpoliert, so dass eine beinahe der Realität entsprechende Bevölkerungsverteilung wiedergegeben werden kann.

Die obere Karte wurde bereits mit Hilfe der im Rahmen dieser Arbeit entworfenen Webanwendung erstellt. Dazu wurde der interpolierte Punktdatensatz, der geographische Koordinaten enthält, in die Applikation eingespielt und die entsprechenden Klassifizierungsparameter eingestellt. Als Farbverlauf wurde ein einpoliger Verlauf von Gelb (bzw. Weiß) nach Rot gewählt. Die untere Karte wurde mit dem gleichen Punktdatensatz auf Basis der vom BEV frei zur Verfügung gestellten Verwaltungsgrenzen erstellt. Um die Vergleichbarkeit der beiden Karten zu gewährleisten, wurden gleiche Klassengrenzen, Maßstäbe und Farbverläufe verwendet.

Anhand dieses Beispiels ist zu beobachten, dass bei Verwendung von 2km-Rasterzellen kleinräumige Verteilungen besser erkennbar sind als bei der Darstellung in Bezug auf Gemeindeebene. Des Weiteren kann die Besiedelung der Täler bei der rasterbasierten Karte besser nachvollzogen werden.



Quelle: ILLMEYER 2014

Abb. 25: Bevölkerungsdichtekarte Österreichs mit den Bezugsebenen Gemeinden und 2 km-Raster

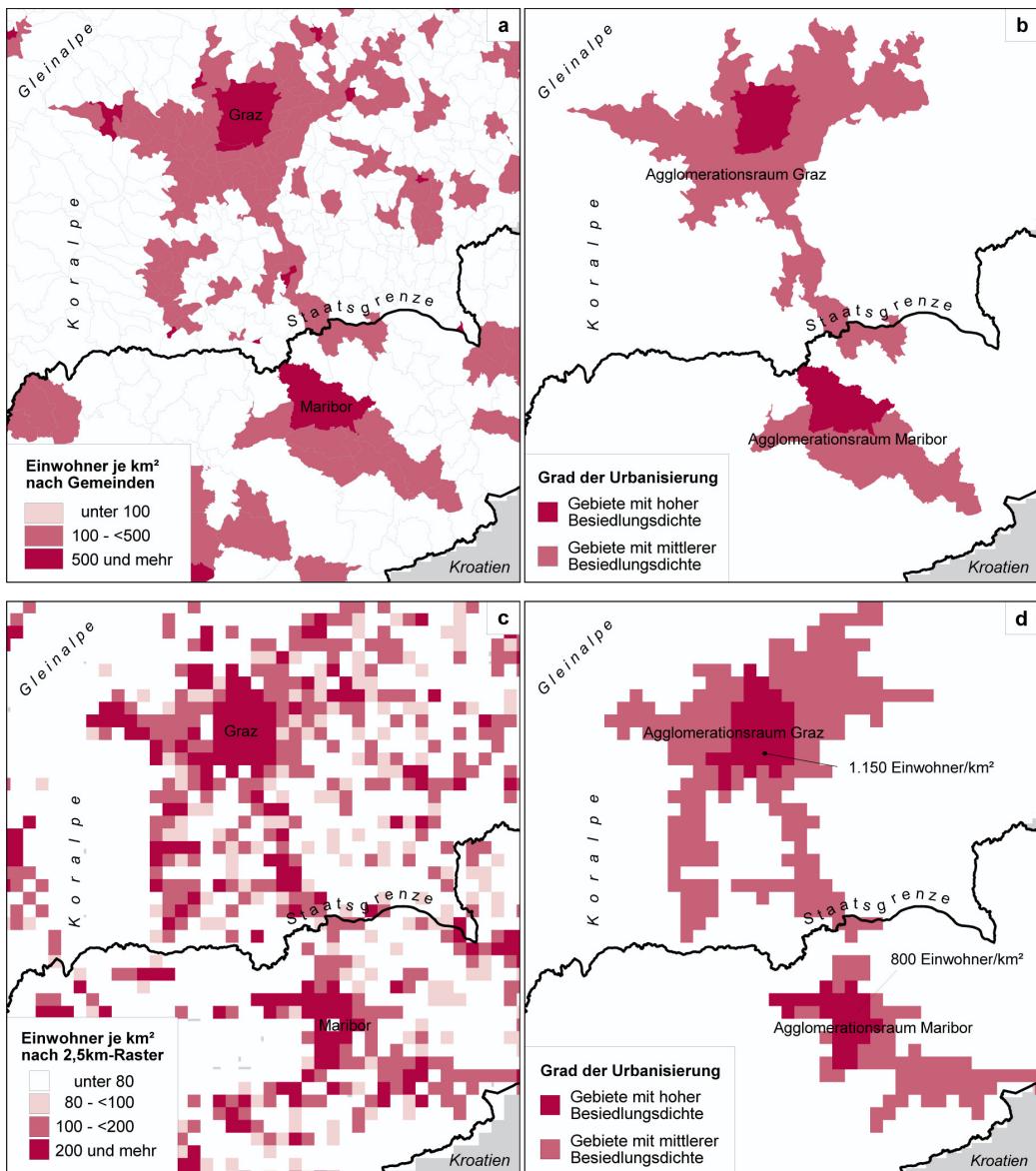
5.2.4 Länderübergreifende Darstellungen

Die Qualität von grenzübergreifenden Abbildungen hängt nicht nur von der Vergleichbarkeit statistischer Indikatoren ab, sondern auch davon, in welcher Weise statistische Daten regionalisiert sind. Da planerische Entscheidungen auf regionaler Ebene möglichst flexible räumliche Ansätze benötigen, eignen sich dafür vor allem rasterbasierte Bezugsflächen. Die Inhomogenität der NUTS- und LAU-Einheiten kann in diesem Zusammenhang als ein Hauptproblem bei statistischen Analysen angesehen werden. Diese Unterschiede vergrößern sich noch bei grenzübergreifenden Darstellungen. Statistische Daten sind

hingegen nur dann regional und zeitlich miteinander vergleichbar, wenn diese auf konstanten und koordinativ gebundenen Gebietsgliederungen, wie etwa regionalstatistischen Rastern, abgebildet werden. (vgl. WONKA et al. 2007: 72ff)

Bei der Darstellung länderübergreifender Thematiken, ergeben sich vor allem dann Probleme, wenn administrative Einheiten als Bezugsfläche gewählt werden. Größere Flächenunterschiede können bereits innerhalb eines Landes vorliegen, wodurch Fehlinterpretationen gemacht werden können. Dieses Problem kann verschärft werden, wenn vergleichbare Gebietsgliederungen einzelner Länder gegenübergestellt werden. Bspw. beträgt die durchschnittliche Größe einer lokalen Einheit (Gemeinde bzw. LAU2) in Frankreich 15 km², eine Fläche, auf der im Schnitt 1.500 Einwohner leben. Im Vergleich dazu weist die durchschnittliche lokale Einheit in Dänemark 156 km² und 19.000 Einwohner auf. (vgl. ebd.: 18)

Beim Vergleich des Grenzgebietes zwischen Slowenien und Österreich, abgebildet auf Verwaltungseinheiten und regionalstatistische Rastern, ist der Unterschied klar erkennbar. Abb. 26 Abb. 26: (a+b) stellt das Ergebnis der Abgrenzung nach LAU-2 Einheiten dar. Die Abgrenzung nach 2,5 km-Rastern (Abb. 26 c+d) zeigt, dass genauere Ergebnissen erzielt werden können. Allein die Tatsache, dass die slowenischen LAU-2 Einheiten größer sind als die österreichischen, lässt keine genauen Vergleiche der Agglomerationsräume zu. Einerseits gehen dicht besiedelte Gebiete auf Grund der Größe einer Verwaltungseinheit verloren, andererseits liegen dicht besiedelte Gebiete laut Darstellung im Ödland.



Quelle: bearbeitet nach WONKA et al. 2007: 77

Abb. 26: Bevölkerungsdichte im Grenzgebiet Österreich/Slowenien

5.3 Zuordnung statistischer Daten zu Rastern

Ein wichtiger Schritt bei der Umsetzung rasterbasierter Karten ist die Zuordnung statistischer Daten zu den zugehörigen Rasterzellen. Da neben punktbasierten Daten (z.B. Bevölkerungszahlen) auch flächenhafte und linienbezogene Elemente in ein Rasternetz überführt werden können, müssen diese Transformationsprozesse näher erläutert werden. Derartige Daten können z.B. aus Geländeerhebungen anhand von Luft- bzw. Satellitenbildern gewonnen werden. In der im Rahmen dieser Arbeit erstellten Webapplikation ist zwar nur die Zuordnung von Punktdaten zu Rastern interaktiv möglich, um die Vollständigkeit zu gewährleisten, werden jedoch die Möglichkeiten für flächen- und linienhafte Elemente ebenfalls aufgezeigt. Diese Zuordnung könnte beispielsweise in einem vorgelagerten Verarbeitungsprozess durchgeführt werden.

5.3.1 Zuordnung von Punktdaten zu Rasterzellen

Für die Erhebung von Punktdaten mit bekannten Koordinaten kann eine Aggregation der Daten auf Basis einer bestimmten Projektion des Rasternetzes durchgeführt werden. Sind Daten ausschließlich auf Basis eines größeren Bezugsgebietes verfügbar, müssen Disaggregationsmethoden angewandt werden.

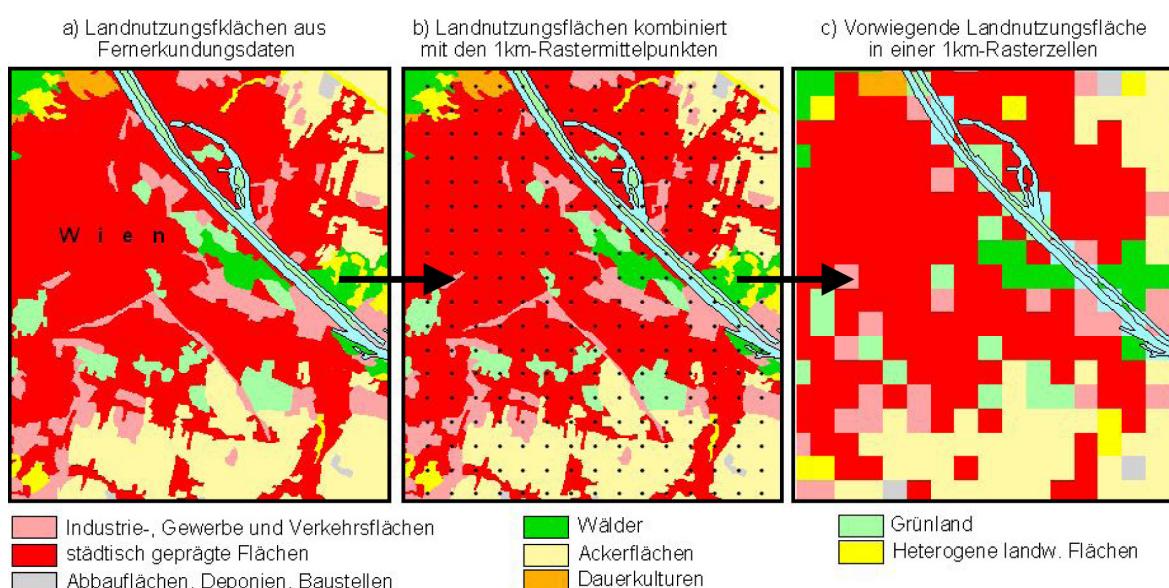
Sind geokodierte Daten auf Basis von Gebäuden oder Adressen verfügbar, können diese aggregiert werden. Da die Koordinaten, z.B. beim Eingang, mitten im Haus oder irgendwo im Grundstück liegen können, ist der maximale Lagefehler auf die breiteste Stelle des Grundstückes begrenzt. Die Gebäudekoordinaten werden dann mit der zugehörigen Rasterzelle verbunden und über diese aggregiert. (vgl. KAMINGER 2010a: 241)

Als Disaggregation wird die Aufschlüsselung von statistischen Daten nach bestimmten Merkmalen bezeichnet. Es ist eine Top-Down-Methode, da bspw. von allgemeinen Bevölkerungsdaten auf kleine Gebiete herunter gerechnet wird. Bei der Disaggregation werden Daten größerer Gebiete durch zusätzliche Informationen auf kleinere Gebiete aufgeteilt. (vgl. ebd.: 242f; KAMINGER 2010b: 10ff)

5.3.2 Zuordnung von Flächendaten zu Rasterzellen

Die Zuordnung von Polygonflächen zu Rastern kann auf verschiedene Art und Weise erfolgen, was im Folgenden anhand des Beispiels einer Landnutzungskategorie aus CORINE gezeigt wird.

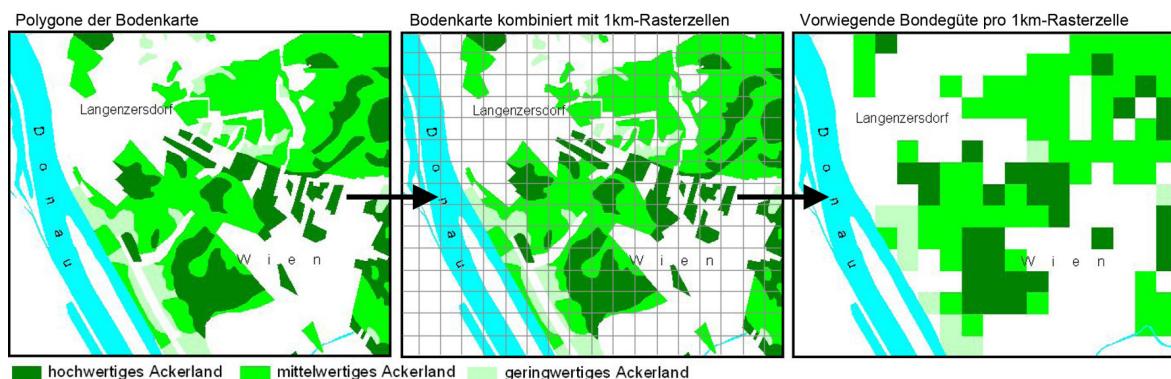
Die schnellste und einfachste Methode zur Zuordnung von Polygonen zu einer Rasterzelle erfolgt über deren Mittelpunkte (siehe Abb. 27). Rastermittelpunkte werden dabei mit den Landnutzungspolygonen verknüpft. Die Zuordnung einer Kategorie zu einer Rasterzelle erfolgt nur dann, wenn der Rastermittelpunkt in der jeweiligen Landnutzungskategorie liegt. (vgl. WONKA 2010: 49f)



Quelle: WONKA 2010: 50

Abb. 27: Zuordnung von Polygonflächen zu Rastern über Rasterzellenmittelpunkt

Um genauere Ergebnisse zu generieren, müssen die Zuordnungen über die Landnutzungsflächen erfolgen. Dabei erhält jede Rasterzelle nur jenes Merkmal, des in ihr liegenden größten Flächenanteils (vgl. ebd.: 51). Abb. 28 Abb. 28: zeigt in diesem Zusammenhang drei Bodenkategorien, für die jeweils die Fläche pro Rasterzelle berechnet werden muss.

