

Master Thesis

im Rahmen des
Universitätslehrganges „Geographical Information Science & Systems“
(UNIGIS MSc) am Zentrum für GeoInformatik (Z_GIS)
der Paris Lodron-Universität Salzburg

zum Thema

„Möglichkeiten der Verwendung statistischer Rasterdaten für die GIS-gestützte
räumliche Analyse“

vorgelegt von

Dipl.-Ing.(FH) Cornelia Herrmann-Hahn
U1318, UNIGIS MSc Jahrgang 2007

Zur Erlangung des Grades
„Master of Science (Geographical Information Science & Systems) – MSc(GIS)“

Gutachter:
Ao. Univ. Prof. Dr. Josef Strobl

Lünen, Juli 2009

ERKLÄRUNG ÜBER DIE EIGENSTÄNDIGE ABFASSUNG DER ARBEIT

„Ich versichere, diese Master Thesis ohne fremde Hilfe und ohne Verwendung anderer als der angeführten Quellen angefertigt zu haben, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat.

Alle Ausführungen der Arbeit, die wörtlich oder sinngemäß übernommen wurden, sind entsprechend gekennzeichnet.“

Lünen, Juli 2009

Cornelia Herrmann-Hahn

KURZFASSUNG

Es steht außer Frage, dass Daten, vorwiegend erhoben von statistischen Ämtern, für eine Vielzahl von Analysen herangezogen werden. Doch sind diese Analysen, insbesondere, wenn sie von den statistischen Ämtern selbst durchgeführt werden, fachbezogen und gehören in den Bereich der deskriptiven Statistik.

Im Sinne der deskriptiven Statistik werden die erhobenen Daten zusammengetragen, aufbereitet und ausgewertet, wobei sich die anschließenden Analyseergebnisse auf den untersuchten Datensatz beziehen.

Präsentiert werden diese Ergebnisse vorrangig in Tabellen und Diagrammen.

Um die Daten für weitergehende Analysen raumbezogener Zusammenhänge verwenden zu können, ist als erster Schritt das Herstellen eines Raumbezuges erforderlich.

Über ihren Raumbezug können statistische Daten mit anderen Daten kombiniert und verglichen werden, wodurch weitergehende Analysen möglich sind und die Daten einen Mehrwert erhalten.

Als Raumbezug können administrative Grenzen aber auch z.B. quadratische Rasterzellen verwendet werden.

Die vorliegende Arbeit beschäftigt sich mit der Frage, ob und wie statistische Daten für räumliche Analysen mit Hilfe von Geographischen Informationssystemen (GIS) verwendet werden können.

Um aufzuzeigen, welche Möglichkeiten und Grenzen bei der Durchführung einer räumlichen Analyse auf der Basis von statistischen Rasterdaten unter Verwendung von Geoinformationssystemen, bestehen, wird als Beispiel eine Betroffenen-Analyse erstellt.

In dieser Analyse wird mit Hilfe der Daten der Bevölkerungsstatistik des Dortmunder Stadtteils Mengede die vom Lärm (insbesondere Straßenlärm) betroffenen Wohngebiete und damit der Anteil der Betroffenen Einwohner ermittelt.

Um die Unterschiede zwischen einer auf Rasterdaten basierten Analyse und einer Analyse von auf administrativen Bezugseinheiten basierenden Analyse zu veranschaulichen, wird die Betroffenen-Analyse für beide Datengrundlagen durchgeführt.

ABSTRACT

There is no doubt that data, predominantly collected from statistic offices, are used for a multiplicity of analyses.

But these analyses, in particular if they are accomplished by the statistic offices themselves, are subject-based and belong into the range of the descriptive statistics.

According to the descriptive statistics the imposed statistical data are collected, processed and evaluated.

Descriptive statistics are used to describe the basic features of the data in a study in which the following analysis results refer to the examined data set.

The analysis results are presented in tables and diagrams.

To be able to use the data for advanced analyses of spatial connections, giving them a spatial reference is necessary as the first step.

About spatial reference statistic data can be combined and compared with other data, whereby advanced analyses are possible and the data obtain an increase in value.

Mostly administrative boundaries are used to present statistical data but also e.g. square raster cells can be used.

This Master Thesis deals with the question, whether/how statistical data can be used for spatial analyses with the help of geographical information systems (GIS).

In order to point out, which possibilities and limits exist at the time of the execution of a spatial analysis, on the basis of statistical raster data using geo information systems, a "person affected analysis" is made.

In this analysis, the populated areas and thus the portion of the inhabitants concerned by noise (in particular road noise) are determined with the help of the data of the population statistics of the Dortmund district Mengede.

To illustrate the differences between an analysis based on raster data and an analysis based on administrative boundaries, the "person affected analysis" is carried out for the two databases.

Inhaltsverzeichnis

ERKLÄRUNG ÜBER DIE EIGENSTÄNDIGE ABFASSUNG DER ARBEIT	I
KURZFASSUNG	II
ABSTRACT	III
INHALTSVERZEICHNIS	IV
ABBILDUNGSVERZEICHNIS	V
TABELLENVERZEICHNIS	VI
ABKÜRZUNGSVERZEICHNIS	VII
1 EINLEITUNG	1
1.1 Problemstellung.....	1
1.2 Zielsetzung	2
1.3 Methodischer Ansatz und Aufbau der Arbeit	4
2 GIS	5
2.1 Geodaten	8
2.2 INSPIRE	10
3 STATISTISCHE DATEN	12
3.1 Amtliche Statistik	12
3.2 Nichtamtliche Statistik	15
3.3 Koordinatengebundene Statistik	16
3.4 Datenschutz bei Geodaten der Statistik	17
4 DATENSTRUKTUR VEKTORDATEN / RASTERDATEN	22
4.1 Geometrie Vektordaten /Rasterdaten	24
4.2 Zuordnung von Daten zu Rasterzellen	27
5 GIS-GESTÜTZTE RÄUMLICHE ANALYSE	29
6 BEZUGSSYSTEM	33

7	METHODEN ZUR WAHRUNG DES DATENSCHUTZES	36
8	BEISPIELGEBIET	40
	8.1 Daten der Lärmkartierung	41
	8.2 Daten der Einwohnerstatistik	45
	8.3 Verwendetes Bezugssystem	47
	8.4 Ausgangsdaten Vektor	47
	8.5 Ausgangsdaten (Vektor-)Raster	49
9	DARSTELLUNG DER BEVÖLKERUNGSDICHTE	53
	9.1 Bevölkerungsdichte – Bezugsfläche Baublock	54
	9.2 Bevölkerungsdichte – Bezugsfläche Rasterzelle	56
10	ERMITTLUNG DER VON LÄRMBELASTUNG BETROFFENEN BEVÖLKERUNG	59
	10.1 Bezugsfläche Baublock	60
	10.2 Bezugsfläche Rasterzelle	63
	10.3 Bezugsfläche Rasterpixel	64
11	METHODEN ZUR WAHRUNG DES DATENSCHUTZES – ANGEWENDET AUF DAS BEISPIELGEBIET	67
	11.1 Löschen von unterbesetzten Zellen	67
	11.2 Verwendung variabler Zellgrößen	69
12	VERGLEICH DER ANALYSEERGEBNISSE	71
13	ERGEBNISDISKUSSION	77
14	ZUSAMMENFASSUNG / AUSBLICK	80
	LITERATURVERZEICHNIS	83

Abbildungsverzeichnis

Abb. 1 : Darstellung von Geoobjekten im Vektor- und Rasterdatenmodell	25
Abb. 2: Rastermodell – Nachbarschaften	26
Abb. 3: Rastermodell – Abstandmessung	26
Abb. 4: Regionalstatistisches Rasternetz auf der Basis des ETRS1989 LAEA	34
Abb. 5: Darstellung von Rasterdaten als Quadtree	37
Abb. 6: Rasterzellen variabler Größe	38
Abb. 7: Bevölkerungsverteilung - Bezugsfläche Baublock	48
Abb. 8: Adresskoordinaten.....	49
Abb. 9: Bevölkerungsverteilung – Bezugsfläche Raster	50
Abb. 10: fiktiver Datensatz: Zuordnung Gebäudemittelpunkt - Zelle	52
Abb. 11: Differenz zwischen Gebäudeschwerpunkt und Bezugsflächenschwerpunkt ..	53
Abb. 12: Bevölkerungsdichte – Bezugsfläche Baublock	55
Abb. 13: Mengede: Bevölkerungsdichte Wohnbebauung	56
Abb. 14: Zuweisen der Einwohnerdichte der Rasterzellen an die Gebäude	57
Abb. 15: Vergleich Einwohnerdichte	58
Abb. 16: Bezugsfläche Baublock: Mengede: Straßenverkehr - Lärmausbreitung	60
Abb. 17: Bezugsfläche Baublock: Mengede – Bevölkerungsdichte	61
Abb. 18: Bezugsfläche Baublock: Mengede: von Straßenlärm betroffene Siedlungsbereiche	61
Abb. 19: Bezugsfläche Rasterzelle: Bevölkerungsdichte lärmbelasteter Bereiche	63
Abb. 20: Bezugsfläche Pixel: Lärmpegel	64
Abb. 21: Bezugsfläche Pixel: von Straßenlärm betroffene Wohnbevölkerung	65
Abb. 22: Anzahl der Zellen mit unter 10 bzw. unter 30 Einwohnern pro ha	67
Abb. 23: Variable Zellgröße / Feste Zellgröße	69
Abb. 24: Bezugsfläche variable Zellen: Lärmbelastete Bereiche	70
Abb. 25: fiktiver Datensatzvariable – Bevölkerungsdichte	71
Abb. 26: fiktiver Datensatz – von Straßenlärm belastete Bereiche (Bezugsfläche Baublock / Rasterzelle)	72
Abb. 27: fiktiver Datensatz – von Straßenlärm belastete Bereiche (Bezugsfläche variable Zellen)	73
Abb. 28: fiktiver Datensatz: Schwerpunkte der von Lärmbelastung betroffenen Bevölkerung	76

Tabellenverzeichnis

Tabelle 1: Gegenüberstellung Vektormodell / Rastermodell	24
Tabelle 2: ArcGIS Operatoren.....	30
Tabelle 3: Anteil der zu löschenden, unterbesetzten Zellen	68
Tabelle 4: Von Lärm belastete Personen.....	73
Tabelle 5: Betroffenenanzahl bei Anwendung einer Mindestfallzahlregel.....	75

Abkürzungsverzeichnis

BDSG	Bundesdatenschutzgesetz
BImSchV	Bundes-Immissionsschutzgesetz Verordnung
BStatG	Bundesstatistikgesetz
BVerfG	Bundesverfassungsgericht
CRS	Coordinate Reference System
dB	Dezibel
DHDN	Deutsches Hauptdreiecksnetz
EBA	Eisenbahnbundesamt
ESDI	European Spatial Data Infrastructure
EU	Europäische Union
GeoZG	Geodatenzugangsgesetz
GIS	Geographische Informationssysteme
INSPIRE	Infrastructure for Spatial Information in Europe
LAEA	Lambert Azimutal Equal Area
LCC	Lambert Conformal Conic
LAU	Local Administrative Units
NUTS	Nomenclature des unités territoriales statistiques / Systematik der Gebietseinheiten für die Statistik
OGC	Open Geospatial Consortium
ULD	Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein
UTM	Universale Transversale Merkator-Projektion
VBEB	Vorläufige Berechnungsmethode zur Ermittlung der Belastetenzahlen durch Umgebungslärm

1 Einleitung

Es steht außer Frage, dass Daten, vorwiegend erhoben von statistischen Ämtern, für eine Vielzahl von Analysen herangezogen werden. Doch sind diese Analysen, insbesondere, wenn sie von den statistischen Ämtern selbst durchgeführt werden, fachbezogen und gehören in den Bereich der deskriptiven Statistik.