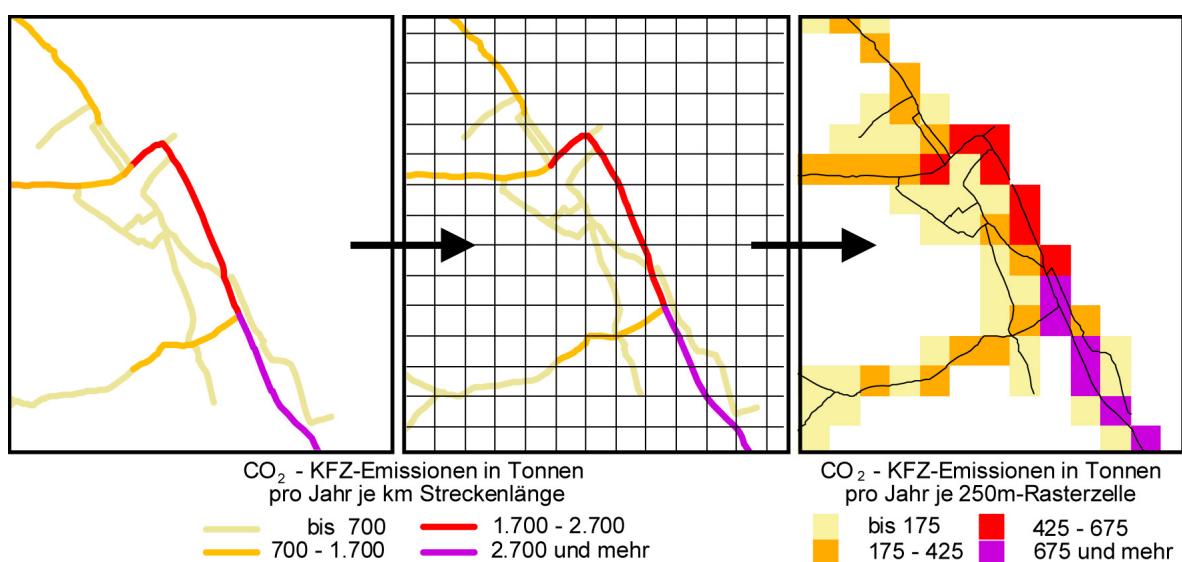


Quelle: bearbeitet nach WONKA 2010: 51

Abb. 28: Zuordnung von Polygonflächen zu Rastern über Rasterzellenschwerpunkte

5.3.3 Zuordnung von linienbezogenen Daten zu Rasterzellen

Auch linienhaft in Erscheinung tretende Thematiken können auf Rasterkarten umgelegt werden. Dies wird hierbei am Beispiel einer Straße mit zugehöriger CO₂-KFZ-Belastung pro Kilometer demonstriert (siehe Abb. 29). Es werden hierfür die Linien mit den Rasterzellen verschnitten sowie die CO₂-Belastung mit den in der jeweiligen Rasterzelle liegenden Straßenkilometern multipliziert. Abschließend werden die Belastungen für jeder Zelle aufsummiert und in der Karte dargestellt (vgl. ebd.: 98).



Quelle: WONKA 2010: 51

Abb. 29: Zuordnung von linienhaften Daten zu Rasterzellen

5.4 Darstellungsarten und kartographische Aspekte

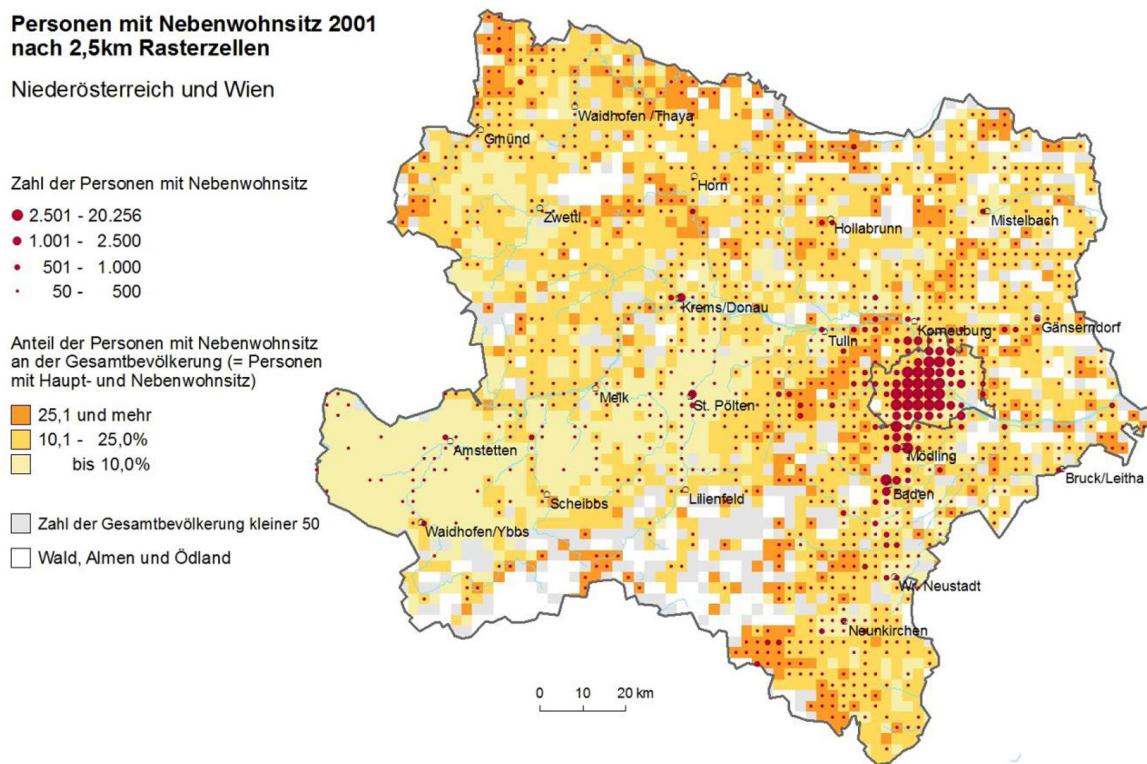
Eine der Aufgaben der Kartographie ist es, statistische Daten in möglichst übersichtlicher Form darzustellen. Die wesentlichen Informationen sollen schnell und verständlich aufgenommen werden können, sodass ein leichter Vergleich von Objekten ermöglicht wird. Der Betrachter der Karte soll Aussagen über die Raumbezogenheit und das Nebeneinander von Sachverhalten treffen können (vgl. WONKA 2006: 49f). Die kartographischen Grundlagen zur Erstellung von thematischen Karten wurde bereits in Kapitel 2 diskutiert. In diesem Abschnitt sollen nun die spezifischen Charakteristika von rasterbasierten Darstellungen, welche bereits aus den vorigen Kapiteln bekannt sind, nochmals zusammengefasst werden.

Prinzipiell kann bei Darstellungen auf Basis regionalstatistischer Raster zwischen flächenhaften Relativwert- und Absolutwertdarstellungen mit gestuften oder flächenproportionalen Kreisen oder Quadraten unterschieden werden. Mehrschichtige thematische Rasterkarten können des Weiteren eine Kombination beider Abbildungsarten beinhalten.

Relativwertdarstellungen mit Farbabstufungen (Farbton, Sättigung und Helligkeit) werden verwendet, um Daten flächendeckend auf einem bestimmten Gebiet darzustellen. Die absoluten Zahlen jeder Rasterzelle (z.B. Bevölkerungszahlen, Kunden, etc.) werden dafür in Relation zu einem Gebiet oder ähnlichem Faktor gesetzt (z.B. Bevölkerung pro Quadratkilometer). Dafür wird eine, am besten für sich selbst sprechende, Farbskala verwendet, die eine bestimmte Verteilung in einer Region verdeutlichen soll. Die einzelnen Rasterzellen nehmen dabei unterschiedliche Farbwerte an, wobei eine Anzahl von 7-8 Klassen nicht überschritten werden sollte, damit die Übersichtlichkeit und eine eindeutige Zuordnung gewährleistet werden kann. Rasterzellen, die keine oder zu wenig Merkmalsausprägungen beinhalten, werden meist ohne Flächenfärbung dargestellt. Hinsichtlich der rasterbasierten Darstellungen überwiegen Relativwertdarstellungen.

Absolutwertdarstellungen werden, wie bereits erwähnt, meist durch gestufte oder flächenproportionale Kreise, Quadrate oder mittels anderer geometrische Formen dargestellt. Die geometrischen Figuren werden dafür im Mittelpunkt einer Rasterzelle oder seltener im Schwerpunkt der Verteilung einer Rasterzelle, platziert. Bei einem Großteil der rasterbasierten Absolutwertdarstellungen werden die Punktsignaturen, entsprechend der zugrundeliegenden Verteilung, nach ihrer Größe variiert. Seltener sind Variationen auf Basis unterschiedlicher Sättigung oder Helligkeit.

Rasterbasierte Absolutwert- und Relativwertdarstellungen können kombiniert in einer mehrschichtigen thematischen Karte verwendet werden. Abb. 30 zeigt in diesem Zusammenhang eine Karte von Nebenwohnsitzen in Niederösterreich auf Basis 2,5 km-Rasterzellen. Die punktförmigen Signaturen stellen die absoluten Zahlen der Personen mit Nebenwohnsitz dar, wohingegen die Flächenfärbung den proportionalen Anteil der Personen mit Nebenwohnsitz an der Gesamtbevölkerung repräsentiert.



Quelle: bearbeitet nach WONKA 2010: 94

Abb. 30: Nebenwohnsitze 2001 in NÖ und Wien auf Basis 2,5km-Rasterzellen

Bei der Erstellung von kartographischen Produkten auf Basis regionalstatistischer Raster ist die Wahl der topographischen Grundlage (Grundkarte) ebenfalls von fundamentaler Bedeutung, da dadurch dem statistischen Inhalt ein räumlicher Bezug zugewiesen wird. Wie eine Grundkarte für thematische Darstellung gestaltet werden kann und welche Elemente sie beinhalten kann wurde bereits im Kapitel 2.2.1 diskutiert.

In Abhängigkeit des Kartenmaßstabes muss die Rastergröße bei rasterbasierten Darstellungen gewählt werden. Bei kleinen Rastermaschen werden Verteilungen besser wiedergegeben, hingegen kommt es bei grobmaschigen Rastergrößen zur Glättung der Daten. In diesem Zusammenhang ist eine Rastergröße zu wählen, die Verteilungen realitätsnah widerspiegelt und gleichzeitig Signaturen noch gut erkennbar sind. (vgl. WONKA 2004: 350ff; KAMINGER und WONKA 2005: 759f)

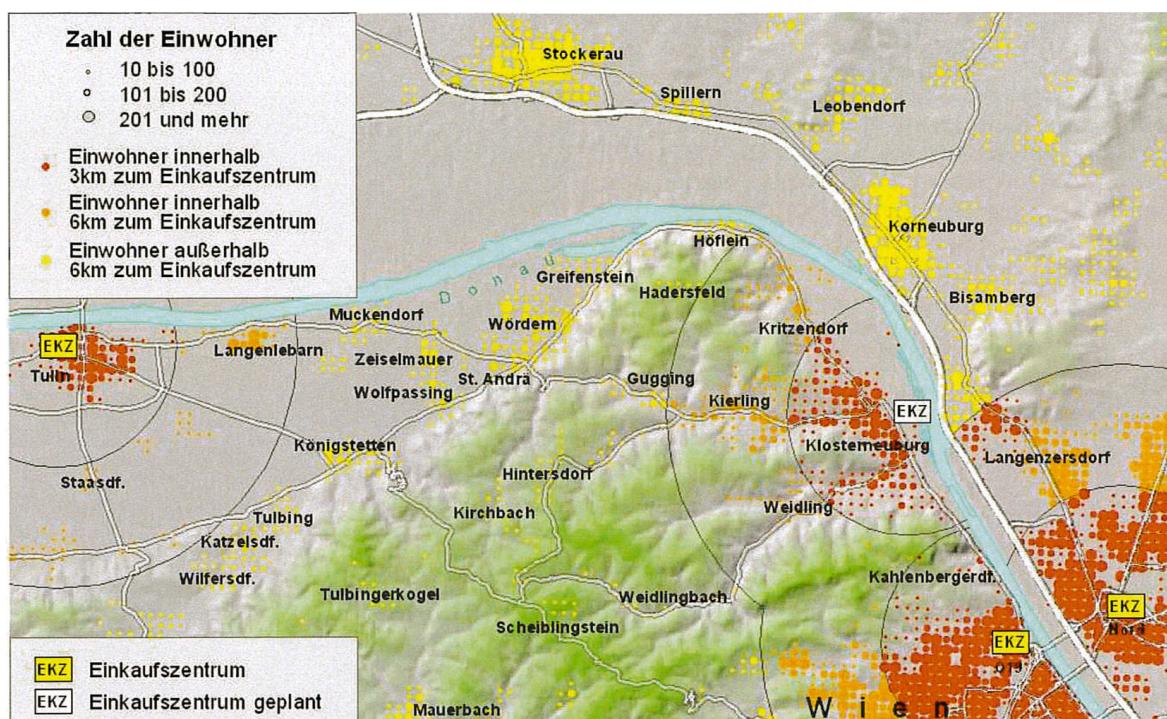
5.5 Anwendungsbereiche rasterbasierter Karten

Es gibt vielfältige Einsatzmöglichkeiten für thematische Karten auf Rasterbasis. Einige Möglichkeiten und Darstellungen wurden bereits in den vorherigen Kapiteln bzw. in den bisher dargestellten Abbildungen erläutert und finden sich oft in der Regionalstatistik wieder. In diesem Abschnitt der Arbeit erfolgt eine Beschreibung von zwei weiteren wichtigen Anwendungsbereichen. Neben der Infrastruktur- und Verkehrsplanung spielen detaillierte rasterbasierte Karten im Bereich des Geomarketing eine große Rolle. Des Weiteren können diese auch für GIS-Analysen herangezogen werden. Die in dieser Arbeit

entwickelte Applikation ist so entworfen, dass sie im Umfeld dieser Anwendungsbereiche zukünftig zum Einsatz kommen könnte. Vor allem im Geomarketing kann ein großer Interessenskreis identifiziert werden.

5.5.1 Infrastrukturplanung und Verkehrsplanung

Obwohl die Infrastruktur- und Verkehrsplanung der Raumplanung zugeordnet werden kann, werden diese Bereiche im Rahmen dieser Arbeit aufgrund ihrer großen Bedeutung gesondert behandelt. Um Infrastrukturleistungen befriedigen zu können, müssen Distanzen überwunden werden. Dieser Umstand wird in Erreichbarkeitsmodellen abgebildet, dessen Ergebnis meist Distanztabellen bzw. Distanzmatrizen darstellen. Abb. 31 zeigt bspw., wie viele Einwohner innerhalb einer fußläufigen Distanz zu einem Supermarkt wohnen. Für diese Fragestellung (fußläufige Entfernung ca. 300 m) müssen Einwohnerzahlen auf 100 m-Raster abgebildet werden, um zu angemessenen Ergebnissen zu gelangen. Da sich die Einwohnerzahlen auf die Rastermittelpunkte beziehen, ist der maximal mögliche Lagefehler gering (höchstens die Hälfte der Diagonale der Rasterzelle). (vgl. WONKA 2011: 59f).



Quelle: Wonka 2011: 83

Abb. 31: Fußläufige Distanz zu Supermärkten

Ausgedehnte Siedlungsgebiete sowie flächenintensive Siedlungsformen führen zu hohem Verkehrsaufkommen. Nicht nur hohe Pendleranteile sowie ein hoher Anteil an Zweitwohnsitzen, sondern auch eine mangelnde Versorgung mit Geschäften des täglichen Bedarfs, führen dazu, dass lange Wege zurückgelegt werden müssen. Die Verkehrspolitik ist somit überaus bedeutsam für die Entwicklung des Individualverkehrs. Die Erstellung

von Verkehrskonzepten, mit dem Ziel den motorisierten Individualverkehr auf umweltfreundliche Verkehrsmittel zu verlagern, ist somit notwendig (vgl. WONKA 2006: 137). Um diese Probleme in der Verkehrsplanung zu erkennen und eine geeignete Planung durchzuführen, können rasterbasierte Darstellung ebenfalls als Grundlage dienen.

5.5.2 Geomarketing

GIS-Tools werden immer häufiger im Bereich Marketing und Vertrieb angewendet. Die rasche Veränderung der Märkte, die Globalisierung des Wettbewerbs, Konzentrationsprozesse in vielen Wirtschaftszweigen und die sinkende Nachfrage mit verstärktem Wettbewerbs- und Kostendruck verlangen einen hohen Grad an Flexibilität. Dafür werden zielgerichtete Informationen über Kunden und Märkte benötigen. Diese Umstände führten zur Entwicklung entsprechender Geomarketingssystemen, die im Wesentlichen aus folgenden Komponenten bestehen: (vgl. WONKA 2006: 145).

- *Geodaten: Bilden die Kartengrundlage und sind in ein Koordinatensystem eingebettet.*
- *Statistische Daten: Sind in der Karte abzubilden bzw. zu analysieren.*
- *Software: Stellt die Verknüpfung von Geodaten zu statistischen Daten her. Ermöglicht die Durchführung räumlicher Analysen.*

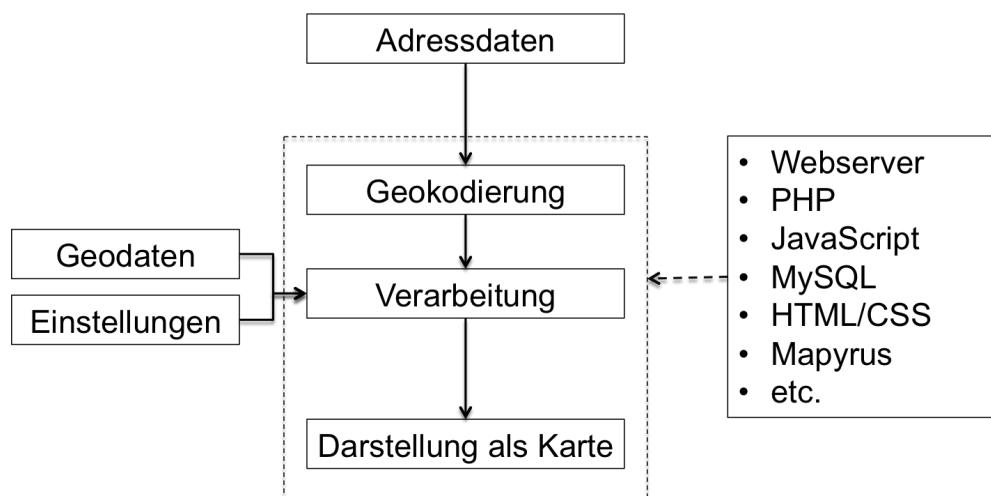
Mit Hilfe der Darstellung statistischer Daten in einer rasterbasierten Karte ist es möglich, Zusammenhänge und Hintergründe von Vertriebsergebnissen transparent zu gestalten. Eine Voraussetzung um räumliche Zusammenhänge zu erkennen ist, dass die Daten entsprechend räumlich detailliert vorliegen. Bei der Interpretation der Daten muss beachtet werden, dass es einen Unterschied zwischen den ursprünglichen (z.B. Einwohnerzahlen) und abgeleiteten (z.B. Marktpotential, Kaufkraft) Daten gibt. Abgeleitete Daten werden durch unterschiedliche Verarbeitungsprozesse bzw. Rechenoperationen erstellt. Ein Problem hierbei ist, dass die Berechnungsprozeduren oft nicht nachvollzogen werden können und somit die Eignung für bestimmte Probleme nicht validiert werden kann. Um betriebsinterne Daten mit externen Sachdaten in Verbindung zu bringen, müssen Kundenadressen geokodiert und mit den Sachdaten in den entsprechenden Rasterzellen dargestellt werden. (vgl. ebd.: 145f)

6 Erstellung eines Prototypen der Online-Applikation

Um die gesammelten Informationen in einen praktischen Kontext zu setzen, wurde ein Prototyp entworfen, der den Prozess der Geokodierung von Adressdaten sowie die Erstellung von Karten, basierend auf regionalstatistischen Rastern im Internet zulässt. Dafür wurde eine Webplattform entwickelt, in der diese Maßnahmen möglich sind. An dieser Stelle muss jedoch festgehalten werden, dass es sich hierbei um einen ersten Entwurf handelt, der die geforderten Funktionalitäten erfüllt. Daher werden auch Verbesserungspotentiale bzw. Erweiterungsmöglichkeiten identifiziert, die im Kapitel 6.3 beschrieben werden.

6.1 Konzept und Idee

Die grundlegende Idee dieser Webplattform ist, den Umwandlungsprozess von Adressdaten in Koordinaten sowie die darauf basierende Erstellung von rasterbasierten statischen Karten innerhalb einer Anwendung zu integrieren. Abb. 32 bietet in diesem Zusammenhang einen Überblick über die Elemente der Webplattform und der dafür benötigte Inputs und Techniken.



Quelle: ILLMEYER, 2013

Abb. 32: Überblick der Webapplikation

Derzeit gibt es zwar bereits einige Webplattformen, mit denen Karten auf Basis administrativer Einheiten erstellt werden können, jedoch existieren noch keine kostenfreien Anwendungen, die sich auf die Möglichkeit der kartographischen Darstellung von Themen auf Basis von gleichmäßigen Rastern stützen und gleichzeitig eine Umwandlung von Adressen in geographische Koordinaten ermöglichen. Da der Prozess der Geokodierung von Adressdaten im Rahmen dieser Webplattform vorgelagert ist, können relativ einfach und rasch Karten für statistische Auswertungen, wie bspw.

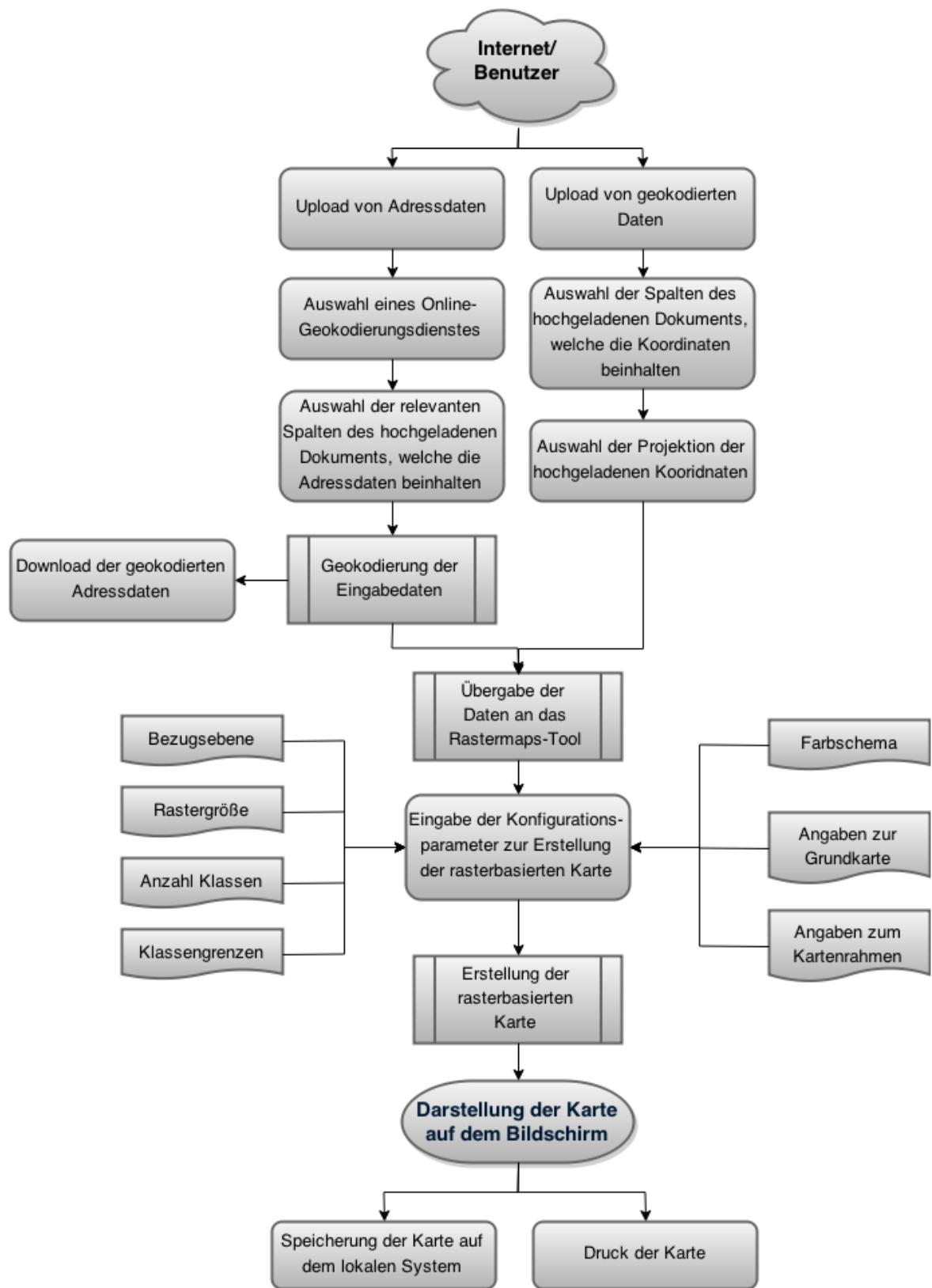
Kundendichten, Bevölkerungsdichten, etc. erstellt werden. In diesem Fall wird lediglich eine Tabelle mit den Adressen von Kunden bzw. Adressen von Standorten etc. benötigt, auf Basis derer eine Karte mit Punktdichten in Rasterzellen berechnet und dargestellt wird.

Vorrangiges Ziel dieser Umsetzung ist es, wie bereits erwähnt, einen ersten Prototyp zu entwerfen bzw. das Grundgerüst für spätere Entwicklungen aufzubauen. Das Programm soll als Webanwendung zur Verfügung stehen und einfach auf andere Systeme portierbar sein um sie in beispielsweise unternehmensinternen Netzwerken verwenden zu können.

Der Ablauf zur Erstellung von rasterbasierten Karten verläuft dabei prinzipiell immer nach einem ähnlichen Schema (siehe Abb. 33). Zuerst müssen Adressdaten mit dem integrierten Online-Geokodierungstool in Koordinaten umgewandelt werden. Anschließend werden diese Ergebnisse an das Tool zur Erstellung der rasterbasierten Karten weitergereicht, wo verschiedene Einstellungen getroffen werden können, welche das Aussehen der Karte beeinflussen (siehe Kapitel 6.3). Als letzter Schritt wird die druckbare Karte auf dem Bildschirm angezeigt und steht zum Download oder für den Druck bereit. Falls Daten bereits in geokodierter Form vorliegen, können diese auch gleich hochgeladen und somit der Schritt der Geokodierung übersprungen werden.

Die graphische Benutzeroberfläche (GUI – Graphical User Interface) der Anwendung soll ansprechend und so einfach als möglich gestaltet werden. Die Bedienung der einzelnen Teil-Elemente muss ohne zusätzliche Erklärung durchgeführt werden können. Dies ermöglicht es auch ungeschulten Benutzern die Applikation schnell zu verstehen und benutzen zu können. Die Anwendung wird in einer ersten Version komplett in Englisch verfasst, da sie die Möglichkeiten beinhaltet, Daten auf Basis von Österreich und Europa darzustellen. Um den Benutzer bei der Erstellung der Karte mit kartographischen Symbolisierungskriterien nicht zu überfordern, werden, abhängig von der Klassenanzahl, bereits vordefinierte und aufeinander abgestimmte Farbschemata vorgegeben. Dadurch hat der Benutzer die Möglichkeit aus verschiedenen ein- und zweipoligen Farbskalen zu wählen.

Mit den begrenzten Möglichkeiten des Vektor-Zeichenprogrammes *Mapyrus*, soll unter Beachtung der in den vorigen Kapiteln dieser Arbeit gesammelten Erkenntnissen und den Einstellungen der Benutzer der Webapplikation eine ansprechende rasterbasierte kartographische Darstellung ausgegeben werden. Diese statische Karte soll für den Druck optimiert werden und die Thematik auf Rasterbasis im Vordergrund stehen. Neben den wichtigen Kartenrandangaben wie Titel, Legende und Maßstab sollen in der Karte selbst lediglich Ländergrenzen und die wichtigsten Beschriftungen ersichtlich sein.



Quelle: ILLMEYER, 2015

Abb. 33: Workflow der Webanwendung

6.2 Technische Umsetzung

Die technische Umsetzung der Online-Plattform erfolgt mit Hilfe aktueller Web-Technologien. Für die eigentliche Webanwendung wurden *HTML*, *PHP*, *JavaScript* und *CSS* verwendet. Für die Kartenerstellung wurde das Tool *Mapyrus* herangezogen.

Die Speicherung der Daten erfolgt mit Hilfe einer *MySQL*-Datenbank. In dieser Datenbank werden nicht nur die Ergebnisse eines Geokodierungsvorgangs gespeichert, sondern es sind auch bereits die in der Webanwendung zur Verfügung stehenden Raster gespeichert. Jede Rasterzelle ist somit mit einem eigenen Namen, welcher der offiziellen Benennung der Rasterzellen laut *INSPIRE* entspricht, und den umgebenden Koordinaten gespeichert. Werden Koordinaten nach einem erfolgreichen Geokodierungsvorgang, oder direkt über den Upload, an das Rastermaps-Tool übergeben, werden diese in einer Tabelle in der Datenbank gespeichert. Nachdem der Benutzer die Konfigurationsparameter eingegeben hat, berechnet die Webapplikation wie viele Punkte in jeder Rasterzelle liegen und speichert diese Anzahl ab. Dies wird benötigt um im Anschluss die verschiedenen Flächenfärbungen der Rasterzellen darstellen zu können.

Das für die Erstellung der Karte verwendete Tool *Mapyrus* ist ein Vektor-Zeichenprogramm, welches in *Java* entwickelt wurde um Punkte, Linien und Flächen zu zeichnen. Durch die Erstellung von Skript-Dateien können Ausgaben in Raster- und Vektorformaten generiert werden. Es ist eine Logo Graphik-Sprache mit GIS-Datenleser für Datenbank-Anbindungen, implementiert als unabhängiges Kommandozeilen-Programm z.B. für Web-Server (GISWIKI 2014). Somit ist es möglich, mit *Mapyrus* Karten aus GIS-Daten zu erstellen. Bspw. können Shape-Dateien oder georeferenzierte Rasterdateien einfach in eine Karte integriert werden. Neben der Einbindung vorhandener Geodaten ist es für die Erstellung der Webapplikation ebenfalls wichtig, Daten aus Datenbanken und aus Textdateien einlesen zu können, was *Mapyrus* ebenfalls unterstützt. Neben diesen Eigenschaften unterstützt es außerdem die Gestaltung der Kartenrandelemente wie z.B. die automatische Erstellung einer Legende, Nordpfeil oder die Berechnung der Maßstabsleiste.

Da *Mapyrus* eine *Java*-Applikation ist, muss diese auf dem Webserver im Hintergrund als Dienst laufen. Die Webapplikation erstellt dabei in Echtzeit *Mapyrus*-Skripte, welche die Benutzereingaben und die benötigten Daten beinhalten, und sendet diese per HTTP-Abfrage an die *Mapyrus*-Applikation, welche auf einem bestimmten Port erreichbar ist. Diese Applikation verarbeitet die Inhalte des Skripts und gibt die gezeichnete Karte auf dem Bildschirm wieder.

Das Frontend der Applikation wurde maßgeblich mit *HTML* und *CSS* gestaltet, da diese den aktuellen Standards im Bereich der Entwicklung von Webseiten entsprechen. Teilweise wurden auch bereits Elemente von *HTML5* und *CSS3* in die Seite integriert. Die Übermittlung von Daten an den Server bzw. die Kommunikation zwischen den verschiedenen Unterkategorien der Webplattform wurde mit *PHP* realisiert. Mit *PHP*

lassen sich Variablen von einer Seite zur nächsten übermitteln und z.B. auch *Mapyrus* Skripte in Echtzeit nach Eingaben der Benutzer erstellen. Die dynamischen Inhalte der Anwendung wurden mit Hilfe von *JavaScript* umgesetzt. Darunter fallen z.B. die Überprüfung von Dateiendungen, Einblendung von Lade-Grafiken, die Überprüfung auf richtige Eingaben im Formular oder das dynamische Aus- und Einblenden von Webseiten-inhalten.

Um verschiedene Funktionen der Anwendung leichter nutzen bzw. entwickeln zu können, wurden einige externe *JavaScript*- und *PHP*-Bibliotheken verwendet. Diese werden hier aufgezählt sowie deren Verwendung erläutert.

Um Tabellenkalkulations-Formate wie *Excel*, *CSV*, *OpenOffice* oder *LiberOffice* lesen und schreiben zu können, wurde die PHP-Bibliothek *PHPExcel* verwendet. Die Bibliothek wurde auf Basis des Microsoft OpenXML Standards und PHP entwickelt. Neben dem Lesen und Schreiben von *Excel*-ähnlichen Formaten beinhaltet die Bibliothek die Möglichkeiten zum Erstellen von Metadaten zu den Dateien, Verwendung unterschiedlicher Schriften und Schriftarten, Möglichkeiten zum gestalten der Tabellenblätter oder die Konvertierung zwischen unterschiedlichen Dateitypen. Im Rahmen des Prototyps wird es verwendet, um Adress-Tabellen (oder Koordinaten-Tabellen) einzulesen, aufzubereiten und wieder als Excel auszugeben bzw. die Inhalte an den Geocoder zu schicken.

Die in der Webapplikation Koordinaten in verschiedenen Projektionssystemen vorliegen können, müssen diese teilweise umgerechnet werden. Dafür wird die Programmbibliothek *PROJ.4* verwendet. *proj4php* ist dabei die zugehörige *PHP*-Bibliothek für die Transformation von Punkt-Koordinaten von einem Koordinatensystem in ein anderes, inkl. Datums-Transformationen. Diese Bibliothek wird verwendet, um komplizierte geodätische Umwandlungen zwischen verschiedenen geodätischen Daten zu vereinfachen. Die Projektionen können über *EPSG*-Codes eingelesen oder auch selbst im *PROJ.4*-Standard definiert werden. Dadurch sind Umwandlungen zwischen allen bekannten Projektionssystemen möglich. Diese Bibliothek wird verwendet, um die Adress-Koordinaten zwischen ETRS89 und WGS84 umzuwandeln. Die Umwandlung kann jedoch in späteren Weiterentwicklungen auf beliebige Koordinatensysteme ausgedehnt werden.