Im Sinne der deskriptiven Statistik werden die erhobenen Daten zusammengetragen, aufbereitet und ausgewertet, wobei sich die anschließenden Analyseergebnisse auf den untersuchten Datensatz beziehen.

Da die Weitergabe von statistischen Daten datenschutzrechtlichen Bestimmungen unterliegt, werden die Daten nur in aggregierter Form an Dritte weitergegeben.

Als Raumbezug werden hierbei administrative Gebietseinheiten verwendet.

1.1 Problemstellung

Im Bereich der Raumplanung stellt die Auswertung von statistischen Daten einen wichtigen Bestandteil dar, räumliche Entwicklungen zu dokumentieren.

Aufgrund ihrer unterschiedlichen Größe, erweisen sich die administrativen Gebietseinheiten als Raumbezug oftmals als wenig geeignet, kleinräumige Entwicklungen aufzuzeigen.

Auch die Veränderlichkeit der Verwaltungsgrenzen kann einen Vergleich der Daten u.a. innerhalb von Zeitreihen erschweren.

Die Alternative stellen Daten auf Grundlage von Raumeinheiten dar, die unabhängig von administrativen Grenzen sind.

Aus diesem Grund wird in den statistischen Ämtern verschiedener Länder Europas die Verwendung von quadratischen Rasterzellen als Raumbezug für statistische Daten diskutiert.

Vorreiter sind hier die Schweiz und Österreich, die schon seit 1970 bzw. 2001 Daten u.a. der Bevölkerungsstatistik auf der Grundlage von Rasterzellen zur Verfügung stellen.

In der vorliegenden Arbeit werden in verschiedenen Analysen raster- und vektordatenbasierende Daten der Einwohnerstatistik verwendet, und die Analyseergebnisse beider Datengrundlagen miteinander verglichen.

Als Datengrundlage dienen hierzu raster- und vektorbasierende Daten der Einwohnerstatistik des Dortmunder Stadtteil Mengede, sowie Daten der Lärmkartierung.

Die Frage ist, wie aussagekräftig kann eine GIS gestützte räumliche Analyse auf Grundlage von rasterbasierenden statistischen Daten sein.

Zur Beantwortung dieser Frage wurde für die vorliegende Arbeit folgende Hypothese aufgestellt:

Hypothese 1

„Mit Hilfe von GIS kann bei Verwendung von Rasterdaten der Einwohnerstatistik die räumliche Verteilung der von Lärmbelastung betroffenen Wohnbevölkerung modelliert werden.“

Da bei der Verwendung von statistischen Daten Datenschutzrichtlinien beachtet werden müssen, schließt sich folgende Hypothese an:

Hypothese 2

„Die Verwendung von auf Raster aggregierte, statistische Daten sichert die Einhaltung der Datenschutzbestimmungen für personenbezogene Daten.“

1.2 Zielsetzung

Die Bevölkerungsverteilung einer zu untersuchenden Region sowie weitergehende soziodemographische Informationen sind in vielen Bereichen der Raumplanung, z.B. Infrastrukturplanungen aber auch für vorbeugende Maßnahmen der Katastrophenplanung von Interesse.

Hierzu stellt die amtliche Statistik Daten zur Verfügung.

Diese Daten beziehen sich meist auf administrative, hierarchisch gegliederte Gebiete. In Einzelfällen werden auch statistische Bereiche, entsprechend dem vom Antragsteller zu analysierenden Gebiet, zusammengestellt.

Die Bereitstellung der Daten, aggregiert auf Rasterzellen, stellt eine Möglichkeit dar, die Daten, aufbereitet in einer standardisierten Form, zur Verfügung zu stellen.

Daher sollen im Folgenden die Möglichkeiten und Grenzen der Verwendung statistischer Daten für GIS-gestützte räumliche Analysen aufgezeigt und diskutiert werden.

Um letztendlich die Eignung rasterbasierter, statistischer Daten bewerten zu können, muss

- 1) die mögliche, geometrische Genauigkeit geprüft werden,
- 2) die Handhabung der Daten untersucht,
- 3) sowie die Interoperabilität der Daten betrachtet werden.

Dazu wird im Einzelnen geprüft, inwieweit sich die Ergebnisse und die Durchführung einer Betroffenen-Analyse, auf Grundlage von Raster- und Vektordaten ähneln oder voneinander abweichen.

Auch wird die Möglichkeit der Einhaltung des Datenschutzes für beide Datenarten betrachtet.

So werden Methoden der Erzeugung von Rasterdaten diskutiert, bei denen es nicht notwendig ist, Zellen zu unterdrücken.

Es ist nicht Ziel dieser Arbeit, die tatsächliche Lärmbelastung der Wohnbevölkerung des Stadtteil Mengede zu analysieren oder zu visualisieren, um von diesen Ergebnissen konkrete Lärmschutz- oder Lärmvermeidungsmaßnahmen abzuleiten.

Auch die Güte der verwendeten Daten der Lärmkartierung soll nicht bewertet werden.

Die verwendeten Datensätze sind lediglich Beispiele, um Methoden der raster- und vektorbasierten Analyse zu testen.

Die Ergebnisse der Analyse werden vorrangig unter diesem Gesichtspunkt diskutiert.

Die Arbeit richtet sich an GIS-Interessierte und GIS-Anwender, die statistische Daten verwenden.

1.3 Methodischer Ansatz und Aufbau der Arbeit

Im ersten Teil der Arbeit werden allgemeine Begrifflichkeiten, die sich mit dem Thema der GIS-gestützten räumlichen Analyse beschäftigen, erläutert.

Anschließend werden amtliche und nicht amtliche, statistische Daten beschrieben und die mit ihrer Verwendung verbundene Datenschutzproblematik.

Nach einem Überblick über die, in einem GIS zu bearbeitenden Datenmodelle, werden Methoden zur Wahrung des Datenschutzes beschrieben.

Der zweite Teil der Arbeit beschäftigt sich mit dem Beispielgebiet, der Visualisierung der Bevölkerungsdichte und den zur Ermittlung der von Lärm betroffenen Bevölkerung durchgeführten Analysen.

2 GIS

[GIS technology] Acronym for *geographic information system*.

An integrated collection of computer software and data used to view and manage information about geographic places, analyze spatial relationships, and model spatial processes. A GIS provides a framework for gathering and organizing spatial data and related information so that it can be displayed and analyzed.

(ESRI GIS Dictionary 2006)

Ein *Geoinformationssystem* dient der Erfassung, Speicherung, Analyse und Darstellung aller Daten, die einen Teil der Erdoberfläche und die darauf befindlichen technischen und administrativen Einrichtungen sowie geowissenschaftliche, ökonomische und ökologische Gegebenheiten beschreiben.

(Bartelme 1995, S.12)

A powerful set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world for a particular set of purposes.

(Burrough 1986, S.6)

Ein Geo-Informationssystem ist ein rechnergestütztes System, das aus Hardware, Software, Daten und den Anwendungen besteht. Mit ihm können raumbezogene Daten digital erfasst und redigiert, gespeichert und reorganisiert, modelliert und analysiert sowie alphanummerisch und graphisch präsentiert werden.

(Bill 2001, S.105)

Es gibt eine Vielzahl von Definitionen für den Begriff Geoinformationssystem.

Den Fokus dessen, was ein Geoinformationssystem ausmacht, legen die verschiedenen Autoren, je nach fachlichem Hintergrund, in einem anderen Bereich.

Während Bartelme festhält, welchem Zweck ein GIS dient, sieht Burrough es als eine Sammlung von Werkzeugen, mit denen raumbezogenen Daten „bearbeitet“ werden können.

Bill beschränkt Geoinformationssysteme auf rechnergestützte Systeme und benennt eine Reihe von Funktionen, die das System zur Verfügung stellen muss.

Als Kernaussage lässt sich aber folgendes festhalten (angelehnt an die Definition von ESRI 2006):

Mit Hilfe von Geoinformationssystemen werden digitale, raumbezogene Daten / Geodaten erfasst, gespeichert, verwaltet, manipuliert und ausgewertet.

Es dient dazu, Raumbeziehungen zu analysieren und Raumprozesse zu modellieren.

Neben der Dokumentation und Analyse, ermöglicht es die Visualisierung räumlicher Phänomene.

Die Definitionen von GIS und somit auch die Ansprüche, die an ein GIS gestellt werden, haben sich dabei im Laufe der Jahre verändert.

Parallel zur Entwicklung der Hardware- und Softwaresysteme und der Möglichkeiten die dadurch entstanden sind, sind auch die Einsatzmöglichkeiten eines GIS gewachsen (vgl. Bartelme 1995, S.7).

Nach Bartelme (1995, S.7) befinden wir uns inzwischen in der 5. Phase, der sich teilweise überlappenden, Entwicklungsstufen der Geoinformation:

Beginnend mit der „*Zeit der Pioniere*“ (Phase 1) (ca. 1955 bis 1975), den Anfängen der Geoinformatik, über die „*Zeit der Behörden*“ (Phase 2) (ca. 1970 bis 1985), in der die Vermessungsverwaltungen begonnen, Geodaten rechnergestützt zu verarbeiten und Phase 3 als „*Zeit der Firmen*“ (ca. 1982 bis 1990), in der Softwarefirmen die Systeme auf die Anforderungen an Geoinformationssysteme zuschnitten, zu Phase 4, der „*Zeit der Nutzer*“ (ab ca. 1988 bis 1995), in der nutzeroptimierte Lösungen für die Verarbeitung der Daten entwickelt wurden.

Die Phase 5 letztlich ist „*Die Zeit des offenen Marktes der Geoinformation*“:

In allen Bereichen von Wirtschaft und (Raum-)Planung werden heute GIS eingesetzt, um Informationen zu gewinnen, die der Entscheidungsfindung dienen.

Dabei wird durch die Verwendung von OGC-konformen Austauschformaten (OGC / Open Geospatial Consortium) und Services der Begriff Geoinformationssystem inzwischen immer mehr durch den Begriff Geo-Services ersetzt.

Die konzeptionellen Grundlagen des GIS beziehen sich auf das von Alfred Hettner (1859 – 1941) 1927 in der Landschaftsforschung eingeführte „Schichtenprinzip“.

Entsprechend dem von ihm beschriebenen „länderkundlichen Schema“ werden die, für die analytische Landeskunde benötigten Sachverhalte, wie z.B. Besiedlung oder Vegetation, in einzelne thematische Schichten aufgeteilt.

Mitte der 1960er Jahre wurde dieses „Layer-Prinzip“ in verschiedenen Software Entwicklungen implementiert.

So entwickelte das Harvard Laboratory for Computer Graphics and Spatial Analysis an der Harvard University, USA, in der Mitte der 1960er Jahre erste Kartographie Programme, die unabhängige Datenebenen verwendeten.

Zur gleichen Zeit wurde in Kanada unter der Leitung von Roger Tomlinson das Canadian Geographic Information System (CGIS) entwickelt, zur Erfassung und Verwaltung der Landressourcen, aber auch zur Unterstützung der Landschaftsplanung.