Für die Geokodierung von Adressdaten wurde ebenfalls eine PHP-Bibliothek, namens *Geocoder*, verwendet. Wie bereits in Kapitel 3 erläutert und evaluiert, gibt es verschiedene Online-Geokodierungsanbieter. Diese können alle über den *Geocoder* angesprochen und verwendet werden. Es werden einfach die unterschiedlichen Services der Geokodierungsanbieter über deren zur Verfügung gestellte APIs integriert. Somit kann über den *Geocoder* jeder Anbieter einzeln angesprochen werden. Die zwei Hauptkomponenten dieser Bibliothek sind der *HttpAdapter* und der *Provider*. *HttpAdapter* sind dafür verantwortlich Daten von den APIs der Anbieter über HTTP-Abfragen zu sammeln. Da diese Resultate in verschiedenen Formaten und Strukturen zurückgeliefert werden, werden die *Provider* benötigt. Diese beinhalten die Logik, um brauchbare Informationen zu sammeln und weiterverarbeiten zu können. Dadurch können die Adressen in Koordinaten umgewandelt werden, was essentiell für die Erstellung von rasterbasierten Karten ist.

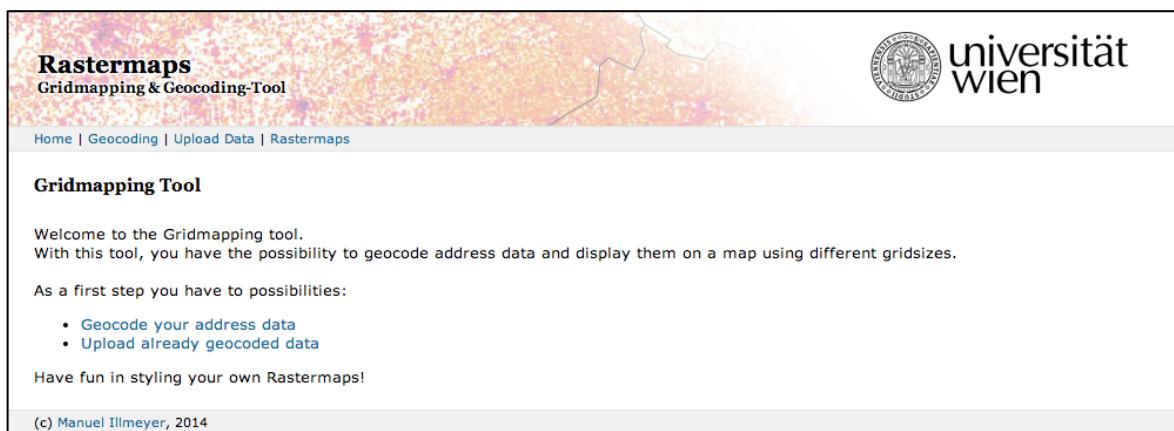
In Kapitel 2.1.2 wurde die Bedeutung der Farbe für die korrekte Darstellung eines Themas auf einer Karte bereits diskutiert. Es wurde ebenfalls bereits erwähnt, dass dem Benutzer vordefinierte Farbskalen, abhängig von der Anzahl der Klassen, zur Verfügung gestellt werden. Die *JavaScript*- und *PHP*-Arrays von *Colorbrewer* bieten dabei eine gute Möglichkeit harmonische Farbverläufe in der Webapplikation anzubieten. Dafür werden Hexadezimal-Codes für die Einfärbung von Klassen zur Verfügung gestellt. Die Arrays können ebenfalls nach Belieben mit eigenen Farbverläufen erweitert werden.

Für die Funktion des Sliders zur Einstellung der Transparenz des Overlays der Thematik wurde die *JavaScript*-Bibliothek *jQuery* verwendet. *jQuery* ist die meistverwendete *JavaScript*-Bibliothek, welche Funktionen zur dynamischen Erstellung von Webinhalten und Funktionen zur *DOM*-Navigation und -Manipulation zur Verfügung stellt. Auf Grund des großen Umfangs dieser Bibliothek wurde eine minimierte Version verwendet um den Slider umzusetzen. Dadurch kann der Slider in der Webapplikation angezeigt und der vom Benutzer eingestellte Wert für die Erstellung der Karte weiterverwendet werden.

6.3 Funktionsweise

Um die Funktionsweise der Webplattform genau erklären zu können, werden im Folgenden einzelne Screenshots der Applikation dargestellt und erläutert.

Die Einstiegsseite (siehe Abb. 34) enthält eine kurze Einführung zur Anwendung. Hierbei besteht die Möglichkeit entweder das Geokodierungs-Tool zu verwenden oder bereits geokodierte Daten hochzuladen, um auf Basis dessen Karten zu erstellen.

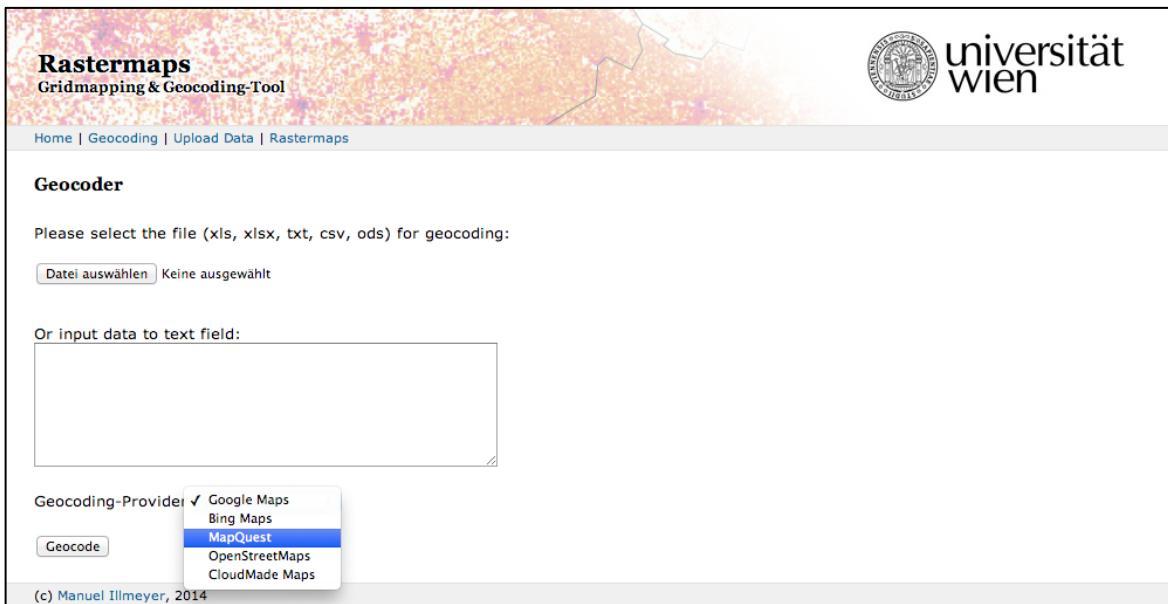


Quelle: ILLMEYER, 2014

Abb. 34: Einstiegsseite der Webapplikation

Im Menüpunkt für die Geokodierung erfolgt zunächst die Aufforderung, dass eine geeignete Datei mit Adressdaten hochgeladen werden soll (siehe Abb. 35). Im Falle der Option *CSV* muss zusätzlich der Delimiter angegeben werden. Neben dem Hochladen von Dateien, besteht des Weiteren die Möglichkeit die Adressen direkt in ein Textfeld zu kopieren. Nach Auswahl der Datenbasis muss nur noch der gewünschten Geokodierungsanbieter ausgewählt werden. Vorausgewählt ist der Dienst von Google Maps, der bei der

Evaluierung in Kapitel 3 das beste Ergebnis lieferte. Wie bereits erwähnt müssen jedoch bei diesem Dienst bestimmte Einschränkungen bei der Anzahl an Anfragen eingehalten werden. Falls sehr viele Adressen geokodiert werden sollen, ist der Dienst von OpenStreetMap zu wählen. Alle in Kapitel 3 beschriebenen Dienste stehen auf der Webapplikation zur Auswahl.



Quelle: ILLMEYER, 2014

Abb. 35: Upload von Adressdaten

Im Anschluss daran müssen die Inhalte der hochgeladenen Datei noch genauer spezifiziert werden (siehe Abb. 36). Die erste Angabe legt dabei fest, ob die Datei in der ersten Zeile bereits Daten oder lediglich den Titel enthält. In den folgenden Auswahlfeldern wird bestimmt, welche Spalten der Datei die jeweiligen benötigten räumlichen Attribute für die Geokodierung enthalten. Dadurch werden die Spalten festgelegt, die beispielsweise die Adresse oder die PLZ beinhalten. Diese Angaben sind ausschlaggebend für eine korrekte Geokodierung der Adressdaten, da die ausgelesenen Werte entweder einzeln oder in einer einzigen Zeichenkette in einer bestimmten Reihenfolge an die Dienste übergeben werden müssen.

Quelle: ILLMEYER, 2014

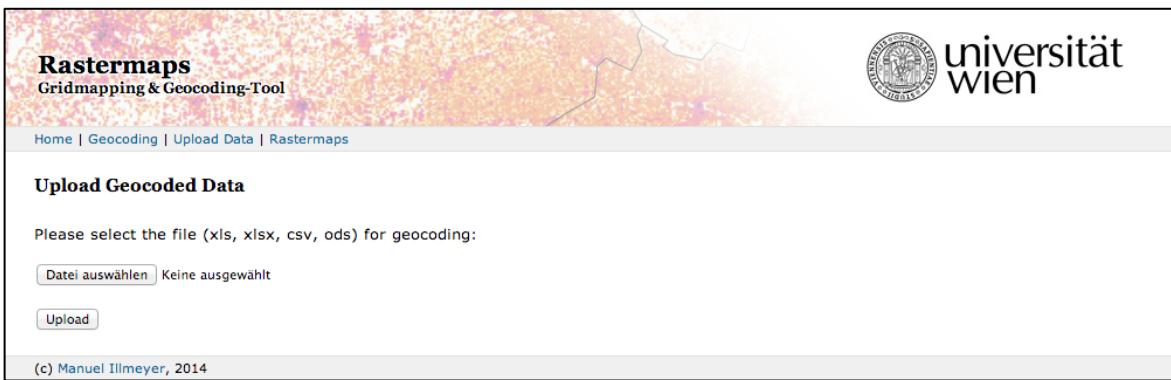
Abb. 36: Einstellungen zur hochgeladenen Adressdatei

Nachdem alle Parameter eingegeben und die Datei abgeschickt wurde, ist die Länge des Geokodierungsprozesses im Wesentlichen von der Anzahl der Adresseinträge sowie vom ausgewählten Geokodierungsanbieter abhängig. Nach erfolgreicher Geokodierung wird eine kurze Zusammenfassung, inkl. Informationen über die Dauer und der Anzahl erfolgreich geokodierter Adressen, angezeigt (siehe Abb. 37). Des Weiteren besteht die Möglichkeit, das Ergebnis als Excel-Datei herunterzuladen oder direkt mit der Erstellung der Karte fortzufahren.

Quelle: ILLMEYER, 2014

Abb. 37: Ergebnis der Geokodierung

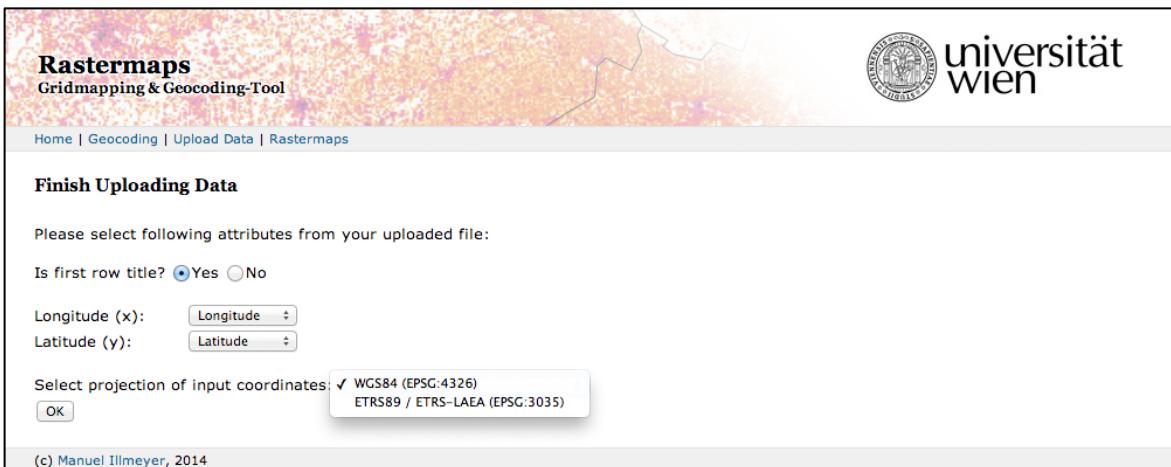
Neben dem Upload von Adressdaten besteht, wie bereits erwähnt, zusätzlich die Option bereits geokodierte Daten bzw. einfache Adresspunkte hochzuladen (siehe Abb. 38). Ähnlich wie beim Upload von Adressdaten können auch hierbei Dateien mit Tabellenkalkulations-Formaten verwendet werden.



Quelle: ILLMEYER, 2014

Abb. 38: Upload von Punktdaten

Im Anschluss daran muss angegeben werden, ob die erste Zeile bereits Daten oder einen Titel enthält. Des Weiteren müssen jene Spalten ausgewählt werden, in denen die räumlichen Informationen gespeichert sind (siehe Abb. 39). Zusätzlich kann ausgewählt werden, ob sich in der Datei geographische WGS84-Koordinaten oder ETRS-LAEA Koordinaten befinden. Nach erfolgreichem Upload kann mit der Erstellung der rasterbasierten Karten fortgefahrene werden.



Quelle: ILLMEYER, 2014

Abb. 39: Einstellungen zur hochgeladenen Koordinaten-Datei

Die Konfigurationsmöglichkeiten für die Erstellung von rasterbasierten Karten sind vielfältig und stellen eine wichtige Möglichkeit zur individuellen Gestaltung der Karte dar. Dadurch kann vor allem, neben dem Aussehen der Karte, auch deren Aussagekraft, durch z.B. unterschiedliche Klassengrenzen, entscheidend beeinflusst werden. Im Folgenden werden die Parameter, in Hinblick auf die gewonnenen Erkenntnissen aus vorhergehenden Abschnitten dieser Arbeit (vor allem Kapitel 2 und 5), aufgezählt und erläutert. Darüber hinaus sind diese in Abb. 40 ersichtlich.

Die Auswahl des darzustellenden Gebietes (*Geographic Area*) ist im Kontext dieser Arbeit beschränkt auf das Gebiet von Österreich und Europa. Der Maßstab ist ebenfalls fix vorgegeben, sodass die Karten, für Österreich im Querformat und für Europa im

Hochformat, auf einer DIN A4 Seite gut abgebildet werden können. Die Auswahl des Gebietes hängt im Allgemeinen von den vorhandenen Daten ab. Wie in Kapitel 6.4 beschrieben, kann die Applikation in einer weiteren Umsetzung so gestaltet werden, dass das darzustellende Gebiet frei gewählt werden kann, um dies an die Daten anzupassen.

In den Kapiteln 5.2 und 5.4 dieser Arbeit wurde bereits darüber diskutiert, dass die Rastergröße in Abhängigkeit des Maßstabes zu wählen ist. Weitere Einflussfaktoren für die Wahl der Rastergröße sind, wie im Kapitel 4.4 angeführt, Datenschutzbestimmungen, welche eingehalten werden müssen. Neben diesen Faktoren ist noch die Verteilung der Daten auf das darzustellende Gebiet zu beachten. Sind zu wenige Daten vorhanden um eine aussagekräftige Darstellung auf Basis kleinerer Rasterzellen zu gewährleisten, sollte eine größere Rasterweite gewählt werden. Auf Basis dieser Informationen und in Anbetracht der Performance des Systems stehen dem Benutzer verschiedene Rastweiten (*Grid Size*) für Österreich und Europa zur Verfügung.

Ebenso von den vorhandenen Daten abhängig ist die Wahl der Anzahl an thematischen Klassen (*Thematic Classes*), in welche die Daten eingeteilt werden sollen. In Kapitel 2.2.2 wurde bereits erwähnt, dass eine bestimmte Anzahl an Klassen nicht überschritten werden sollte, um die Unterscheidbarkeit von einzelnen Farbwerten noch gewährleisten zu können. Im Hinblick darauf steht dem Benutzer frei bis zu sieben thematische Klassen zu bilden.

Der wohl wichtigste Parameter, um die Aussagekraft der Karte zu gewährleisten, ist die Wahl der Klassengrenzen (*Class Limits*). Durch die Wahl der Grenzen kann die Aussage der Karte stark beeinflusst werden (siehe Kapitel 2.2.2). In den meisten Geoinformationssystemen bzw. kartographischen Informationssystemen stehen statistische Berechnungen im Hintergrund der Wahl von Klassengrenzen. In der im Rahmen dieser Arbeit entwickelten Webapplikation können die Werte lediglich manuell eingegeben werden. Die Grenzen müssen daher bereits im Vorhinein berechnet werden. In der Anwendung ist neben den Klassenobergrenzen noch die Benennung jeder Kategorie einzugeben, welche auch in der Legende angezeigt wird.

Ebenfalls wichtig für eine aussagekräftige Karte ist die Wahl des Farbschemas (*Color Scheme*) (siehe Kapitel 2.1.2). Durch die Auswahl eines geeigneten Farbschemas können auf einen Blick eindeutige Assoziationen des Betrachters der Karte mit dem dargestellten Thema erzeugt werden. So rufen beispielsweise kräftige Farbe andere Assoziationen hervor als blasses Farbtöne. Der Benutzer hat, wie bereits erwähnt, die Möglichkeit, aus verschiedenen vordefinierten ein- und mehrpoligen Farbskalen auszuwählen. Mit der Auswahl „*Flip Color Scheme*“ ist außerdem die Möglichkeit gegeben, die Anordnung der Farbabstufungen umzudrehen.

Der letzte Abschnitt der Konfigurationsparameter kennzeichnet den Bereich, in dem die Grundkarte (siehe Kapitel 2.2.1) und die Kartenrandangaben (siehe Kapitel 2.1.1) gestaltet werden können. Der Kartentitel (*Map Title*) kann frei gewählt werden und wird über der Karte angezeigt. Als ein Element der Grundkarte, neben den Ländergrenzen und der Beschriftung, kann optional eine halbtransparente Schummerung auf der Karte angezeigt

werden (*Show Hillshade*). Dabei ist jedoch darauf zu achten, dass die Transparenz der Thematik möglichst gering gewählt wird, um die Farbwiedergabe nicht zu stark zu beeinflussen. Eine weitere Gestaltungsmöglichkeit der Europakarte bietet die Auswahl der Darstellung der Meere (*Depiction of Seas*). Hierbei kann zwischen einer weißen oder einer blauen Färbung der Meere bzw. einer gestreiften Küstendarstellung gewählt werden. Der letzte Parameter ist die bereits genannte Einstellung zur Auswahl für die Transparenz der Rasterzellenfärbung (*Transparency*). Dabei kann mittels eines Sliders festgelegt werden, wie hoch die Durchsichtigkeit der Farben des darzustellenden Themas ist.

Nach der Eingabe aller benötigten Parameter werden diese auf Fehler überprüft. Anschließend wird berechnet, wie viele Adresspunkte in den jeweiligen Rasterzellen liegen, woraufhin die Input- und Skript-Dateien für *Mapyrus* erstellt werden. In der abschließenden Anzeige kann die erstellte Karte heruntergeladen werden, die allerdings erst beim Klick des Download-Buttons gezeichnet wird. Einige Beispiele für Karten, welche mit der entworfenen Webapplikation erstellt wurden, sind in Kapitel 6.5 aufgeführt und beschrieben.

Rastermaps
Gridmapping & Geocoding Tool

Home | Geocoding | Upload Data | Rastermaps

Rastermaps

Which geographic area would you like to display?

Europe Austria

Which grid size would you like to display your data on?

10km 50km 100km

How many thematic classes would you like to define?

Please select the class limit and name the classes (shown in Legend).

Class 1: Upper class limit <input type="text" value="5"/>	Class name <input type="text" value="<5 Bewohner"/>
Class 2: Upper class limit <input type="text" value="10"/>	Class name <input type="text" value="5-10 Bewohner"/>
Class 3: Upper class limit <input type="text" value="20"/>	Class name <input type="text" value="10-20 Bewohner"/>
Class 4: Upper class limit <input type="text" value="9999"/>	Class name <input type="text" value=">20 Bewohner"/>

Please select your color scheme:



Flip Color Scheme? No Yes

Map Title:

Show Hillshade: Yes No

Depiction of Seas: None Blue Dashed Shore

Transparency:

(c) Manuel Illmeyer, 2014

Quelle: ILLMEYER, 2014

Abb. 40: Konfigurationsmöglichkeiten für rasterbasierte Karten

6.4 Weiterentwicklungen

Wie bereits erwähnt, stellt die, im Rahmen dieser Arbeit entwickelte, Applikation lediglich einen ersten Prototypen für die Erstellung von rasterbasierten Karten im Internet dar. Im Folgenden werden einige Punkte diskutiert, in denen die Anwendung noch weiterentwickelt werden könnte und sollte.

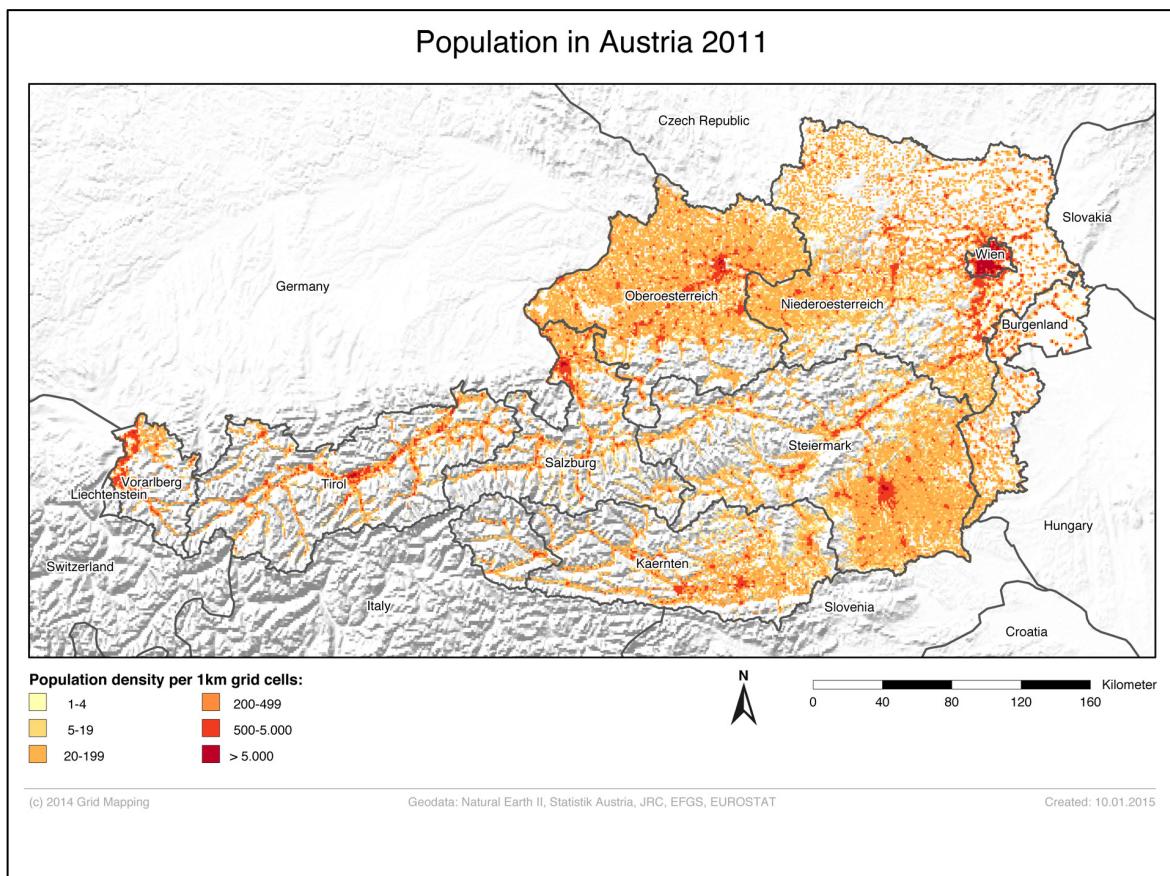
- *Berechnung von Statistiken für Klassengrenzen*: Derzeit ist es nur möglich, die Klassengrenzen manuell einzugeben. In einem weiteren Schritt sollten die Klassengrenzen auch durch die Anwendung von Statistiken aus den Eingangsdaten berechnet werden können (z.B. gleiche Intervalle, Quantile, natürliche Unterbrechungen, Standardabweichung, etc.).
- *Benutzerfreundlichkeit*: Die Benutzerfreundlichkeit kann bspw. durch eine bessere Darstellung der Konfigurationsparameter der Karte erhöht werden.
- *Performance*: Die Performance kann erheblich verbessert werden, indem für die Geokodierung nicht einzelne Abfragen an die jeweiligen Geokodierungsdienste geschickt werden, sondern diese so aufbereitet werden, dass eine Batch-Verarbeitung möglich ist. Dies erfordert jedoch für die meisten Dienste den entgeltlichen Ankauf von Lizenzrechten.
- *Gestaltungsmöglichkeiten der Karte*: Die Darstellung der Karte ist derzeit vorgegeben. Durch Einstellungsmöglichkeiten der Platzierung von Elementen, Auswahl von Schriftarten, Definition von eigenen Farbschemen, Auswahl von Projektionen, Anordnung der Legendeneinträge, etc. können hier weitere Verbesserungen vorgenommen werden.
- *Eingabehilfen*: Für die einzelnen Eingabe- und Konfigurationsparameter sollten Hilfen für den Benutzer vorhanden sein. Des Weiteren wäre eine detailliertere Abfrage von Benutzerfehlern sinnvoll.
- *Gleichzeitige Zugriffe*: Derzeit ist die Applikation noch nicht für gleichzeitige Zugriffe von mehreren Benutzern optimiert.
- *Koordinatensysteme*: Die Möglichkeiten für den Upload von Adresspunkten in unterschiedlichen Koordinatensystemen sollte erweitert werden.
- *Darstellungsflächen*: Derzeit ist es nur möglich, die Daten österreich- oder europaweit darzustellen. In einem weiteren Schritt könnte die Anwendung insofern erweitert werden, dass auch zu den Daten passende Geometrien hochgeladen werden können.
- *Mapyrus*: *Mapyrus* sollte ebenfalls noch weiter verbessert werden. So gibt es derzeit noch Fehler bei der Darstellung von Sonderzeichen (z.B. Umlaute) im Zusammenspiel mit Servlets. Außerdem sollte es möglich sein, Beschriftungen so einzustellen, dass diese sich gegenseitig nicht überlappen.

6.5 Beispiele

Nachfolgend werden noch einige Beispiele, welche mit Hilfe der im Rahmen dieser Arbeit entwickelten Webapplikation erzeugt wurden, präsentiert, die verwendeten Daten erläutert und die jeweilige Methodik der Erstellung der Karte erklärt.

Abb. 41 zeigt die Bevölkerungsdichte Österreichs auf Basis von 1km Rasterzellen. Als Basis der Bevölkerungsdaten wurde der Datensatz des EFGS verwendet, welcher auf deren Webseite frei zur Verfügung steht. Es handelt sich dabei um Daten, welche 2011 europaweit erhoben wurden und vom *Austrian Institute of Technology (AIT)* in einen konsolidierten Datenbestand zusammengefügt wurden. Die Daten liegen bereits in einem 1km ETRS-LAEA Raster vor, in welchem die Anzahl der Bewohner pro Rasterzelle gespeichert ist. Somit mussten nur noch die Daten von Österreich herausgeschnitten und in die Webapplikation hochgeladen werden.

Um die Vergleichbarkeit mit der originalen Darstellung (vgl. Abb. 21) der Bevölkerungsdaten zu gewährleisten, wurden für die Erstellung der Karte ebenfalls sechs Klassen gebildet und die Klassengrenzen gleich gewählt (siehe Legende in Abb. 41). Die Einteilung der Klassen erfolgt somit nach dem Schema von Sinnklassen (vgl. Kapitel 2.2.2), da eine Einteilung bereits von den nationalen Statistikinstituten Europas vorgenommen wurde. Das Farbschema unterscheidet sich jedoch von dem der originalen Karte. In Abb. 41 wurde ein Farbverlauf von Gelb (niedrige Bevölkerungsdichte), über Orange, bis hin zu Rot (hohe Bevölkerungsdichte) gewählt. Dieses Farbschema findet sich in vielen Bevölkerungsdichtekarten wieder, da das kräftige Rot auf den ersten Blick mit einer hohen Dichte assoziiert werden kann. Weiters wurde der Grundkarte eine Geländeschummerung hinzugefügt, um die höhere Dichte der Bevölkerungsverteilung in den Tälern Österreichs noch zu verdeutlichen.

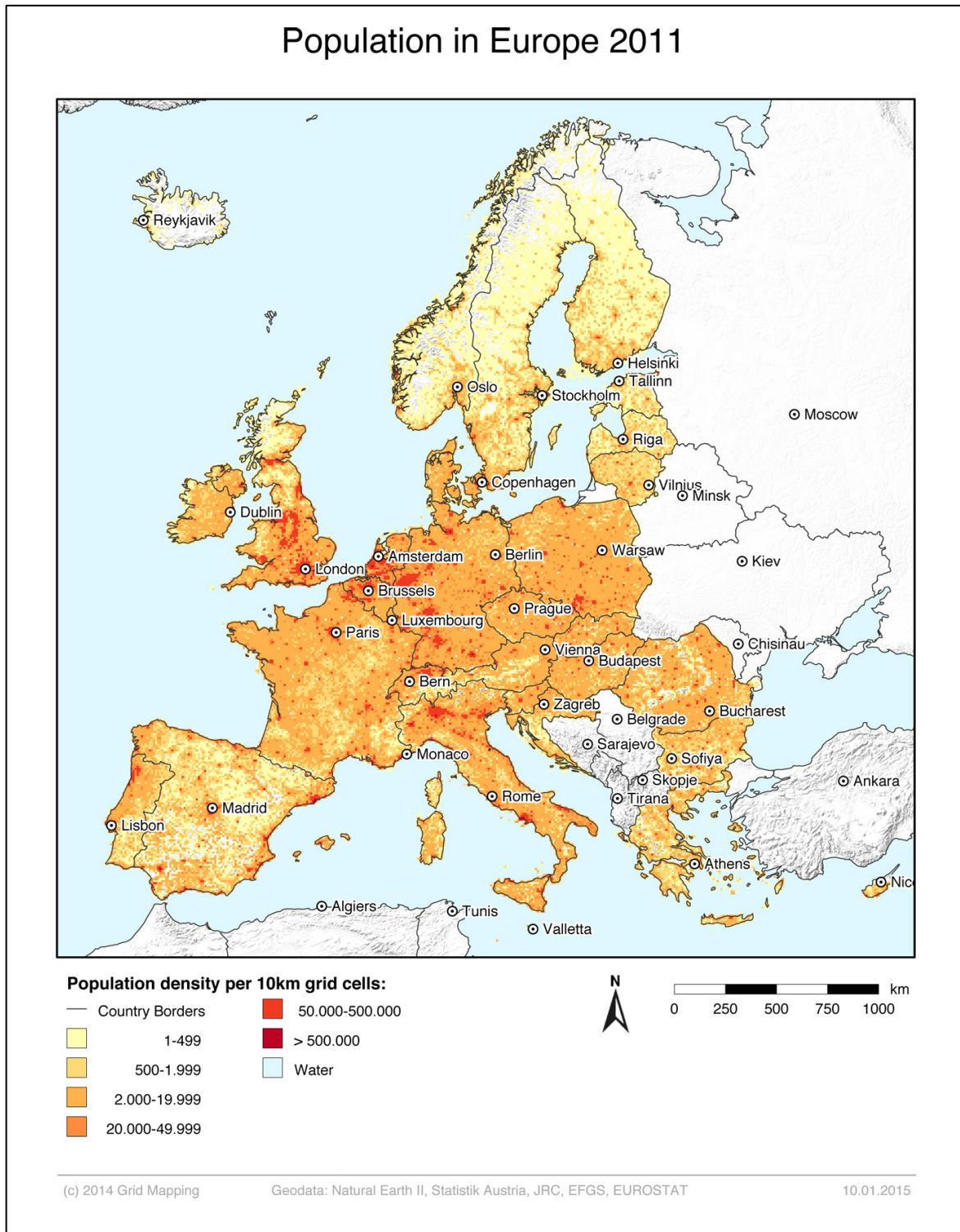


Quelle: ILLMEYER, 2015

Abb. 41: Bevölkerungsdichte Österreichs 2011 auf Basis 1km Rasterzellen

Abb. 42 zeigt die Bevölkerungsdichte Europas auf Basis von 10km Rasterzellen. Als Datenbasis wurde, wie im vorigen Beispiel, ebenfalls der Datensatz des EFGS von 2011 verwendet. Da die Daten in einem 1km ETRS-LAEA Raster vorliegen, die Karte jedoch auf Basis von 10km Rasterzellen erstellt wurde, mussten die Daten noch aggregiert werden. Dafür wurden die Rasterzellenmittelpunkte, inkl. der darin lebenden Bevölkerung, abgespeichert und mit Hilfe der Webapplikation auf 10km-Rasterzellen umgerechnet.