Ein weiterer Meilenstein in dieser Zeit war die Gründung des Environmental Science Research Institute (ESRI) 1969 in Kalifornien.

Durch diese Trennung der Geometriedaten einschließlich ihrer Attributdaten in verschiedene thematische Informationsebenen (Layer) entstehen vielfältige Möglichkeiten der Separation und Kombination der Daten (vgl. Bill 2001, S.77).

Diese Layer-Technik wird auch heute noch verwendet und stellt die Grundlage dar für die in GIS-Systemen implementierte „overlay“-Technik, die die Verschneidung von Polygonen ermöglicht.

Der entscheidende Unterschied zwischen einem Geoinformationssystem und einem graphisch orientierten CAD-System, ist die Verwaltung der Sachdaten:

Die räumlichen Daten und ihre Attribute werden in einem GIS innerhalb eines Datenbanksystems in separaten Dateien gespeichert, die über einen Identifikator oder ein Schlüsselattribut miteinander verknüpft sind.

So kann der Benutzer ein Objekt über seine Geometrie auswählen, und anschließend über die Attributtabelle Informationen über das Objekt erhalten.

Er kann aber auch über ausgewählte Attribute zu bestimmten Objekten gelangen.

Somit kann in einem GIS ein Element über eine thematische Abfrage anhand eines Attributes / Eigenschaft identifiziert werden oder anhand seiner Lage im Raum.

Damit stellt das GIS das Handwerkszeug zur Verfügung.

Entscheidend für die Aussagekraft der Analyse Ergebnisse sind aber nicht nur die Art der durchgeführten Analyse, sondern auch die Qualität der verwendeten (Input-) Daten:

Je kleinräumiger die statistischen Daten, die für die Analyse verwendet werden, sind desto präziser können die Ergebnisse der Analyse sein.

2.1 Geodaten

Unter Geodaten versteht man Daten über Gegenstände, Geländeformen oder Infrastrukturen der Erde, die einen Raumbezug aufweisen.

Dazu zählen Geometriedaten, die die Lage und Form von Objekten im Raum beschreiben, topologische Daten, die die räumliche Beziehung von Objekten (Topologie) darstellen, aber auch Sachdaten, die der Beschreibung des thematischen Inhaltes des Objektes dienen. (vgl. Bill et.al. 2001, S.106)

Auf Grundlage der INSPIRE Direktive ist der Zugang zu digitalen Geodaten im Geodatenzugangsgesetz (GeoZG) vom 10.02.2009 festgehalten.

Es regelt den Zugang zu vorhandenen Geodaten, Geodatendiensten und Metadaten der Bundesbehörden und Anstalten öffentlichen Rechts der Bundesrepublik Deutschland.

Geodaten werden in zwei Bereiche unterteilt: die Geobasisdaten und die Geofachdaten.

Zu den Geobasisdaten zählen vorrangig die Daten der amtlichen Vermessungsverwaltungen, die die Topographie beschreiben.

Fachdaten sind nichtgeometrische Daten oder Attributdaten (z.B. demografische oder sozioökonomische Angaben), die themabezogen die Eigenschaften einer Entität beschreiben.

Zu den Geofachdaten gehören alle thematischen Daten, die einen Raumbezug aufweisen.

Hierbei kann der Raumbezug direkt über geografische Koordinaten oder andere Gitterangaben, aber auch indirekt über die Anschrift (Ort, Straße) oder einen amtlichen Gemeindegemeinschaftsschlüssel gegeben sein.

Die in dieser Arbeit verwendeten Einwohner-Daten der amtlichen Statistik gehören in den Bereich der Geofachdaten. Über die jeweilige Adresse ist den Daten eine geographische Koordinate zugewiesen worden, wodurch die Daten einen eindeutigen Raumbezug erhalten haben.

Für jede Datenerhebung oder –analyse ist es erforderlich, die Qualität der verwendeten Daten zu dokumentieren, dies erfolgt über die Metadaten.

Nur so lässt sich auch die Güte der Analyseergebnisse beurteilen.

Qualitätskriterien sind 2002 von der International Organisation for Standardization (ISO) als ISO-Norm 19113 festgehalten (ISO 19113 data quality elements and subelements).

Sie umfassen

- Completeness (Vollständigkeit)
- Logical consistency (Logische Konsistenz)
- Positional accuracy (Positionsgenauigkeit)
- Temporal accuracy (Zeitliche Genauigkeit)
- Thematic accuracy (Thematische Genauigkeit)

Vollständigkeit: Vorhandensein oder Fehlen von Objekten, ihrer Attribute oder Beziehungen

Logische Konsistenz: Einhalten von logischen Regeln der konzeptionellen, logischen und physikalischen Datenstruktur

Positionsgenauigkeit: Genauigkeit der Lage von Objekten

Zeitliche Genauigkeit: Genauigkeit von Zeitangaben und der zeitlichen Beziehungen der Objekte

Thematische Genauigkeit: Genauigkeit von quantitativen und nicht-quantitativen Attributen, Zuordnung der Objekte zu Objektklassen, sowie die Richtigkeit dieser Zuordnungen

(vgl. Müllegger 2007, S.10f.)

2.2 INSPIRE

Im Mai 2007 wurde die EU-Richtlinie mit dem Titel „INSPIRE- Infrastructure for Spatial Information in Europe“ erlassen.

Ursprünglich initiiert, um Geodaten für eine nachhaltige Umweltpolitik verfügbar zu machen, verpflichtet die INSPIRE Richtlinie als Rahmengesetz alle EU-Länder zum Ausbau einer eigenen nationalen Geodateninfrastruktur.

Dabei steht die Harmonisierung von Geodaten, Metadaten, Diensten und Standards im Vordergrund.

Es ist vorgesehen, bis Mitte 2009 nationale Gesetze zu erlassen, die die verpflichtende Bereitstellung der nationalen Geobasisdaten durch die öffentlichen Verwaltungen des Bundes, der Länder und der Kommunen ermöglichen und regeln.

Ziel ist es, den Aufbau einer europäischen Geodateninfrastruktur (European Spatial Data Infrastructure = ESDI) bis 2019 abzuschließen.

Die Richtlinie teilt die Geodaten in drei verschiedene Themenbereiche:

Anhang I (Fertigstellung der Implementierungsrichtlinien bis 15.05.2009, Umsetzung geplant bis 2010):

Koordinatenreferenzsysteme, Geogittersysteme, Geographische Namen, Administrative Grenzen, Transportwege, Hydrographie, Schutzgebiete.

Anhang II (Fertigstellung der Implementierungsrichtlinien bis 15.05.2012, Umsetzung geplant bis 2013):

Höhenmodelle, Adressen, Grundstücke, Topographie, Orthophotos.

Anhang III (Fertigstellung der Implementierungsrichtlinien bis 15.05.2012, Umsetzung geplant bis 2013):

Statistische Einheiten, Gebäude, Bodenschätzung, Geologie, Landnutzung, Gesundheit, industrielle und agrarische Produktionsstätten, öffentliche Einrichtungen, Bevölkerung, Meteorologie, etc.

Der, durch die INSPIRE-Direktive gelenkte Aufbau einer einheitlichen Geodateninfrastruktur innerhalb der Mitgliedsstaaten der EU (Europäische Union), ist damit auch eine Aufforderung an die amtliche Statistik, geeignete Geofachdaten zur Verfügung zu stellen. (vgl. von Eschwege et.al. 2006, S.135)

Wie Anhang I, Ziffer 2 zeigt, ist die Definition eines harmonisierten Rastersystems multipler Zellgröße mit festem Ursprung in einer europaweit anzuwendenden Projektion das Ziel:

*“ANNEX I
SPATIAL DATA THEMES REFERRED TO IN ARTICLES 9(A), 13(1) AND
14(A)
1. Coordinate reference systems
Systems for uniquely referencing spatial information in space as a set of
coordinates (x,y,z)
and/or latitude and longitude and height, based on a geodetic horizontal and
vertical datum
2. Geographical grid systems
Harmonised multi-resolution grid with a common point of origin and
standardised location and size of grid cells.”
(vgl. commission of the european communities 2004)*

Durch Umsetzung der Richtlinien der INSPIRE Direktive werden die Voraussetzungen geschaffen für einen reibungslosen Austausch von Geodaten, innerhalb der einzelnen europäischen Länder, aber auch länderübergreifend. Zur Harmonisierung der EU-weit vorhandenen Daten, ist eine Reihe von technischen Standards definiert, die die Interoperabilität der Daten sichern.

3 Statistische Daten

Statistische Daten werden in vielen Bereichen als Handlungsgrundlage oder zur Kontrolle von Entwicklungen oder durchgeführter Maßnahmen erhoben.

Die amtliche Statistik, erhoben vom Amt für Bundesstatistik, den statistischen Landesämtern oder den statistischen Ämtern der Kommunen hat den Auftrag, statistische Informationen auf wissenschaftlicher Grundlage zu erheben, aufzubereiten, zu analysieren und darzustellen, die objektiv und unabhängig sind.

Durch die Ergebnisse werden gesellschaftliche, ökologische und wirtschaftliche Zusammenhänge aufgeschlüsselt und dienen so dem Bund, den Ländern und Kommunen als Planungsgrundlage.

In den Bereich der nichtamtlichen Statistik fallen Unternehmensstatistiken, erhoben von Unternehmen oder Verbänden vorrangig im Bereich Geomarketing, oder forschungsorientierte Statistiken der Wirtschaftsforschungsinstitute.

Daneben gibt es noch eine Reihe von kommerziellen Instituten, wie Marktforschungs- oder Meinungsforschungsinstitute, die statistische Daten erfassen.

3.1 Amtliche Statistik

Um statistische Informationen, die einen Raumbezug aufweisen, eindeutig einer räumlichen Bezugseinheit zuordnen zu können, wurde für die Mitgliedsländer der europäischen Union 1980 vom Europäischen Amt für Statistik in Luxemburg ein hierarchisches System zur eindeutigen Identifizierung entwickelt.

Seither werden in Europa regionale Raumeinheiten in sog. NUTS- Regionen (fr. Nomenclature des unités territoriales statistiques / Systematik der Gebietseinheiten für die Statistik) unterteilt.

Dieses System ist hierarchisch aufgebaut und verwendet 4 regionale Hierarchieebenen, sowie zwei lokale Hierarchieebenen

Wobei z.B. NUTS 3 in Deutschland 429 Regionen: Landkreise, Kreise oder kreisfreie Städte beschreibt.

Mit LAU 1 (Local Administrative Units) bzw. LAU 2 werden Gemeindeverbände bzw. Gemeinden bezeichnet.

Vorrangig beziehen sich die Daten der amtlichen Statistik auf diese hierarchisch gegliederten, administrativen Gebietseinheiten oder auf Gebiete, die nach fachlichen Kriterien aggregiert wurden.

Die nach administrativen Grenzen gegliederten Daten werden primär für kommunale Planungen verwendet, da hier regionalstatistische Verteilungen und Regionen abgebildet werden können (vgl. Wonka 2008, S.1f).

Die amtliche Statistik selbst beschränkt sich meist auf die kartographische Darstellung der Daten (vgl. Szibalski 2006, S.207).

Thematische Karten, die Daten der amtlichen Statistik präsentieren, sind vorwiegend Choroplethenkarten.

In diesen Karten werden die verwendeten Gebietseinheiten, entsprechend dem für das jeweilige Gebiet ermittelten relativen Zahlenwert, farblich abgestuft eingefärbt.