Um die Vergleichbarkeit mit der originalen Darstellung (vgl. Abb. 21) der Bevölkerungsdaten zu gewährleisten, wurden für die Erstellung der Karte sechs Klassen gebildet und die Klassengrenzen um ein Hundertfaches nach oben gesetzt, da 100 1km-Rasterzellen innerhalb einer 10km-Rasterzelle liegen (siehe Legende in Abb. 42). Die Einteilung der Klassen erfolgt somit ebenfalls nach dem Schema von Sinnklassen (vgl. Kapitel 2.2.2). Die Farbskala ist ident zu der in Abb. 41 und es wird ebenfalls eine Geländeschummerung in der Grundkarte angezeigt.



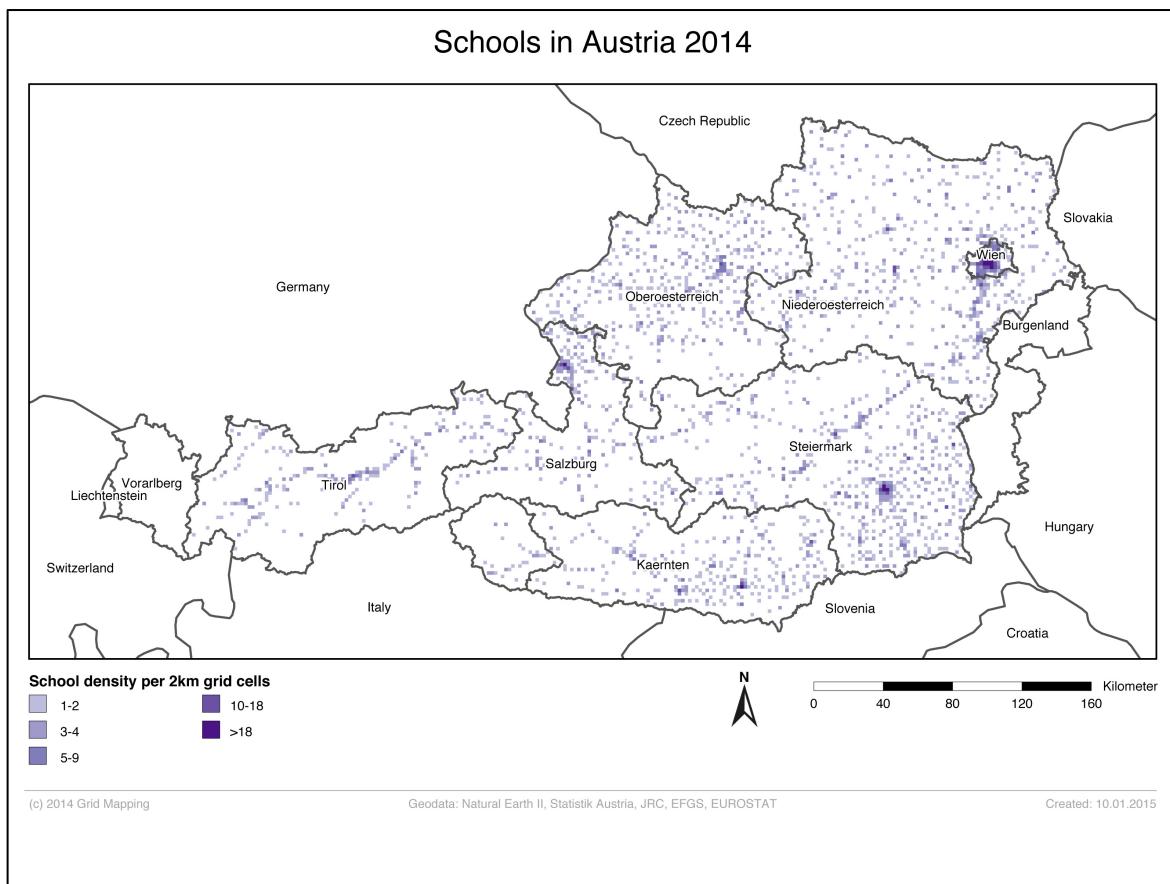
Quelle: ILLMEYER, 2015

Abb. 42: Bevölkerungsdichte Europas 2011 auf Basis 10km Rasterzellen

In Abb. 43 sind die Schulen in Österreich auf Basis von 2km-Rasterzellen ersichtlich. Die Daten, mit denen die Karte erstellt worden ist, stammen aus dem Open Government Data (OGD) Portal Österreichs (<http://data.gv.at/>). Auf diesem Portal sind offene Geodaten der

Bundesländer unter Einhaltung bestimmter Lizenzierungsrichtlinien erhältlich. Die Daten der Schulen waren, außer für Burgenland und Vorarlberg, flächendeckend verfügbar.

Die Erstellung der Karte erfolgte mit den Adressen der Schulen, welche mit Hilfe der Webapplikation in einem ersten Schritt geokodiert wurden. Anschließend wurden die geokodierten Daten auf Basis des 2km-Rasters Österreichs in fünf Gruppen klassifiziert. Die Klassifizierung erfolgte dabei auf Basis von natürlichen Grenzen (siehe Kapitel 2.2.2). Weiters wurde ein neutrales Farbschema gewählt, welches von einem hellem bis zu einem dunklem Violett reicht, wobei der dunklere Farbton wiederum eine höhere Dichte darstellt. Auf eine Schummerung wurde bei dieser Karte verzichtet um einen guten Überblick über die relative geringe Dichte an Schulen noch gewährleisten zu können. Es ist trotzdem gut sichtbar, dass die Schulen im Bergland entlang von Tälern errichtet wurden. Weiters sind die großen Städte wie Wien, Graz, Salzburg, Innsbruck oder Linz auf Grund einer hohen Schuldichte gut zu erkennen. Im Flachland sind die Schulen relativ verstreut auf das gesamte Gebiet.



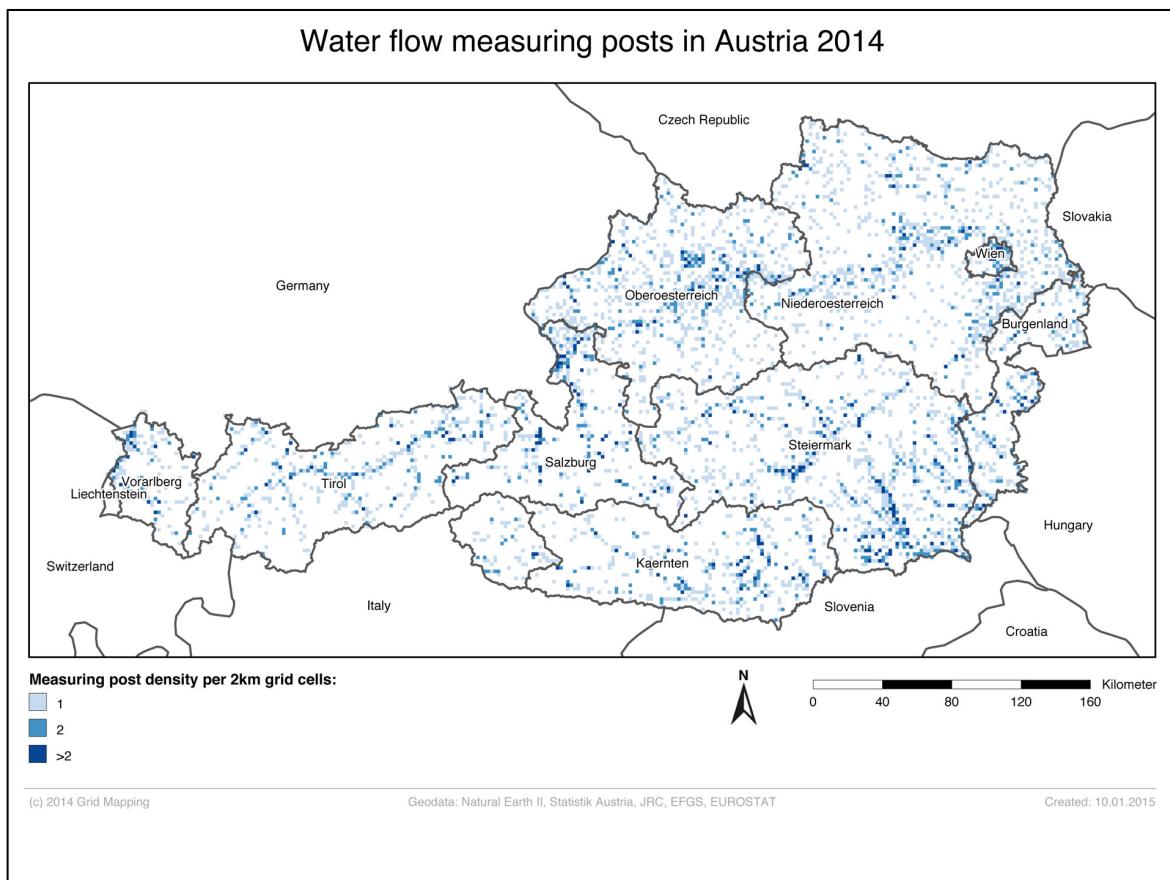
Quelle: ILLMEYER, 2015

Abb. 43: Schulen in Österreich auf Basis von 2km-Rasterzellen

Abb. 44 soll zeigen, dass jegliche punktbasierte oder adressbasierte Daten mit Hilfe der Webapplikation auf einer rasterbasierten Karte dargestellt werden können. In diesem Fall sind es Grundwasser- und Fließgewässer-Messstellen in Österreich. Die Daten dazu stammen von der Geometadatenbank der Plattform INSPIRE Österreich

(<http://geometadatensuche.inspire.gv.at/>) und stehen in Downloaddienst des Umweltbundesamtes zum Thema Wasser zur Verfügung. Es handelt sich dabei um einfache Punktdaten, welche den Ort einer Messstelle kennzeichnen.

Die Erstellung der Karte erfolgte mit den Koordinaten der Punktdaten, welche mit Hilfe der Webapplikation nur noch den zugehörigen Rasterzellen zugewiesen werden mussten. Auf Grund der eher geringen Anzahl an Punkten wurden lediglich drei Klassen gebildet. Die Klassifizierung erfolgte dabei auf Basis von Natürlichen Grenzen (siehe Kapitel 2.2.2). Um eine Assoziation zu den Wasserdaten herstellen zu können, wurde ein blaues Farbschema gewählt. Auffällig in dieser Abbildung ist dabei eine erhöhte Dichte an Messstellen im Bereich der Donau.



Quelle: ILLMEYER, 2015

Abb. 44: Grundwasser- und Fließgewässer-Messstellen in Österreich auf Basis von 2km-Rasterzellen

7 Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Aufbereitung von Adressdaten bzw. thematischen Daten, die auf Adressen oder Adresspunkten basieren, sowie auf deren Darstellung auf Basis geometrischer Rasterzellen. Während in der Regionalstatistik diese kartographische Abbildungsart bereits weit verbreitet ist, sind rasterbasierte Karten in der klassischen Kartographie noch eher selten. Aus diesem Grund wurde im Zuge dieser Arbeit großer Wert auf die verschiedenen Anwendungsbereiche der Kartographie sowie auf die kartographische Gestaltung gelegt und ein Online-Tool entwickelt, mit dem man diese Art von Karten einfach erstellen kann.

Zur Erfassung der Grundlagen zu diesem Thema wurde zu Beginn dieser Arbeit die Kartographie in ihren Grundlagen und Prinzipien erläutert. Neben einer Beschreibung, aus welchen Elementen eine Karte besteht und welche Informationen eine Karte vermitteln kann oder soll, wurde der Fokus auf die Gestaltung und Visualisierung von Karten gelegt. Durch eine angemessene Wahl von Farben und Signaturen sowie entsprechende Generalisierungsmaßnahmen kann die Lesbarkeit einer Karte gewährleistet werden.

Im Rahmen der Umsetzung der Webapplikation wurde diskutiert, welche Elemente eine Karte beinhalten sollte. Somit sind Bestandteile, wie die Legende, der Maßstabsbalken und ein Nordpfeil, vorerst fix vorgegeben. Diese werden auf jeder Karte angezeigt und können in einer ersten Umsetzung nicht verändert werden. Die Wahl des Projektionssystems, in der die resultierenden Karten dargestellt werden, fiel auf Grund der vorhandenen Daten, des verwendeten Rasternetzes und der Vorgaben der *INSPIRE*-Initiative auf das Referenzsystem „ETRS89/LAEA“.

Um die Daten auf rasterbasierten Karten nach kartographischen Kriterien anzeigen zu können, wurden die kartographischen Variablen beschrieben. Von den sechs Variablen Größe, Form, Füllung, Tonwert, Richtung und Farbe lassen sich jedoch nur die drei Variablen Füllung, Tonwert und Farbe auf rasterbasierte Karten anwenden. Dem Ersteller einer rasterbasierten Karte stehen dafür in der entworfenen Webanwendung vorgefertigte Farbskalen zur Verfügung, welche sich in Tonwert und Farbe unterscheiden. Da die erstellten Karten in einem ersten Schritt auf dem Bildschirm angezeigt werden, jedoch auch für den Druck optimiert sind, müssen Richtlinien der analogen und der digitalen Kartographie eingehalten werden.

Da das Endprodukt dieser Arbeit eine Applikation darstellt, welche thematische rasterbasierte Karten erstellt, wurde die thematische Kartographie im Detail betrachtet, wobei die Erkenntnisse, die aus den Grundlagen der Kartographie gewonnen wurden, einflossen. Der Erstellung der Grundkarte wurde in diesem Zusammenhang ebenfalls erhöhte Aufmerksamkeit geschenkt, da diese dazu beiträgt, die thematischen Inhalte besser erfassen zu können. Der Benutzer der Webapplikation hat dabei die Möglichkeit, zwischen verschiedenen Gestaltungselementen zu wählen. So kann bspw. eine Geländeschummierung optional angezeigt und die Darstellung von Meeresflächen beeinflusst werden. Die restlichen Grundkartenelemente wie z.B. Ländergrenzen und Beschriftungen sind

vorgegeben. Aufgrund der Vielzahl an in Kartenform darstellbaren Themen, wurden die verschiedenen Kategorien thematischer Karten erläutert. Bei Karten basierend auf regionalstatistischen Rastern handelt es sich nach Meinung des Autors um spezielle Choroplethenkarten, auch Flächenkartogramme genannt, da die thematische Grenzziehung einem definierten Muster folgt und Absolutwerte, auf Grund gleichbleibender Bezugsflächen, auch als Flächenfärbung angezeigt werden können. Die Anzahl der Klassen und die Wahl der Klassifizierungsmethode sind ebenfalls von großer Bedeutung, um eine korrekte und aussagekräftige Darstellung thematischer Inhalte zu gewährleisten. In der entworfenen Webapplikation sind die Klassengrenzen frei wählbar. Folgende Klassifizierungsmethoden können dabei unterschieden werden: (vgl. OLBRICH et al. 2002: 42ff)

- Äquidistante Klassen: Alle Klassen haben die gleiche Breite, d.h. den gleichen Abstand zwischen der oberen und unteren Klassengrenze.
- Mathematische Progression: Alle Verfahren, die auf Grundlage von mathematischer Reihenbildung entstehen (z.B. arithmetische oder geometrische Reihe).
- Quantile: Hier werden die Grenzen so gewählt, dass in jede Klasse gleich viele Werte fallen.
- Sinnklassen: Die Klassenbildung erfolgt auf Basis einer sachlogischen Begründung
- Natürliche Klassen: Mit Hilfe von mathematischen Algorithmen werden Datenlücken genutzt, um Klassen zu definieren.
- Statistische Parameter: Klassenbildung mit Hilfe statistischer Parameter (z.B. arithmetisches Mittel und Standardabweichung)

Zur Beantwortung der Forschungsfrage wurden weitere (Unter-) Forschungsfragen gebildet, die in den verschiedenen Kapiteln beantwortet wurden. So beschäftigt sich eine Forschungsfrage mit dem Thema der Geokodierung, wobei eruiert wurde, welche Anbieter es bereits gibt und wie genau diese Dienste arbeiten. Zur adäquaten Auseinandersetzung mit dieser Fragestellung war es zunächst notwendig den Begriff „Geokodierung“ zu definieren. Demnach kann unter diesem Begriff jener Vorgang verstanden werden, bei dem postalische Anschriften in Koordinatenpaare umgewandelt werden. Die Methoden der Geokodierung wurden in diesem Zusammenhang ebenfalls kurz beschrieben, wobei bedingt durch das Thema in dieser Arbeit hauptsächlich Online-Dienste genauer betrachtet wurden. Diese haben den Vorteil, dass keine externe Software benötigt wird und die Anschaffung, Wartung und Aktualisierung der Referenzdaten vom Anbieter des Dienstes vorgenommen wird. Weiters sind diese Online-Services leicht zu bedienen und man erhält schnelle Ergebnisse in Form von Koordinaten. Um Online-Geokodierungsdienste evaluieren zu können wurden in einem ersten Schritt die Anbieter identifiziert, welche für das Gebiet von Europa in Frage kommen und deren Limitierungen festgehalten. Im Rahmen der Arbeit sind dies die Dienste von *Bing Maps*, *CloudMade*, *Google*, *MapQuest* und *OpenStreetMap*.