Wie gut statistische Daten geeignet sind, raumplanerische Entscheidungsfindungen zu unterstützen, hängt aber nicht nur von der Vielfalt oder vom Umfang der Daten ab, sondern auch von der verwendeten räumlichen Bezugseinheit (vgl. Wonka 2008, S.1).

Die kleinste räumliche Auflösung, die für räumliche Analysen öffentlich zugänglich ist, stellen statistische Unterbezirke, vereinzelt auch Baublöcke dar.

Baublöcke setzen sich aus Baublockseiten zusammen. Bei der Darstellung werden öffentliche Straßen nicht als Baublockfläche definiert.

Entsprechend dem Bundesstatistikgesetz §10 Abs. 2 ist die Blockseite die kleinste räumliche Einheit, für die dauerhaft Informationen gespeichert werden dürfen.

„Blockseite ist innerhalb eines Gemeindegebiets die Seite mit gleicher Straßenbezeichnung von der durch Straßeneinmündungen oder vergleichbare Begrenzungen umschlossenen Fläche“ (BStatG §10, Abs. 3).

Als Bezugseinheit ist die Blockseite allerdings weniger geeignet, da die Seiten in der Länge variieren und, eng mit der Geometrie der Straßen verbunden, ebenfalls verändert werden, wenn sich an der Verlauf der Straße ändert. (von Eschwege et.al. 2006, S.129).

So ist auch die Verwendung einer geometrischen Gebietsgliederung, die Verwendung von geographischen Rastern, denkbar.

Für alle in der amtlichen Statistik erhobenen Individualdaten, die Adressdaten zugeordnet werden können, kann durch eine Georeferenzierung über die Verknüpfung mit Koordinaten ein Raumbezug hergestellt werden.

Diese koordinatengebundenen Daten können anschließend auf Rasterzellen aggregiert werden.

Raster stellen, gegenüber den administrativen Gebietseinheiten, durch ihre einheitliche Flächengröße eine neutrale Bezugsgröße dar.

Sie sind hierarchisch unterteilbar und können in unterschiedlichen Zellengrößen definiert werden.

Der Lagefehler der dargestellten statistischen Inhalte kann dabei maximal die Größe der Rasterweite erreichen, während er bei der Verwendung von administrativen Bezugseinheiten je nach Größe der Bezugseinheit variiert und mehrere Kilometer betragen kann (vgl. Wonka 2008, S.2). Das Raster kann entweder koordinatengebunden gewählt oder koordinatenunabhängig frei platziert werden.

Bei einem koordinatengebundenen Raster müssen zur eindeutigen Identifizierung der Rasterzellen im Raum, das verwendete geodätische Bezugssystem, die Rasterweite sowie der Ursprung des Rastergitters bekannt sein.

In einigen Ländern Europas werden bereits Rasterdaten der amtlichen Statistik in unterschiedlichen Auflösungen erstellt und der Öffentlichkeit zur Verfügung gestellt. So werden in der Schweiz, Österreich und den Niederlanden, sowie in Estland, Slowenien und den skandinavischen Ländern Dänemark, Norwegen, Schweden und Finnland auf Raster aggregierte statistische Daten verwendet und publiziert. Für Personendaten werden dabei in allen Ländern andere datenschutzrechtliche Bestimmungen angewendet als bei Daten der Unternehmensstatistik.

Um die statistische Geheimhaltung zu gewährleisten, werden in allen Ländern Rasterzellen, die mit weniger, als der vorher festgelegten Mindestanzahl von Fällen belegt sind, gesperrt.

Die Höhe dieser Mindestfallzahl variiert allerdings, ebenso wie die verwendete Rasterweite.

Die Schweiz z.B. verwendet eine Rasterweite von 100m und eine Mindestfallzahl von 4. Wobei allerdings sensible Merkmale durch eine Vorauswahl entfernt werden.

Estland dagegen stellt Daten mit einer Rasterweite nicht unter 500m zur Verfügung und sperrt nur die Zellen, die eine Fallzahl von unter 3 aufweisen (vgl. Szibalski 2007, S.142).

In den genannten Ländern ist die dauerhafte Speicherung von Adressdaten im Zusammenhang mit Personendaten erlaubt.

Um rasterbasierte Kartierungsmöglichkeiten für statistische und Landnutzungsdaten besser harmonisieren zu können, haben die skandinavischen Länder den European Grid Club gegründet.

Der European Grid Club hat sich das Ziel gesetzt, alle europäischen Länder zum Speichern der statistischen Daten auf Rasterbasis zu motivieren.

Wobei sich ihr Bestreben auf den kleinmaßstäbigen Bereich bezieht und die Größe der Rasterzellen 1km x 1km betragen soll.

3.2 Nichtamtliche Statistik

Außerhalb der amtlichen Statistik, findet die Darstellung raumbezogener Statistikdaten in Karten in verschiedenen Bereichen Verwendung.

So werden nicht nur in Schulatlanten statistische Daten in Karten präsentiert, auch im Bereich des Geomarketing werden (eigene) statistische Erhebungen raumbezogen visualisiert und analysiert (vgl. Szibalski 2006, S.206).

Hierbei steht die Standortsuche und –optimierung im Mittelpunkt.

Die Unternehmen verwenden vielfach adressbezogene Informationen aus Kundenbefragungen oder entsprechenden Kundendatenbanken, um Erkenntnisse zur Optimierung ihrer Logistik, Filialnetzoptimierung oder Standortwahl zu gewinnen (vgl. Strobl 2005, S.164).

Entscheidend für die Aussagekraft dieser Analysen ist die Definition geeigneter Raumbezugseinheiten.

Neben Gemeindegrenzen oder Mikromarktabgrenzungen ist hier die Verwendung von geometrischen Flächeneinheiten (Quadratgitter, Rechteck- oder Sechseckgitter) denkbar. Ziel ist das Herausfinden von Gebieten, in denen sich die Käuferstruktur als homogen erweist. Dazu werden häufig Point Pattern Analysen verwendet.

Mit Hilfe dieser Analysen werden die punkthaft vorliegenden Daten auf Punktmuster untersucht, um die räumliche Verteilung von Ereignissen aufzuzeigen und die Ursache der Punktverteilung zu erkennen.

Dabei wird zwischen Effekten erster Ordnung (first order effects) und Effekten zweiter Ordnung (second order effects) unterschieden. Die Effekte erster Ordnung beziehen sich auf die Anzahl der Ereignisse pro Fläche, untersuchen somit die Dichte.

Als Effekte zweiter Ordnung werden die Beziehungen der Punkte zueinander bezeichnet.

So können durch Segregation Marketingstrategien entwickelt werden.

(vgl. Dickmann et. al. 2008, S.228).

3.3 Koordinatengebundene Statistik

Voraussetzung für die Erzeugung von rasterbezogenen, statistischen Daten ist das Vorhalten von koordinatengebundenen statistischen Datenbanken (vgl. Wonka 2008, S.14).

In der Bundesrepublik Deutschland schreibt das Bundesstatistikgesetz (BStatG) vor, dass der Adressbezug bei der Weiterverwendung der Daten so früh wie möglich gelöscht werden muss.

Aus diesem Grund gibt es Überlegungen, bei der Erstellung von rasterbasierten, statistischen Daten den einzelnen Fallzahlen anstelle einer Gebäudekoordinate (Adresskoordinate) die Bezeichnung (Kennziffer) der Gitterzelle anzuhängen, der sie geographisch zuzuordnen sind.

So wäre die räumliche Zuordnung der einzelnen Gitterzellen nur über die Kennziffer möglich (vgl. Szibalski 2007, S. 138).

Eine dauerhafte Speicherung von Adressdaten im Zusammenhang mit Personendaten ist in der Bundesrepublik Deutschland nicht erlaubt.

In Österreich z.B. arbeitet die STATISTIK AUSTRIA mit koordinatengebundenen Datenbanken, in denen nicht nur Bezeichnung der Rasterzelle, in der das jeweilige Gebäude liegt, sondern auch die Koordinaten der Gebäudemittelpunkte abgespeichert sind.

Die Bevölkerungsdichte je Zelle wird so als Summe aller statistischen Daten mit gleicher Rasterbezeichnung ermittelt (vgl. Wonka 2007, S. 14).

3.4 Datenschutz bei Geodaten der Statistik

Im sog. Volkszählungsurteil des Bundesverfassungsgerichts (BVerfG) vom 15.12.1983 wird das Recht des Einzelnen auf „informationelle Selbstbestimmbarkeit“ festgehalten und definiert, wie und in welcher Form personenbezogene Daten erhoben, gespeichert oder weitergegeben werden dürfen.

Ob und inwieweit bei der Bearbeitung oder Weitergabe von Geofachdaten oder allgemein Geodaten das Datenschutzrecht zur Anwendung kommen muss, hängt davon ab, ob die gespeicherten Informationen einen Personenbezug aufweisen.

Geodaten sind raumbezogene Informationen über Gegenstände, Geländeformen oder Infrastrukturen der Erde. Sie gehören damit in den Bereich der Sachdaten.

Einen Personenbezug erhalten sie erst, wenn sie mit anderen Informationen angereichert werden. Hierbei stellt die direkte Verknüpfung von Namen oder Adressinformationen mit den Geodaten eindeutig das Vorhandensein eines Personenbezuges dar.

Schwieriger zu beurteilen, ob ein Personenbezug vorliegt ist es, wenn Geodaten (ursprünglich ohne Personenbezug) mit anderen Daten, wie z.B. soziodemographischen Informationen über Einkommen oder Bildung verbunden werden.

Erlaubt ist gemäß Bundesdatenschutzgesetz (BDSG) § 28, Abs. 1 Nr.3 das Speichern und Verarbeiten personenbezogener Daten, „wenn die Daten allgemein zugänglich sind oder die verantwortliche Stelle sie veröffentlichen dürfte“.

Sprechen allerdings schutzwürdige Interessen des Einzelnen gegen die Verarbeitung, so muss abgewogen werden, ob diese Interessen höher zu bewerten sind, als die Interessen der verarbeitenden Stelle.

Erlaubt ist ebenfalls die Erhebung, Verarbeitung und Nutzung personenbezogener Daten, wenn der Betroffene seine Einwilligung gegeben hat und über den Zweck der Datenerhebung informiert worden ist. Daher ist der vorgesehene Verwendungszweck der Daten immer mit anzugeben. (vgl. BDGS § 4 und § 4a, S.5f).

Laut BDSG § 3 Abs. 1 sind „Personenbezogene Daten [...] Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlichen Person (Betroffener).“ (BDSG Fassung 1990, S.4)

Die Bestimmbarkeit wird dabei als gegeben angesehen, sobald nur die Möglichkeit besteht, aus vorliegenden Daten eine Verbindung zu einer Person herzustellen.

Wobei die reine Bestimmbarkeit, also z.B. die Zuordnung von Adressinformationen zu Hauskoordinaten nicht als ausschlaggebendes Merkmal betrachtet werden kann, zwischen Sachdaten und personenbezogene Daten zu unterscheiden (vgl. Forgó et.al. 2008). In vielen Bereichen werden Programme zum Geokodieren von Adressen angeboten, damit ist die Zuweisung von Koordinaten zu einer Adressinformation so einfach wie noch nie.

Vielmehr steht hier die Zuordnung von Angaben zu oder über einzelne Personen im Vordergrund.

Der Personenbezug kann als gegeben betrachtet werden, wenn die Daten „... die Identität, die Merkmale oder das Verhalten einer Person betreffen oder wenn sie verwendet werden, um die Art festzulegen oder zu beeinflussen, in der die Person behandelt oder beurteilt wird“ (Artikel-29-Datenschutzgruppe 2007, S. 11).

Ob ein Personenbezug hergestellt werden kann, wird durch den Inhalt der Daten und den Zweck, zu dem sie erhoben oder zusammengetragen worden sind, mitbestimmt.

So kann die Bestimmbarkeit mit Hilfe von drei Kategorien beurteilt werden:

- Inhaltskontext
- Zweckkontext
- Ergebniskontext

Geodaten, die eindeutig Informationen über die Identität einzelner Personen enthalten, stehen in einem Inhaltskontext zur Person.

Ein Zweckkontext der Daten liegt vor, wenn Daten verwendet werden können, z.B. um einzelne Personen zu beurteilen oder ihr Verhalten zu beeinflussen. Sie beschreiben zwar nicht die Person selbst, geben aber Auskunft über das Umfeld der Person.

Geodaten, die einen Ergebniskontext aufweisen, wirken sich auf die Rechte und Interessen von einzelnen Personen aus oder könnten sich darauf auswirken.

Ein Personenbezug der Daten liegt vor, wenn die Daten in eine dieser drei Kategorien fallen.

Geodaten hingegen, die in keinem Zusammenhang zu den genannten Kontexten zu einer Person stehen, unterliegen nicht den Datenschutzbestimmungen.

Damit entscheidet die Beantwortung der Frage, welche Möglichkeiten bestehen, die vorhandenen Informationen direkt einer einzelnen Person zuzuordnen, über die Verarbeitung und/oder Weitergabe von Daten.

Um das Persönlichkeitsrecht des Einzelnen zu schützen, sieht das Datenschutzgesetz vor, Daten zu anonymisieren:

„Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können.“ (BDSG § 3 Abs.6, S.4)

Eine weitere Maßnahme, die Herstellung eines Personenbezuges zu verhindern, ist das Zusammenfassen von Individualdaten zu Gruppen.

Schon das Zusammenfassen von 4 Personen zu einer Gruppe, lässt die Zuordnung eines Merkmals zu einer einzelnen Person mit einer Wahrscheinlichkeit von nur 25% zu (vgl. Weichert 2007, S.19).

Es ist allerdings im BDSG nicht genau geregelt, von wie vielen Personen man die Daten zusammenfassen muss, um durch diese Aggregation den Rückschluss auf Einzelangaben zu verhindern. Während Weichert (2007) eine Aggregation von 4 Personen/Grundstücken für ausreichend hält, weist Karg (2008, S.23) im ULD-Gutachten (Gutachten herausgegeben vom Unabhängigen Landeszentrum für Datenschutz Schleswig-Holstein) darauf hin, dass Aufsichtsbehörden eine Aggregation von 10 Grundstücken verlangen.

Um zu beurteilen, ob Daten datenschutzrechtlichen Bestimmungen unterliegen, sollten daher folgende Fragen beantwortet werden (nach Forgó et.al. 2008):

1. Sind einzelne Personen bestimmbar, können sie ohne großen Aufwand identifiziert werden?
2. Handelt es sich bei der vorliegenden Information um eine Einzelangabe oder einer Sammelangabe, die nur einer Person zugeordnet wird?
3. Stehen die Informationen in einem Inhalts-, Zweck- oder Ergebniskontext zu einer Person? Sind die Informationen geeignet, um eine Aussage über eine einzelne Person zu treffen?

Können alle drei Fragen verneint werden, besteht kein Personenbezug, der Datenschutz findet keine Anwendung.

Muss eine der Fragen bejaht werden, liegt ein Personenbezug vor und die Daten dürfen nur weiterverarbeitet werden, wenn die Einwilligung des Betroffenen vorliegt oder eine gesetzliche Ermächtigung die Datenverarbeitung erlaubt.

Um die primäre und sekundäre Geheimhaltung von Einzeldaten in den Tabellen der amtlichen Statistik zu gewährleisten, werden verschiedene Methoden angewendet.

So werden zur Wahrung der primären Geheimhaltung Einzelangaben in den Tabellen gelöscht oder gesperrt.

Die sekundäre Geheimhaltung erfordert Maßnahmen, die dafür sorgen, dass primär geheim zu haltende Angaben nicht über Summen- oder Differenzbildung ermittelt werden können.

Dazu werden die Daten z.B. durch Mittelwertbildung oder Runden der Tabelleneinträge verändert.

In der nichtamtlichen Statistik können sich Unternehmen, die im Rahmen ihres Direktmarketing personenbezogenen Daten in (eigene) Statistiken verwenden, auf das sog. „Listenprivileg“ berufen.

Laut § 28 des BDSG ist „das Erheben, Speichern, Verändern oder Übermitteln personenbezogener Daten oder ihre Nutzung als Mittel für die Erfüllung eigener Geschäftszwecke [...] zulässig

[...] wenn es der Zweckbestimmung eines Vertragsverhältnisses oder vertragsähnlichen Vertrauensverhältnis mit dem Betroffenen dient“.

Personenbezogenen Daten dürfen „für Zwecke der Werbung, der Markt- und Meinungsforschung“ verwendet werden, „wenn es sich um listenmäßig oder sonst zusammengefasste Daten über Angehörige einer Personengruppe handelt“ und sich die gespeicherten Daten lediglich auf Name, Titel, akademischer Grad, Beruf, Anschrift, Geburtsjahr und die Zugehörigkeit des Betroffenen zu dieser Personengruppe beziehen.

Wichtig ist, dass kein schutzwürdiges Interesse des Betroffenen verletzt wird.

Die Bundesregierung hat am 10. Juli 2009 die Datenschutznovelle II (Datenhandel) verabschiedet. Befürchtungen, das Listenprivileg würde gänzlich abgeschafft, sind damit nicht eingetreten.

Allerdings muss künftig die Herkunft der verwendeten Adressen deutlich angezeigt werden.

Im Gegensatz zu den Daten der Bevölkerungsstatistik unterliegen die Daten des Unternehmensregisters weniger datenschutzrechtlichen Bestimmungen.

Auf Basis des Unternehmensregisters werden Wirtschafts- und Umweltstatistiken erstellt, aus denen wirtschaftliche Entwicklungen abgeleitet werden.

Das Bundesstatistikgesetz (BStatG) erlaubt hier die dauerhafte Speicherung von Adressdaten, somit ist es möglich, auch die Adresskoordinate dauerhaft hinzuzufügen. (vgl. BStatG 1987, §13)

Bei den für diese Arbeit verwendeten Daten der Einwohnerstatistik, handelt es sich um relativ unsensible Daten. Es wurde die Gesamtzahl der Wohnbevölkerung verwendet, keine Unterscheidungen nach Alter oder Geschlecht, religiöse Zugehörigkeit oder Schulbildung.

Im Sinne der Statistik sensible Daten liegen immer dann vor, wenn eine geringe Grundgesamtheit vorliegt.

Wenn also bei einer Analyse aus dem gesamten, zur Verfügung stehenden Datensatz durch Auswahlkriterien nur eine kleine Anzahl selektiert wird.

U.U. werden dabei Attribute lediglich einer einzelnen Person zugewiesen.

Da aus Gründen des Datenschutzes personenbezogene Daten nicht an Dritte weitergegeben werden dürfen, kann das für Daten mit geringer Grundgesamtheit

bedeuten, dass sie, entsprechend der primären oder sekundären Geheimhaltungspflicht, unterdrückt werden und aus dem Gesamtdatensatz entfernt werden.

Nur so ist zu verhindern, dass die jeweiligen Daten einer einzelnen Person zugeordnet werden können.

4 Datenstruktur Vektordaten /Rasterdaten

In Geographischen Informationssystemen (GIS) werden die zu speichernden Phänomene der realen Welt (Entitäten) nach ihrer Struktur unterschieden:

- kontinuierliche Entitäten, die eine flächenhafte Ausbreitung im Raum haben und

- diskrete Entitäten, eindeutig abzugrenzende Flächen, linienhafte Phänomene wie Straßenverläufe oder Gewässerlinien, sowie punkthafte Informationen.

Zur Speicherung dieser Datenstrukturen werden unterschiedliche Datenmodelle verwendet.

Kontinuierlich auftretende Entitäten werden in einem Rastermodell / Grid gespeichert, diskrete Entitäten in einem Vektormodell.

So können in einem Geoinformationssystem Vektordaten, wie auch Rasterdaten verarbeitet werden.

Die meisten mit einem GIS durchgeführten Arbeiten sind vektororientiert, daher nehmen die Vektordaten und die Bearbeitung von Vektordaten den mengenmäßig größeren Teil ein.

Im Vektormodell werden räumlich diskrete Objekte abgebildet.

Objekte / Entitäten, wie z.B. Straßenverläufe, Gebäude oder Gewässer, sind im Raum eindeutig abzugrenzen.

Sie werden mit Anfangs- und Endpunkt, sowie den dazwischenliegenden gerichteten Strecken (Vektoren) beschrieben.

Somit wird in einem Vektormodell die Lage von Punkten innerhalb eines Koordinatensystems dargestellt, sowie die Verbindungen zwischen den Punkten.

Die Darstellung des räumlichen Diskreta erfolgt mittels einer Reihe von sog. Features.

Zu diesen Features gehören die elementaren, geometrischen Elemente:

Punkt, Linie, und Fläche.

Diese geometrischen Primitiva lassen sich durch Gruppierung zu höherwertigen Geometrien zusammenfassen.

So stellt die Linie die Verbindung zwischen den Punkten dar und die Fläche ein Gebiet, welches von Linienzügen begrenzt wird.

Darüberhinaus gehören auch Text-Features, Komplexe Features oder Netz-Features zu den Vektordaten.

Diesen Features können aber nicht nur geographische Koordinaten für die jeweiligen Eckpunkte zugeordnet werden, es können ihnen auch beliebig viele Attribute zugewiesen werden.

Die Geometrie der dargestellten Objekte ergibt sich aus den Koordinaten der Eckpunkte.

Die topologische Beziehung zwischen den Objekten muss dabei explizit gespeichert werden: Nur über die Angabe, ob Koordinaten aufeinander folgen, kann erkannt werden, ob sie benachbart sind und/oder eine Linie bilden oder eine Fläche umgrenzen.

Rasterdaten/Griddaten werden oft nur in Form von Bilddateien benutzt, wie Orthophotos oder gescannte, topographische Karten, die als Hintergrundlayer im GIS verwendet werden.

In einem Rastermodell wird der Untersuchungsraum in gleich große, regelmäßig angeordnete Gittermaschen unterteilt (Matrix). Es dient so der Darstellung kontinuierlich auftretender räumlicher Entitäten.

Für die lagemäßige Zuordnung der Maschen wird nicht jede einzelne Masche mit einer Koordinate versehen, es reicht, die Koordinaten des Ursprungs des Rasters zu kennen, sowie die Maschenweite, die Anzahl der Zeilen und Kolonnen, sowie die Orientierung des Rasters im Raum.