Zur Evaluierung der Anbieter dieser Dienste, wurden die drei Kriterien „Genauigkeit“, „Zuverlässigkeit“ und „Schnelligkeit“ überprüft, wobei ersteres in die Subkriterien „Lagefehler“ und „Trefferraten“ sowie das zweite Kriterium in die Subkriterien „Fehlerhafte Datensätze“ und „Limitierungen“ unterteilt wurden. Auf Basis dessen wurden

in einem analytischen Hierarchieprozess die Kriterien in einem paarweisen Vergleich in Hinblick auf dessen Wichtigkeit für die Webapplikation bewertet und die Dienste eingehend untersucht. Dabei konnte festgestellt werden, dass die „Trefferraten“ und die „Schnelligkeit“ am meisten aussagekräftig sind (siehe Tab. 9). Das Ergebnis ist nachvollziehbar, da auf Basis ungenauer Geokodierungsergebnisse keine adäquaten Karten erzeugt werden können. Des Weiteren sollte die Bearbeitung über das Online-Tool keine langen Wartezeiten aufweisen.

Die Schnelligkeit der Dienste wurde dadurch ermittelt, in dem, über den Zeitraum von einer Woche, der zur Verfügung gestellte Adressdatensatz an die fünf Dienste übermittelt und auf ein Ergebnis gewartet wurde. Der Lagefehler eines einzelnen Adresseintrages wurde durch die Entfernung der bekannten Koordinaten aus den Quelldaten zu den resultierenden Koordinaten eines Geokodierungsvorganges berechnet. Die Trefferraten wurden durch die Anzahl der gültigen Geokodierungsergebnisse bestimmt. Das Kriterium der Limitierung wurde auf Basis der Angaben der jeweiligen Betreiber qualitativ bewertet. Das letzte Kriterium der fehlerhaften Datensätze wurde insofern ermittelt, in dem ein Datensatz mit fehlerhaften Adresseinträge an die Dienste geschickt und das Resultat mit den zu erwartenden Ergebnissen verglichen wurde.

Die umfassende Evaluierung kam zu dem Ergebnis, dass das Service von *Google* im Vergleich die besten Ergebnisse hinsichtlich „Trefferraten“, „Schnelligkeit“ und „Lagefehler“ aufweist, wobei der Dienst von *Bing* fast annähernd gleichwertig abschneiden konnte. Allerdings sind die Requests pro Tag bei beiden Diensten limitiert, ein Nachteil der mittels Entgeltzahlungen aufgehoben werden könnte. In diesem Fall wäre zudem die Nutzung von Batch-Requests möglich, da diese an den Erwerb der API gebunden sind. Im Rahmen der vorliegenden Arbeit wird für die Geokodierung von kleineren Adressdatensätzen somit der Dienst von *Google* empfohlen. Für eine größere Anzahl an Adressen muss jedoch das Service von *OpenStreetMap* verwendet werden. Dieser Dienst belegte im Rahmen der Evaluierung Platz 3, weist allerdings vergleichbar gute Ergebnisse wie *Google* bezüglich der „Trefferraten“ auf. Hinsichtlich der „Schnelligkeit“ ist *OpenStreetMap* dem Dienst von *Google* jedoch klar unterlegen. Falls die Applikation im Echtbetrieb, z.B. in einem Unternehmen, verwendet wird, ist anzuraten, den kommerziellen Dienst von *Google* zu verwenden. Dieser weist, wie bereits erwähnt, Vorteile gegenüber den anderen Diensten auf und hat in der bezahlten Version die Möglichkeit mehrere Adressen in einer Anfrage geokodieren zu lassen, was die benötigte Zeit noch deutlich reduziert. Die Ergebnisse der Evaluierung ergeben folgende Reihenfolge der Geokodierungsdienste:

1. Google (29,32 %)
2. Bing (24,83 %)
3. OpenStreetMap (16,30 %)
4. CloudMade (15,04 %)
5. MapQuest (14,51 %)

Hinsichtlich der Frage, wo adressgenaue Daten zu Verfügung stehen, müssen nicht nur die preislichen Konditionen sondern auch gesetzliche Aspekte beachtet werden. So sind Daten

eines Landes meist nicht in adressgenauer Form erhältlich, allerdings gibt es jedoch einige Datensätze, die auf Basis von regionalstatistischen Rastern erstellt wurden. Weiters sind auf Grund von Datenschutzbestimmungen Merkmalsausprägungen einzelner Rasterzellen, welche auf einzelne Individuen zurückgeführt werden können, nicht zulässig. Um dies zu verhindern wurden verschiedene Maßnahmen erläutert. So werden beispielsweise Rasterzellen mit zu geringen Fallzahlen nicht weitergegeben oder die Rasterzellengröße erhöht. Im Zuge der *INSPIRE* Initiative wurde für diese Daten ein einheitliches Projektionssystem sowie ein einheitliches Raster definiert, was länderübergreifende Vergleiche und den Datenaustausch erheblich erleichtert. In Österreich sind entsprechende Daten hauptsächlich bei der Bundesanstalt *Statistik Austria* erhältlich. Europaweite Datensätze können teilweise auch vom *EFGS* bzw. von den statistischen Einrichtungen der einzelnen Mitgliedsländer bezogen werden. Zusätzlich verfügen auch Unternehmen über adressgenaue Daten ihrer Kunden, auf Basis deren Kundendatenbanken ebenfalls rasterbasierte Karten erstellt werden können. Letzteres gilt als eines der wichtigsten Einsatzgebiete für die im Rahmen dieser Arbeit erstelle Online-Plattform.

Da die Daten in der Webapplikation auf Basis regionalstatistischer Raster dargestellt werden, musste dieser Begriff in einem ersten Schritt genauer definiert werden. Geometrische Raster können demnach als flächendeckende, regelmäßige, hierarchisch unterteilbare bzw. zusammenfassbare räumliche Einheiten beschrieben werden. Sie entsprechen dabei einem regelmäßigen Netz an Quadranten, die sinngemäß über die topographische Bezugsfläche der Karte gelegt werden und dabei den thematischen Inhalt der Daten kommunizieren.

Im Zuge eines Vergleichs mit administrativen Bezugseinheiten konnte festgestellt werden, dass mit Hilfe von Rastern aufgrund deren einheitlicher Größe und neutraler Form, bessere Vergleichsmöglichkeiten hinsichtlich der Fläche vorliegen. Des Weiteren können zeitliche Veränderungen besser erfasst werden, da Raster nicht an administrativen Grenzen gebunden sind und somit Grenzänderungen vernachlässigt werden können. Dies wirkt sich vor allem auch bei länderübergreifenden Darstellungen aus, da sich die Verwaltungseinheiten zwischen Ländern stark unterscheiden können. Bei Darstellungen auf Basis geometrischer Raster wirkt sich dies nicht aus, da der gleichbleibende Raster über das gesamte Gebiet gelegt wird. Die Vorteile werden in nachfolgender Aufstellung nochmals zusammengefasst: (vgl. KAMINGER u. WONKA, o.J.: 1)

- Rastereinheiten ermöglichen durch die einheitliche Flächengröße und Form neutrale Raumgliederungen und bessere flächenmäßige Vergleiche.
- Da sich die statistischen Daten auf gleiche Flächen beziehen sind Absolutwerte gleichzeitig auch Dichtewerte.
- Bei Rasterzellen gleicher Dichteklasse kann die trennende Zwischenlinie weggelassen werden um einheitliche Dichtegebiete zu erhalten.
- Da bei Rastern keine Grenzänderungen auftreten können zeitliche Veränderungen besser festgestellt werden.
- Es können räumliche Analysen durchgeführt werden die auf Basis von Gebietsgliederungen gleicher Flächengröße und Form aufgebaut sind.

Die Wahl der Rastergröße hängt im Wesentlichen vom Kartenmaßstab, dem darzustellenden Thema sowie dem Gebiet, auf das sich die Daten beziehen, ab. Bspw. kann sich die Rasterzellengröße in urbanen Räumen von jenen in alpinen Gebieten unterscheiden, obwohl in beiden Fällen der gleiche Maßstab vorliegt sowie das gleiche Thema präsentiert wird. In diesem Zusammenhang wurden im Rahmen dieser Arbeit Beispiele für unterschiedliche Darstellungsarten von rasterbasierten Karten gezeigt. So existieren einerseits Darstellungen, in denen die Rasterzellen vollständig eingefärbt sind. Andererseits kann die Thematik auch mit Punkten, die im Mittelpunkt der Rasterzellen platziert und nach Farbe und Größe variiert werden können, veranschaulicht werden. Im Rahmen der Webapplikation können in der ersten Umsetzungsvariante jedoch lediglich rasterbasierte Karten mit eingefärbten Rasterzellen erstellt werden. Derartige rasterbasierte Darstellungsformen werden bereits in den Bereichen der Infrastruktur-, Verkehrs- und Raumplanung sowie der Regionalstatistik oder auch im Geomarketing eingesetzt. Die Zuordnung von Adresspunkten erfolgt in der entworfenen Webanwendung über Aggregationsmethoden, da die Grundlagendaten in adressgenauer Form vorliegen. Die Zuordnung von Flächen oder Linien zu Rasterzellen ist derzeit noch nicht geplant.

Auf Basis der gewonnenen Erkenntnisse wurde ein Prototyp einer Webapplikation entwickelt, die den Umwandlungsprozess von Adressen in Koordinaten sowie die Darstellung der Thematik auf einer rasterbasierten Karte umfasst. Mit Hilfe dieser Anwendung ist es möglich, eine digitale Tabellendatei mit einer beliebigen Anzahl von Adressen auf die Plattform zu laden, wobei diese bearbeitet und mit den Koordinatenpaaren wieder ausgegeben wird. Wie bereits erwähnt, erfolgt die Geokodierung wahlweise mit einem der in Kapitel 3 beschriebenen Online-Diensten. Nach Abschluss der Geokodierung können die ermittelten Koordinaten auf einer Karte dargestellt werden. Der Benutzer hat dabei verschiedene Möglichkeiten, die Darstellung der Karte zu beeinflussen. Dazu werden unterschiedlich große Rasterzellen als Bezugsebene angeboten. Weiters wurde die Wahl der Klassenanzahl auf drei bis sieben beschränkt, wobei die Klassengrenzen frei wählbar sind. Die Einteilung der Klassen beeinflusst das Kartenbild, neben dem gewählten Farbschema, am meisten. Hinsichtlich des Farbschemas werden harmonische ein- und zweipolige Skalen zur Auswahl angeboten. Weiters kann die Grundkarte durch die Anzeige einer Höhenschummerung oder durch die Veränderung der Meeresdarstellung gestaltet werden. Um die Wichtigkeit der Konfigurationsparameter für die Erstellung rasterbasierter Karten mit Hilfe der Webapplikation hervorzuheben, werden diese hier nochmals taxativ aufgezählt:

- Auswahl des darzustellenden Gebietes
- Auswahl der Rasterzellengröße
- Auswahl der Anzahl an thematischen Klassen
- Eingabe der Klassenobergrenzen (und zugehörige Klassenbezeichnungen)
- Auswahl eines Farbschemas (inkl. der Möglichkeit zur Umkehr der Anordnung der Farben im Farbschema)
- Eingabe des Kartentitels
- Optionale Anzeige einer Geländehöhenschummerung
- Auswahl an Möglichkeiten zur Darstellung der Meere (nur im Maßstab Europa)

- Einstellung der Transparenz, mit der das Thema, und somit die Flächenfärbung der Rasterzellen, gezeichnet wird

Wie bereits erwähnt, handelt es sich bei der entwickelten Anwendung um einen Prototyp für die Erstellung rasterbasierter Karten auf Basis adressgenauer Daten. Aus diesem Grund bestehen selbstverständlich Entwicklungspotentiale, die vor allem folgende Bereiche umfassen: Performance, automatische Berechnung von Klassengrenzen, Gestaltungsmöglichkeiten der Karte, Auswahl mehrerer Projektionssystem, Verbesserung bei parallelen Zugriffen sowie die individuelle Definition von Darstellungsflächen.

Im Rahmen dieser Arbeit erfolgte eine schrittweise Annäherung an die zentrale Forschungsfrage. In diesem Zusammenhang konnten auf Basis der Beantwortung der einzelnen Subfragen einerseits die Vorteile regionalstatistischer Raster gegenüber administrativen Einheiten verdeutlicht und aufgelistet werden. Andererseits konnten kontextbezogene Empfehlungen gegeben werden, wie diese spezielle Darstellungsform angewandt werden sollte. Darüber hinaus flossen alle gewonnenen Erkenntnisse dieser Forschungsarbeit in die Entwicklung der präsentierten Webapplikation ein, die somit als konkrete Beantwortung der Forschungsfrage angesehen werden kann. Die vorliegende Arbeit soll außerdem als Beitrag zur weiteren Entwicklung regionalstatistischer Rasterdaten sowie der dafür notwendigen Infrastrukturen angesehen werden, damit die Vorteile und Potentiale dieser speziellen kartographischen Darstellungsform nicht unberücksichtigt bleiben.