In einem Rastermodell werden die darzustellenden Objekte abstrahiert:

Die gespeicherten Attribute beziehen sich auf die Fläche der Gittermasche / Rasterzelle bzw. Pixel. (vgl. Bartelme 1995, S.46)

4.1 Geometrie Vektordaten /Rasterdaten

Die wesentlichen Unterschiede zwischen einem Vektor- und einem Rastermodell beschreibt folgende Gegenüberstellung (vgl. de Lange 2002):

Tabelle 1: Gegenüberstellung Vektormodell / Rastermodell

Vektormodell	Rastermodell
+ Hohe geometrische Genauigkeit	+ Einfache Datenstrukturen
+ Eindeutige Objektbeschreibung	+ Geringer Aufwand bei Erfassung von Geometrie und Topologie
+ Geringe Datenmenge	+ Einfaches Überlagern und Verschneiden von Geobjekten
	+ Einfache logische und algebraische Operationen
- komplexe Datenstruktur	- keine Form- und Lagetreue der Geobjekte
- aufwändige Erfassung von Geometrie und Topographie	- hoher Speicheraufwand
- aufwändige und rechenintensive logische und algebraische Operationen	- Güte der Form- und Lagebeschreibung der Geobjekte abhängig von der verwendeten Pixelgröße
- Geometrie und Topologie der Geobjekte wird parallel beschrieben	

Hierbei ist die geometrische Genauigkeit von Vektordaten kritisch zu betrachten.

Die exakten Koordinaten, mit denen in einem Vektormodell die dargestellten Entitäten beschrieben werden, suggerieren ein hohes Maß an Genauigkeit.

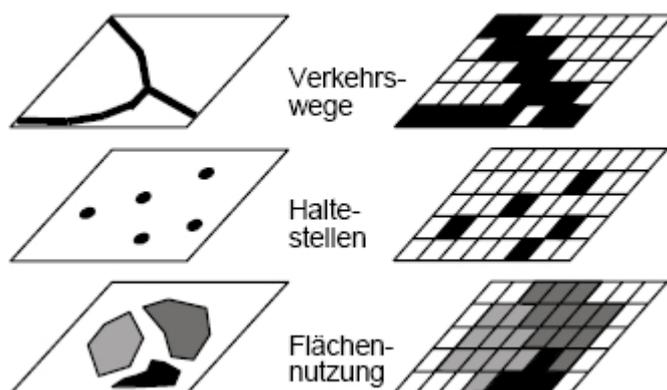
Diese ist aber in vielen Fällen nicht gegeben.

Der Grund hierfür sind Erfassungsungenauigkeiten bei der Datenerhebung oder die Darstellung von Entitäten, die in der realen Welt nicht scharf abzugrenzen sind, wie Flächen von Nutzungsarten. Diese Grenzen sind oftmals nicht exakt bestimmbar, da hier Übergangszonen von einer Nutzungsart zu anderen auftreten.

Auch können Objekte, die scharf abzugrenzen sind, wie z.B. Gebäude, im Modell durch unscharfe Geometrien präsentiert werden.

Folgende Abbildung veranschaulicht die Unterschiede in der Darstellung der Geometrie der verschiedenen Daten.

Abb. 1 : Darstellung von Geoobjekten im Vektor- und Rasterdatenmodell



(Quelle: de Lange 2002, S. 335)

Hierbei steigt der Speicherbedarf der Rasterdaten mit zunehmender Auflösung an. Verschiedene Algorithmen versuchen das Speicherplatzproblem zu lösen, in dem sie im Rasterdatensatz den Datenumfang an den Stellen reduzieren, an denen benachbarte Rasterzellen die gleichen Attributwerte haben und somit zu größeren Gebieten zusammengefasst werden können.

So wird beim run length encoding (Lauflängencodierung) zeilenweise der Wert einer Zelle erfasst und die Anzahl benachbarter Zellen mit gleichem Wert gespeichert.

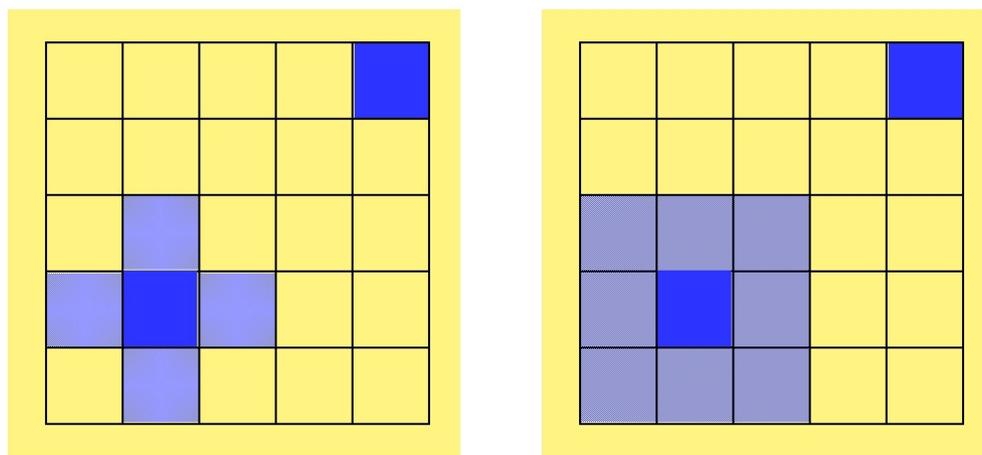
Beim chain encoding (Kettencodierung) wird durch Zeilen- und Spaltenindizes das Anfangspixel einer Linie erfasst. Über die Angabe der Richtungen R_1 bis R_n werden die, die Linie beschreibenden Folgepixel beschrieben.

Die Folgepixel werden dabei über die N.4-Nachbarschaft oder die N.8-Nachbarschaft erkannt. (vgl. de Lange 2002, S. 333)

Im Rastermodell ergibt sich die topologische Beziehung der Rasterzellen zueinander durch ihre Geometrie.

So sind Zellen benachbart, wenn sie eine gemeinsame Zellkante (Kanten-Kanten-Topologie, N.4 Nachbarschaft) oder eine gemeinsame Zellecke (Ecken-Kanten-Topologie, N.8-Nachbarn) besitzen.

Abb. 2: Rastermodell - Nachbarschaften



(Quelle: eigene Darstellung)

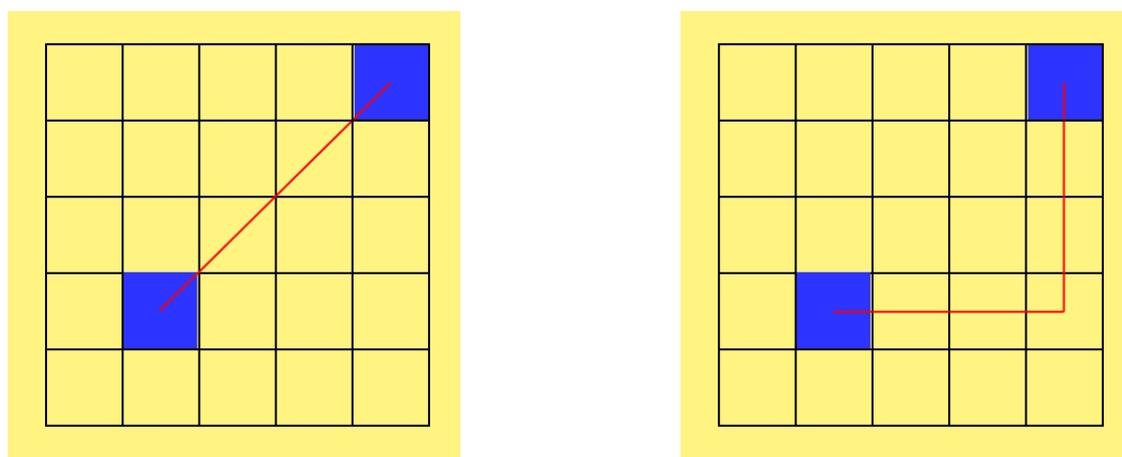
Da Vektor- und Rasterdaten unterschiedlich in ihrer Struktur sind, unterscheidet sich auch der Algorithmus der Abfragen, mit der z.B. bei einer geometrischen Abfrage eine Distanz ermittelt wird.

In einem Vektordatensatz kann der kürzeste Abstand zwischen zwei Punkten mit Hilfe der jeweiligen Punktkoordinaten und des Satz des Pythagoras errechnet werden.

In einem Rasterdatensatz wird die euklidische Distanz zwischen zwei Rasterzellen durch bestimmen des Abstands der beiden Zellmittelpunkte zueinander berechnet.

Oder es wird der Abstand zwischen zwei Punkten durch die sog. City-Block-Metrik (oder Manhattan-Metrik) berechnet.

Abb. 3: Rastermodell - Abstandmessung



(Quelle: eigene Darstellung)

Die jeweilige Datenstruktur kann in einem Geographischen Informationssystem durch Konvertierung verändert werden.

Objekte des Vektormodells können als Rasterdaten gespeichert werden, ebenso können Rasterdaten zu Vektorlinien konvertiert werden.

4.2 Zuordnung von Daten zu Rasterzellen

Bei der Verwendung von auf Rastern bezogene Daten ist grundsätzlich zwischen einem Vektoraster und einem Rasterdatensatz / Grid zu unterscheiden.

Vektoraster bestehen aus einer Vielzahl von gleich großen Polygonen (z.B. Quadraten). Dabei werden die einzelnen Rasterquadrate getrennt voneinander betrachtet.

Für jede Zelle gibt es in der Attributtabelle eine Angabe über Größe und Lage der Zelle im Raum.

Für Analysen können hier die gleichen Vektor-Analyse Funktionen verwendet werden, wie für alle Vektordaten.

In einem Rasterdatensatz /Grid ist die einzelne Zelle / das einzelne Pixel Teil einer Matrix, die über ihre Anzahl von Zeilen und Spalten beschrieben wird.

Als Georeferenz dienen die Koordinaten des Ursprungs der Matrix, sowie die Zellgröße. Die Lage des einzelnen Pixels wird über seine Lage innerhalb der Matrix bestimmt.

Zur Erzeugung eines Rasterdatensatzes ist zunächst der Einsatz eines Rastergitters (Vektoraster) erforderlich.

Dieses Rastergitter wird über den Untersuchungsraum gelegt und mit den zu untersuchenden (Vektor-) Daten verschnitten.

So erhält man gleich große Rasterzellen, denen Werte / Attribute zugewiesen werden.

Für die Verwendung dieser Daten als „echte“ Rasterdaten/ Griddaten, kann dieses Vektoraster in einen Rasterdatensatz konvertiert werden.

Für rasterbasierte Analysen können die in einem Vektormodell vorliegenden punkthaften, linienhaften oder flächenhaften Objekte den Rasterzellen zugeordnet werden (vgl. Wonka 2008, S.25ff.):

Flächenhafte Sachverhalte werden in einem Vektormodell durch die Lage und Form einzelner Polygone beschrieben.

Sollen diese flächenhaften Sachverhalte in einem Rastergitter abgebildet werden, so ist eine Zuordnung der Werte zu den jeweiligen Rasterzellen erforderlich.

Dies kann entweder über den Rastermittelpunkt erfolgen, d.h. die Rasterzelle erhält den Wert des Polygons zugewiesen, in dem sich der Mittelpunkt der Zelle befindet oder der Zelle wird der Wert des Polygons mit dem flächenmäßig größten Anteil innerhalb der Zelle zugewiesen.

Hierbei stellt die zweite Methode den größeren Aufwand dar.

Die Polygone müssen mit dem Rastergitter verschnitten und anschließend die jeweils größte Fläche pro Zelle ermittelt werden.

Punkthaft Sachverhalte werden einer Rasterzelle zugewiesen, in dem entweder die Anzahl der Punkte, die innerhalb einer Rasterzelle liegen, gezählt werden, oder es werden die, den Punkten zugeordneten, Werte addiert. Da sich diese Werte auf einheitliche Flächen (Rasterzellen) beziehen, können sie als Dichtewerte betrachtet werden.

Um linienhafte Sachverhalte auf Rasterzellen zu übertragen, muss das Linienthema mit dem Rastergitter verschnitten werden.

Anschließend wird die Länge der Linie ermittelt. Als Wert der Zelle wird dann die Länge der Linie mit dem Merkmal multipliziert (Beispiel: Linien = Wasserstraßen, Merkmal = Güterverkehr auf Wasserstraßen in Tonnen).

5 GIS gestützte räumliche Analyse

Die GIS gestützte räumliche Analyse ist ein Prozess, bei dem mittels analytischer Techniken raumbezogene Daten analysiert werden, um neue geographische Informationen zu gewinnen.

Die Analyse von Daten kann in zwei Methoden unterteilt werden:

-Abfrage (Query)

Hierbei werden die Eigenschaften der Daten ermittelt. Es werden keine neuen Daten erzeugt, vorhandene nicht verändert.

-Manipulation

Bei der Manipulation werden vorhandene Daten verändert oder aus ihnen neue Daten erzeugt.

Analyse von Vektordaten / Abfrage

In einem Geoinformationssystem werden mit Hilfe von räumlichen Abfragen (spatial queries) die lagemäßigen Beziehungen der Entitäten im Raum ermittelt und analysiert.

Dabei wird zwischen der qualitativen und der quantitativen räumlichen Analyse unterschieden.

Die qualitative Analyse ermittelt die Art und / oder der Aufbau einer Entität, während die quantitative Analyse die Menge oder Größe bestimmt.

Wobei zu jeder räumlichen Analyse die anschließende, fachgerechte Interpretation der Ergebnisse gehört (vgl. Bill et. al. 2001, S.221).

Die räumlichen Abfragen können unterschieden werden in geometrische, thematische und topologische Abfragen.

Mit Hilfe der geometrischen Abfragen wird die Form oder Ausdehnung der Objekte gemessen, z.B. die Fläche, der Umfang oder die Distanz zwischen zwei Objekten.

Thematische Abfragen selektieren die Objekte nach Attributen, wobei einfache Abfragen möglich sind, aber auch verschachtelte Abfragen, bei denen mehrere Abfragen durch logische Operatoren miteinander verbunden werden.

Topologische Abfragen zeigen die räumlichen und strukturellen Eigenschaften eines Objektes auf.

Mit ihnen werden Eigenschaften wie Nachbarschaft, Enthalten sein, Aneinandergrenzen bzw. Berühren, Schneiden bzw. Überlappen oder Distanz ermittelt.

So bietet ArcGIS (ESRI) z.B. folgende Operatoren zur Bestimmung der Nachbarschaftsbeziehungen der Objekte an (anzuwenden bei Vektordaten):

Tabelle 2: ArcGIS Operatoren

Topologische Beziehung	ArcGIS räumlicher Operator
overlap	intersect
disjoint	are within a distance of
contain	completely contain
covered by	are completely within
covers	have their centroid in
equal	are identical to
contain	contain
inside	are contained by

Analyse von Vektordaten / Manipulation

Analysefunktionen, die vorhandene Daten verändern, sind in ArcGIS (ESRI) z.B.:

Extract / Ausschnitt

Hierbei werden aus den Daten bestimmte Teilbereiche ausgeschnitten.

Overlay / Überlagerung

Mit Hilfe des Overlay können zwei unterschiedliche Thematiken überlagert und verschnitten werden.

Proximity / Umgebung

Proximity Funktionen ermöglichen Nachbarschaftsbeziehungen der Objekte zu untersuchen.

So können z.B. durch die Definition eines Buffer, bei der eine Pufferfläche in einem vorgegebenen Abstand um das zu untersuchende Objekt gelegt wird, Elemente ermittelt werden, die innerhalb dieser Distanz liegen.

Statistics / (Statistik)

Über die Statistik Funktionen können z.B. Minimum-, Maximum-, Mittelwerte oder die Standardabweichung bestimmt werden.

Analyse von Rasterdaten / Griddaten

In ESRI ArcGIS werden die Rasterdaten in zwei verschiedenen Formaten gespeichert:

Integer Grid oder Floating-Point Grid.

Das Grid wird mit Hilfe von drei Parametern bestimmt:

Den Ursprungskordinaten (engl. origin), der Rastergröße (engl. resolution, cell size) und der Anzahl der Zeilen (engl. rows) und Spalten (engl. columns) (engl. extent).

Im Integer Grid werden diskrete Daten gespeichert.

In der Attribut Tabelle wird dabei für jeden im Grid auftretenden Wert (Value) angegeben, wie häufig er auftritt.

Mit Hilfe eines Floating-Point Grid dagegen werden kontinuierliche Daten dargestellt wie z.B. Schadstoffkonzentrationen im Boden oder Geländehöhen.

Da jede Grid Zelle einen beliebigen Wert innerhalb eines vorgegebenen Wertebereiches haben kann, wird für ein Floating-Point Grid keine Attribut Tabelle erzeugt.

(Für eine Grid Zelle, für die kein Wert vorliegt, wird der sog. NoData-Wert verwendet.)

In einem Raster können Bereiche durch die Angabe eines Rechteckes oder die Verwendung einer Maske ausgeschnitten werden.

Für die Verschneidung von zwei oder mehr Rasterdatensätzen miteinander können folgende Funktionen verwendet werden (anlehnend an die von Tomlin 1991 beschriebene Map-Algebra):

Local functions

Mittels Local Functions wird ein Output-Raster erzeugt, bei sich die Werte der einzelnen Zellen durch Berechnung von lagemäßig identischen Input-Rasterzellen Werten ergeben.

Focal Functions

Bei dieser Funktion wird der Wert der Output-Rasterzellen aus den Werten des Input-Rasters unter Einbeziehung der Nachbarzellen errechnet.

Zonal Functions

Mit Hilfe dieser Funktion wird der Wert einer Zelle im Output-Raster unter Einbeziehung von Zellen innerhalb einer Zone, die im Input-Raster definiert sind.

Global Functions

Bei der Berechnung des Wertes einer Zelle im Output-Raster werden alle Zellen des Input-Rasters mit einbezogen.

Für die Verwendung dieser Funktionen ist es erforderlich, dass das Input-Raster und das Output-Raster die gleiche Rasterweite haben, so dass die Layer Rasterzellen „deckungsgleich“ übereinander liegen.

6 Bezugssystem

Entsprechend der INSPIRE Direktive zur Schaffung einer einheitlichen Geodaten-Basis, Annex I sind die europäischen Länder aufgefordert, für den Raumbezug zukünftig ein einheitliches Koordinatenreferenzsystem (Coordinate Reference System = CRS) zu verwenden.

Bislang nutzen die einzelnen europäischen Länder unterschiedliche Projektionssysteme, so dass das Zusammenfügen rasterbezogener Daten für länderübergreifende Projekte einen hohen Konvertierungsaufwand darstellt.

Durch Orientierung an der INSPIRE Direktive sind die europäischen Länder angehalten, zu Gunsten eines europaweit anzuwendenden Projektionssystems auf eigene, nur regional verwendbare Systeme zu verzichten.

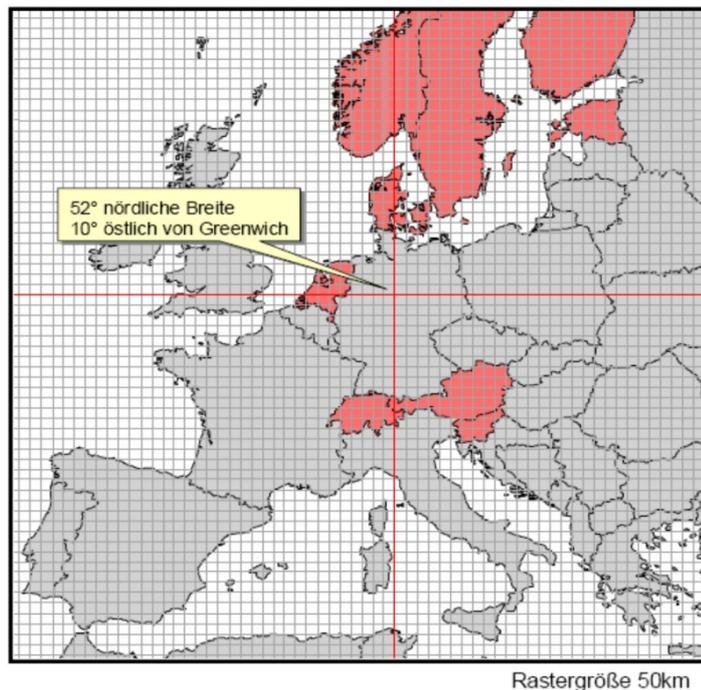
Da Abbildungen ellipsoidischer Koordinaten in die Ebene je nach Projektionsmethode unterschiedliche Verzerrungen aufweisen, sieht INSPIRE drei unterschiedliche Kartenprojektionen vor, so dass für die Abbildungen Längen-, Winkel- oder Flächentreue gegeben ist.

Alle drei Koordinatenreferenzsysteme basieren auf dem European Terrestrial System 1989 (ETRS 89).

Vorgesehen sind:

- Lambert Azimutal Equal Area (ETRS-LAEA), flächentreu, für statistische Zwecke.
- Lambert Conformal Conic (ETRS-LCC), eine nahezu längen- und winkeltreue Projektion für Maßstäbe 1 : 500 000 und kleiner.
- Transverse Mercator (ETRS-TMzn), eine Projektion im UTM-System mit Bezugssystem ETRS89

Abb. 4: Regionalstatistisches Rasternetz auf der Basis des ETRS1989 LAEA (European Terrestrial Reference System)



(Quelle: Wonka 2008)

In den rot markierten Ländern ist die dauerhafte Speicherung von Adresdaten (Gebäudekoordinaten) in den statistischen Datenbanken im Zusammenhang mit Personendaten erlaubt.

Zurzeit findet in der gesamten Bundesrepublik Deutschland der Umstieg vom bisher verwendeten DHDN Gauß-Krüger Koordinatensystem auf das Koordinatensystem UTM Zone 32 N auf Basis des Bezugssystems ETRS 89 statt. Die Umstellung des Koordinatensystems ist am 15.Mai 2007 durch eine EU-Richtlinie, anlehnend an die INSPIRE Direktive, beschlossen worden.

Für die Definition von Rasterzellen ist die Verwendung eines metrischen Koordinatensystems notwendig.

Durch die Verwendung eines fixen, koordinatengebundenen Rasternetzes können statistische Daten untereinander aber auch mit Daten anderer Thematiken problemlos kombiniert werden.

Die europaweite Verwendung eines einheitlichen Koordinatenreferenzsystems ermöglicht darüberhinaus länderübergreifende Analysen.

Ein individuell gelegtes Rasternetz, wie es für die verwendeten Beispieldaten definiert wurde, bietet den Vorteil, den zu untersuchenden Raum optimal abzudecken.

Für wiederkehrende Untersuchungen des gleichen Raumes (z.B. Zeitreihen) oder für den Vergleich benachbarter Untersuchungsräume, ist die Verwendung eines standardisierten Rasternetzes empfehlenswert.