8 Literatur

- ARNBERGER E. (1977): Thematische Kartographie – Mit einer Kurzeinführung über Automation in der thematische Kartographie. Braunschweig.
- BEV (2013): Bundesamt für Eich- und Vermessungswesen. Adressregister. http://www.bev.gv.at/portal/page?_pageid=713,2167982&_dad=portal&_schema=PORTAL (28.03.2013)
- BLOCH V. V. H. (2010): The GEOSTAT project. Dissemination of results from 2010/2011 censuses. <http://212.70.168.20/wp-content/uploads/2012/10/2010JurmalaPT5.4.pdf> (17.03.2013)
- BLOCH V. V. H., GUNDERSEN G. und THORSDALEN B. (2010): A state of the Art report Norway. http://www.efgs.info/workshops/efgs-2010-tallinn-estonia/15_Population%20and%20Housing%20Census%20GIS%20in%20Norway_Vilni%20verner%20Holst%20Bloch%20_%20Bjorn%20Thorsdalen.pdf (at_download/file) (22.03.2013)
- CARTO CIUDAD (2013): Catálogo de productos. <http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do> (28.03.2013)
- CAYO M. R. und TALBOT T. O. (2003): Positional error in automated geocoding of residential addresses. – In: International Journal of Health Geographics 2 (10).
- CLOUDMADE (2013): Geocoding and Geosearch. <http://developers.cloudmade.com/projects/show/geocoding-http-api/> (27.02.2013)
- CORCORAN D. (2013): E-Mail Korrespondenz vom 12.03.2013. Central Statistics Office Ireland.
- CRKVENČIĆ B. (2013): E-Mail Korrespondenz vom 01.03.2013. Croatian Bureau of Statistics.
- DAVIS C. A. und ALENCAR R. O. (2011): Evaluation of the quality of an online geocoding resource in the context of a large Brazilian city. – In: Transactions in GIS 15 (6), 851-868.
- DEARWENT S. M., JACOBS R. R. und HALBERT J. B. (2001): Locational uncertainty in georeferencing public health datasets. – In: Journal of Exposure Analysis and Environmental Epidemiology 11, 329-334.
- DUVANDER A. (2012): 7 Free Geocoding APIs: Google, Bing, Yahoo and MapQuest. <http://blog.programmableweb.com/2012/06/21/7-free-geocoding-apis-google-bing-yahoo-and-mapquest/> (03.03.2013)
- EFGS (2012): GEOSTAT 1A – Representing Census data in a European population grid. http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/documents/ESSnet%20project%20GEOSTAT1A%20final%20report_0.pdf (17.03.2013)
- FAHRNER W. (2012): Geodatenkatalog des BMLFUW – Übersicht über Geodatensätze in der GDI-L GDS (GeoDatenInfrastruktur-Lebensministerium GeoDatenSammlung). <http://www.gruenerbericht.at/cm3/download/finish/124-gis/569-gis-geodatenkatalog-gdi-l-gds-r05-2012/0.html> (16.03.2013)
- GISWIKI (2014): Mapyrus. <http://giswiki.org/wiki/Mapyrus> (27.08.2014)
- GOOGLE (2013): Google Geocoding API. <https://developers.google.com/maps/documentation/geocoding/> (26.02.2013)

- HAKE G. (1982): Kartographie I – Allgemeines zur Kartographie und zur Karte, Herkunft und Erfassung der Informationen, Kartennetzentwürfe, Merkmale und Mittel kartographischer Gestaltung, topographische Karten. Berlin, New York.
- HAKE G., GRÜNREICH D. UND MENG L. (2002): Kartographie – Visualisierung raum-zeitlicher Informationen. Berlin, New York.
- HENRIOD S. (2013): E-Mail Korrespondenz vom 08.03.2013. Statistik Schweiz.
- IMHOF E. (1972): Thematische Kartographie. Berlin, New York.
- KAMINGER I. (2010a): Auf dem Weg zu einer rasterbasierten Regionalstatistik in Europa. – In: MEINEL G. und SCHUMACHER U. (Hrsg.): Flächennutzungsmonitoring II. Konzepte – Indikatoren – Statistik. IÖR Schriften Band 52. – Dresden, 237-248.
- KAMINGER I. (2010b): Auf dem Weg zu einer rasterbasierten Regionalstatistik in Europa. http://www.efgs.info/presentations/kaminger2.pdf/at_download/file (17.03.2013)
- KAMINGER I. (2010c): Kleinräumige Auswertungen auf Basis von regionalstatistischen Rastern in AT und Europa (GEOSTAT). http://www.efgs.info/presentations/kaminger.pdf/at_download/file (17.03.2013)
- KAMINGER I., KATZLBERGER G. und RAINER N. (2010): Statistical grids as tool for small area analysis. – In: 27th SCORUS Conference. Creating a knowledge base for successful cities and regions, along the statistical value chain. http://scorus.csb.gov.lv/images/stories/papers/4_5_kaminger_katzlberger_rainer.pdf (04.04.2013)
- KAMINGER I. und WONKA E. (o.J.): Regionalstatistik auf der Basis von geographischen Rastern. https://www.statistik.at/web_de/static/artikel Regionalstatistik_auf_der_basis_von_geographischen_rastern_026400.pdf (04.04.2013)
- KAMINGER I. und WONKA E. (2005): Regionalstatistische Gebietsgliederungen nach geographischen Rastern. – In: Statistische Nachrichten 8/2005, 753-760.
- KARTVERKET (2013): Norgeskart. <http://www.norgeskart.no/> (28.03.2013)
- KATZLBERGER G. (2012): E-Mail Korrespondenz vom 27.11.2012. Statistik Austria.
- KELNHOFER F. (1971): Beiträge zur Systematik und allgemeinen Strukturlehre der thematischen Kartographie. – In: Arnberger (Hrsg.): Forschungen zur Theoretischen Kartographie, Band 1. Wien.
- KUZMA I. (2013): E-Mail Korrespondenz vom 01.03.2013. Statistical Office Slovenia.
- LARUE S. (2013): E-Mail Korrespondenz vom 13.03.2013. STATEC Luxembourg.
- LECHTHALER M. und STADLER A. (2006): Kartographische Gestaltung einer bildschirmgerechten Visualisierung von Geobasisdaten. In: KRIZ K., CARTWRIGHT W., PUCHER A. und KINBERGER M. (Hrsg.): Wiener Schriften zur Geographie und Kartographie. Band 17. 248-255.
- LIPATZ J. L. (2011): Gridded data from the French census 2007 – Aggregation without coordinates, coordinates but disaggregation. http://www.efgs.info/data/france/gridded-data-from-the-census/Gridded%20data%20from%20the%20French%20census.doc/at_download/file (25.03.2013)
- LIPATZ J. L. (2013): E-Mail Korrespondenz vom 01.03.2013. National Institute of Statistics and Economic Studies.

- LIND M. (2013): E-Mail Korrespondenz vom 20.03.2013. Ministry of Housing, Urban and Rural Affairs.
- MASIK K. (2011): Production of grid based statistics in Statistics Estonia. http://www.efgs.info/workshops/efgs-2011-lisboa-portugal/efgs-2011-conference-in-lisboa/kreet_masik_grid_production_statistics_estonia_efgs2011.ppt (26.03.2013)
- MAPQUEST (2012): Geocoding Service Developer's Guide. <http://www.mapquestapi.com/geocoding/> (27.02.2013)
- MEIXNER O. und HAAS R. (2012): Wissensmanagement und Entscheidungstheorie: Theorien, Methoden, Anwendungen und Fallbeispiele. – Wien.
- MICROSOFT (2013): Bing Maps REST Services – Locations API. <http://msdn.microsoft.com/de-at/library/ff701715.aspx> (27.02.2013)
- MONMONIER M. (1996): Eins zu einer Million – die Tricks und Lügen der Kartographen. How to lie with maps. Basel.
- MORAVEC Š. (2013): E-Mail Korrespondenz vom 12.03.2013. Czech Statistical Office.
- MOSTRÖM J. (2013): E-Mail Korrespondenz vom 07.03.2013. Statistics Sweden.
- MOVABLE TYPE (2013a): Calculate distance, bearing and more between Latitude/Longitude points. <http://www.movable-type.co.uk/scripts/latlong.html> (28.03.2013)
- MOVABLE TYPE (2013b): Vincenty formula for distance between two Latitude/Longitude points. <http://www.movable-type.co.uk/scripts/latlong-vincenty.html> (28.03.2013)
- NORDBECK O. (2013): GEOSTAT 1B – Representing census data in European population grid. <http://www.efgs.info/geostat/1B/geostat-1b/geostat1b-intermediary-report> (26.03.2013)
- ÖBERG S. und SPRINGFELDT P. (1991): The Population. The National Atlas of Sweden. Statistics Sweden.
- OLBRICH G., QUICK M. UND SCHWEIKART J. (2002): Desktop Mapping – Grundlagen und Praxis in Kartographie und GIS. Berlin, Heidelberg.
- OPENSTREETMAP (2013): Nominatim. <http://wiki.openstreetmap.org/wiki/Nominatim> (27.02.2013)
- PRINZ T., STROBL J. und WONKA E. (2004): Flexible Aggregation regionalstatistischer Erhebungen – Neue Produkte der Statistik Austria. – In: Strobl J., Blaschke T. und Griesebner G. (Hrsg.): Angewandte Geoinformatik 2004. Beiträge zum 16. AGIT-Symposium 2004. – Heidelberg, 556-561.
- REJC I. (2013): E-Mail Korrespondenz vom 19.03.2013. Slowenische Landesvermessung.
- ROONGPIOONSOPIT D. und KARIMI H. A. (2010): Comparative evaluation and analysis of online geocoding services. – In: International Journal of Geographical Information Science 24 (7), 1081-1100.
- SCHNORR-BÄCKER S. (2013): E-Mail Korrespondenz vom 01.03.2013. Statistisches Bundesamt Deutschland.
- SCHULTE B., LIPPmann F. und SCHWEIKART J. (2010): Geokodierung mit Webkartendiensten – Möglichkeiten, Unterschiede und Grenzen. – In: Strobl J., Blaschke T. und Griesebner G.

- (Hrsg.): Angewandte Geoinformatik 2010. Beiträge zum 22. AGIT-Symposium Salzburg. – Berlin und Offenbach, 773-778.
- SEHLIN J. (2011): Production of grid statistics at Statistics Sweden. http://www.efgs.info/workshops/efgs-2011-lisboa-portugal/efgs-2011-conference-in-lisboa/johnny_sehlin_productiongridssweden.docx (26.03.2013)
- SINNOTT R. W. (1984): Deriving the haversine formula. In: Sky and Telescope 68 (2), 159.
- ŠMÍDOVÁ R. (2013): E-Mail Korrespondenz vom 21.03.2013. Czech statistical office.
- SOMMER E. (2007): Presentation of Danish method to take care of confidentiality providing statistical data on grids. http://www.efgs.info/workshops/efgs-2007-helsinki-finland/session2_sommer_presentation.pdf (26.03.2013)
- SOMMER E. (2013): E-Mail Korrespondenz vom 01.03.2013. Statistics Denmark.
- STATISTIK AUSTRIA (2013a): NUTS-Einheiten. http://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/nuts_einheiten/index.html (09.04.2013)
- STATISTIK AUSTRIA (2013b): Regionalstatistische Rastereinheiten. http://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/regionalstatistische_rastereinheiten/ (16.03.2013)
- TAMMILEHTO-LUODE M. (2010): A state of the Art report Finland. http://www.efgs.info/workshops/efgs-2010-tallinn-estonia/efgs2010/13_Status%20report%20WP%202%20Geostatistics%20_Marja%20Tammilehto%20-%20Luode.pdf (22.03.2013)
- TICCA P. (2013): E-Mail Korrespondenz vom 28.02.2013. Italian National Institute of Statistics.
- VALGMA Ü. (2013): E-Mail Korrespondenz vom 05.03.2013. Statistics Estonia.
- VAN LEEUWEN N. F. M. (2013): E-Mail Korrespondenz vom 07.03.2013. Statistics Netherlands.
- WIKIPEDIA (2015): European grid. http://en.wikipedia.org/wiki/European_grid (10.01.2015)
- WONKA E. (2004): Österreich untergliedert in Gemeinden, Zählsprengel oder Raster – Ein regionalstatistischer Vergleich. – In: ÖGG Österreichische Geographische Gesellschaft (Hrsg.): Mitteilungen der Österreichischen Geographischen Gesellschaft 146/2004. – Wien, 345-356.
- WONKA E. (2006): Regionalstatistik in Österreich. Von der Tabelle zu räumlicher Analyse und Visualisierung. – In: Strobl J. (Hrsg.): Salzburger Geographische Arbeiten, Band 39. Salzburg.
- WONKA E. (2010): Regionalstatistik in Österreich auf der räumlichen Bezugsbasis von regionalstatistischen Rastereinheiten. Salzburg.
- WONKA E. (2011): Wienerwaldatlas. – In: Österreichische Akademie der Wissenschaften, Institut für Geographic Information Science und Stadtgemeinde Klosterneuburg (Hrsg.): Siedlungsausbreitung des Ballungsraumes Wien und deren Auswirkungen auf die Wienerwaldgemeinde Klosterneuburg, Band II. Salzburg und Wien.
- WONKA E., KAMINGER I. und KUZMA I. (2007): Regionalstatistisches grenzübergreifendes Projekt zwischen Slowenien und Österreich zum Thema Raster. – In: Statistische Nachrichten 1/2007, 72- 78.

- XU S., FLEXNER S. und CARVALHO V. (2012): Geocoding Billions of Addresses: Toward a Spatial Record Linkage System with Big Data. http://stko.geog.ucsb.edu/gibda2012/gibda2012_submission_2.pdf (03.03.2013)
- YAHOO (2012): Yahoo! BOSS – Pricing. <http://info.yahoo.com/legal/us/yahoo/boss/pricing/> (28.02.2013)
- ZANDBERGEN P. A. (2008): A comparison of address point, parcel and street geocoding techniques. – In: Computers, Environment and Urban Systems 32 (2008), 214-232.
- ZANDBERGEN P. A. (2009): Geocoding Quality and Implications for Spatial Analysis. – In: Geography Compass 3/2, 647-680.

Abkürzungsverzeichnis

AHP	Analytischer Hierarchieprozess
AMA	Agrarmarkt Austria
API	Application Programming Interface
BEV	Bundesamt für Eich- und Vermessungswesen in Österreich
CORINE	Coordination of Information on the Environment – Ein Projekt zur einheitlichen Klassifizierung von Landnutzungsformen
CSS	Cascading Style Sheets – Deklarative Sprache für Stilvorlagen; vor allem zusammen mit HTML und XML eingesetzt.
CSV	Comma-separated Values – Dateiformat mit Werten die durch Beistrich getrennt sind
DOM	Document Object Model ist eine Spezifikation einer Schnittstelle für den Zugriff auf HTML- oder XML-Dokumente
DSS	Decision Support System
EFGS	European Forum for Geostatistics
EFTA	Europäische Freihandelszone (Island, Liechtenstein, Norwegen, Schweiz)
ESSnet	European Statistical System Network
ETRS-LAEA	Das Lambert Azimuthal Equal Area Koordinatensystem basierend auf dem Europäischen Terrestrischen Referenzsystem
EU27	27 Mitgliedsstaaten der Europäischen Union von 2007
EUROSTAT	Statistische Amt der Europäischen Union
GEOSTAT	Projekt mit dem Ziel rasterbasierte Statistiken zu erstellen
GIS	Geographisches Informationssystem
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
INSPIRE	Infrastructure for Spatial Information in the European Community
INVEKOS	Integriertes Verwaltungs- und Kontrollsyste in Österreich
JRC	Joint Research Centre

JSON	JavaScript Object Notation
LAU	Local administrative unit
NGO	Non-governmental Organisation
NUTS	Nomenclature des unités territoriales statistiques (dt. Systematik der Gebietseinheiten für die Statistik)
ÖK	Österreichische Karte
OSM	OpenStreetMap
PHP	rekursives Akronym und Backronym für Hypertext Preprocessor – Skriptsprache mit einer an C und Perl angelehnten Syntax, die hauptsächlich zur Erstellung dynamischer Webseiten oder Webanwendungen verwendet wird
POPREG	Bevölkerungsstatistische Datenbank Österreichs
POI	Point of Interest
URL	Uniform Resource Locator
XML	Extensible Markup Language
ZAMG	Zentralanstalt für Meteorologie und Geodynamik in Österreich
ZMR	Zentrales Melderegister in Österreich

Lebenslauf

Persönliche Daten

Name	Manuel Illmeyer
Titel	Bakk. techn.

Ausbildung

seit 03/2010	Masterstudium „Kartographie und Geoinformation“ an der Universität Wien
03/2010	Abschluss des Bachelorstudiums „Umwelt- und Bioressourcenmanagement“ und Verleihung des Bakkalaureus der technischen Wissenschaften (Bakk. techn.)
10/2006 - 03/2010	Bachelorstudium an der Universität für Bodenkultur, Studienrichtung „Umwelt- und Bioressourcenmanagement“
10/2005 - 09/2006	Zivildienst beim Roten Kreuz, Bezirksstelle Lilienfeld
07/2005	Reife- und Diplomprüfung an der HTL St. Pölten mit gutem Erfolg bestanden
2000 - 2005	HTBLuVA St. Pölten, Abteilung EDV und Organisation
1996 - 2000	BG/BRG Lilienfeld
1992 - 1996	Volksschule Lilienfeld

Fachbezogene / Berufliche Erfahrung

seit 04/2013	Projektleiter im Bereich GIS bei Land-, forst- und wasserwirtschaftliches Rechenzentrum GmbH (LFRZ)
11/2011 - 03/2013	Projektmitarbeiter an der Universität Wien, Institut für Geographie und Regionalforschung, Arbeitsgruppe Kartographie und Geoinformation. Mitarbeit an Projekten vor allem in der Planung, Umsetzung und Programmierung von Web-GIS Applikationen.
10/2011 - 02/2012	Tutor für die Proseminare „Einführung in die Geoinformation“ und „Einführung in die Kartographie“ an der Universität Wien
07/2011 - 08/2011	Praktikum bei „Land-, forst- und wasserwirtschaftliches Rechenzentrum GmbH“

03/2011 - 10/2011	Freier Mitarbeiter bei Freytag-Berndt und Artaria KG
03/2011 - 07/2011	Tutor für die Proseminare „Methoden der GIS-gestützten Datenerfassung“ und „Grundlagen der kartographischen Gestaltung“ an der Universität Wien
02/2011	Praktikum bei Schubert & Franzke GmbH – Kartographischer Verlag
10/2010 - 02/2011	Tutor für das Proseminar „Einführung in die Kartographie“ an der Universität Wien
08/2009 - 09/2009	Berufs-Praktikum bei „Umwelt Management Austria“

Eigenständigkeitserklärung

Ich versichere:

- dass ich die Masterarbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.
- dass alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Publikationen entnommen sind, als solche kenntlich gemacht sind.
- dass ich dieses Masterarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin/ einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe.
- dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Datum

Unterschrift