7 Methoden zur Wahrung des Datenschutzes

In Deutschland wird zur Gewährleistung der statistischen Geheimhaltung von Tabelleneinträgen die Mindestfallzahlregel angewendet.

Dabei wird jeder Tabelleneintrag gesperrt, wenn er weniger als 3 Einheiten enthält (vgl. Giessing et.al. 2006, S.806).

Aber auch das „Vergrößern“ der Daten wird praktiziert (vgl. Szibalski 2007, S.143).

D. h. die Tabellenwerte werden zu Klassen zusammengefasst.

Dabei werden die Klassenobergrenzen und -untergrenzen so gewählt, dass sie zu gering besetzte Tabelleneinträge mit einbeziehen.

Diese Methoden lassen sich auch auf die Erstellung von Rastern aus statistischen Daten anwenden.

Der entscheidende Faktor für die Methode des Sperrens von gering besetzten Zellen ist die Mindestfallzahl.

Denn:

Das Entfernen von Zellen, die die Mindestfallzahl nicht aufweisen, kann einen nicht unerheblichen Informationsverlust bedeuten.

Gerade in ländlichen Gebieten kann die Einwohnerzahl pro Flächeneinheit gering ausfallen. Aber auch bei Daten, die nach soziodemographischen Gesichtspunkten ausgewählt werden, können niedrige Fallzahlen auftreten.

Bei statistischen Daten, bezogen auf administrative Grenzen, treten aufgrund der generell größeren Fläche der Bezugseinheit, Unterschreitungen der Mindestfallzahl kaum auf. Zumindest werden sie in der Literatur nicht diskutiert.

In den verschiedenen europäischen Ländern, die Rasterdaten aus den Daten der amtlichen Statistik erzeugen, werden unterschiedliche Mindestfallzahlen zur Wahrung der statistischen Geheimhaltung verwendet.

Eine Abhängigkeit der Fallzahl zur verwendeten Rasterweite ist dabei nicht eindeutig zu erkennen. (vgl. Szibalski 2007, S.142f).

Die Verwendung von Klassen, die dem Datensatz angepasst sind, ist für die kartographische Darstellung der statistischen Daten unproblematisch.

Sollen diese Daten für weitere (räumliche) Analysen verwendet werden, sind Zellen, die nicht die Mindestfallzahl aufweisen, zu löschen.

Alternativ kann die geringe Zahl durch die Mindestfallzahl ersetzt / verfälscht werden.

In beiden Fällen werden jedoch die absoluten Zahlen verändert.

Sollen keine Zellen unterdrückt werden, können folgende Methoden angewendet werden (nach Kaminger et.al. 2007, S. 305):

Hierarchische Aggregation

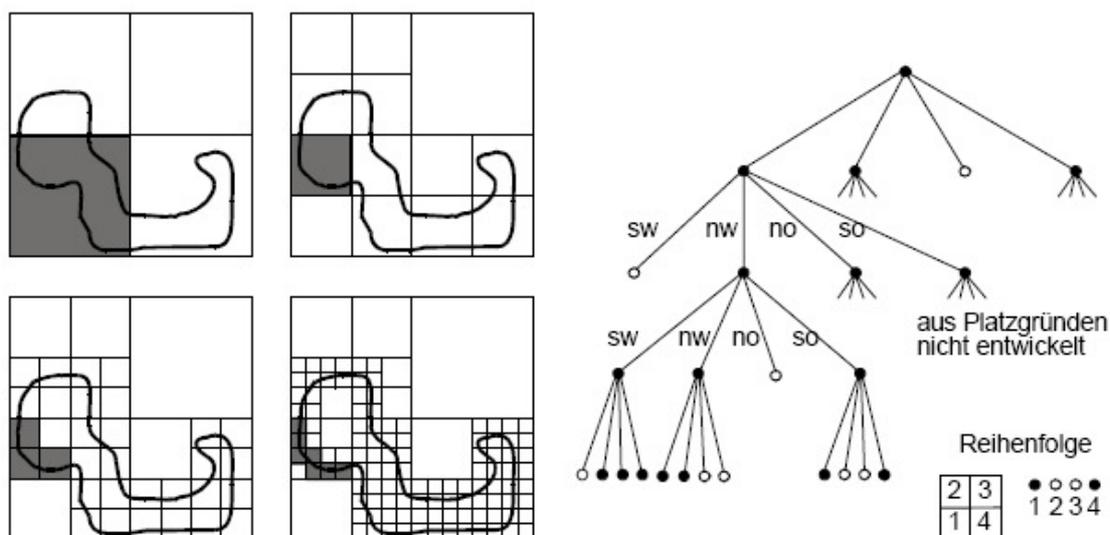
Bei der hierarchischen Aggregation werden Zellen, die zu geringe Fallzahlen aufweisen, mit ihren Nachbarzellen zu einem Block zusammengefasst.

Die hierarchische Aggregation erfolgt in Anlehnung an das Quadtree-Prinzip.

Ein Quadtree ist eine baumartige Datenstruktur, die die Speicherung eines Rasterdatensatzes beschreibt. Dabei wird die gesamte Rastermatrix rekursiv in gleich große Viertel unterteilt. Diese Teilung erfolgt so lange, bis alle Quadranten bezüglich eines ausgewählten Attributes homogen sind.

(vgl. Bill et.al. 2001, S. 216).

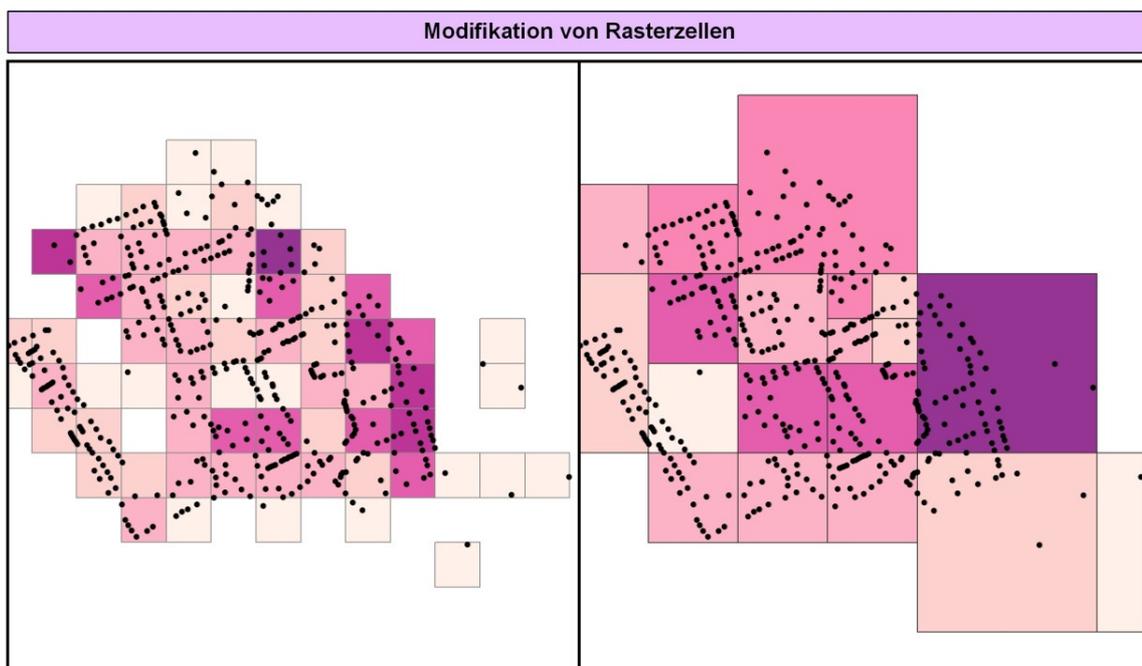
Abb.5: Darstellung von Rasterdaten als Quadtree



(Quelle: de Lange 2002, S.334)

Durch das Zusammenfassen von unterbesetzten Zellen entstehen übergeordnete, größere Zellen, die die kleineren, gering besetzten Zellen überlagern. Dies erfolgt solange, bis alle Zellen die Mindestfallzahl aufweisen.

Abb.6: Rasterzellen variabler Größe



(C) C. Herrmann-Hahn, 2009

Abbildung 6 zeigt beispielhaft die so entstehenden Zellen variabler Größe.

Hierbei sind jeweils 4 kleinere Zellen zu einer nächstgrößeren Zelle zusammengefasst, sobald die Summe der, innerhalb einer Zelle liegenden Werte, nicht die Mindestzahl erreicht.

Nachteilig bei dieser Methode ist, dass ausreichend besetzte kleinere Zellen von der nächstgrößeren Zelle überdeckt werden, sobald nur eine Zelle des „4-er Blocks“ nicht die Mindestzahl aufweist.

Damit gehen Detailinformationen verloren.

Zusammenfassen von benachbarten Zellen

Bei dieser Methode werden unterbesetzte Zellen solange mit Nachbarzellen zusammengefasst, bis in den so entstehenden Flächen der Schwellwert überschritten wird.

Individuelles Clustering

Hierbei werden Zellen nach ausgewählten Kriterien zusammengefasst, wobei die Zellen nicht zwingend benachbart sein müssen.

Bei dieser Methode könnte allerdings, bei zweimaliger Anwendung, durch Differenzbildung auf Einzeldaten geschlossen und so die Datenschutzrichtlinien verletzt werden.

Die drei vorgestellten Methoden sind aufgrund ihrer variablen Zellgröße nur bedingt für zeitliche Vergleiche geeignet, da die Zellgrößen der Daten verschiedener Zeiträume variieren.

8 Beispielgebiet

Die für diese Arbeit verwendeten Daten wurden freundlicherweise von der Stadt Dortmund zur Verfügung gestellt.

Als Kartenhintergrundbilder werden für die Abbildungen die Kartenblätter 2594714, 2594716, 2594714 und 2596716 der DTK10, Bezugssystem Gauß-Krüger 2. Meridianstreifen verwendet.

Quelle: Bezirksregierung Köln, Dezernat 74 / Geodatenzentrum, Geodateninfrastruktur.

© GEObasis.nrw

Um die Unterschiede zwischen einer vektorbasierten Analyse gegenüber einer rasterbasierten Analyse zu untersuchen, wurde als Beispielgebiet der Stadtteil Mengede gewählt.

Mengede ist einer der 12 Stadtbezirke der kreisfreien Stadt Dortmund, knapp 28,77 km² groß und hat mit einer Bevölkerungszahl von insgesamt 37.722 (Stand 31.12.2007) eine Einwohnerdichte von durchschnittlich 13,1 Personen pro ha.

(Die Gesamtbevölkerungszahl wurde aus den vorliegenden Daten ermittelt und weicht u.U. von der Bevölkerungszahl anderer Quellen ab.)

Es liegt im Nordwesten von Dortmund und grenzt an die Städte Castrop-Rauxel, Waltrop und Lünen. Zum Stadtbezirk Mengede gehören die Stadtteile Bodelschwingh, Mengede, Nette, Oestrich, Schwieringhausen und Westerfilde.

Im Norden wird Mengede begrenzt von der Autobahn A2, im Westen verläuft die A 45 durch den Stadtteil.

Auch führen durch Mengede mehrere stark befahrene Hauptverkehrsstraßen.

Durch Mengede führen ebenfalls einige Schienenwege der Stadtbahn und des überregionalen Eisenbahnverkehrs.

Die durch diesen Schienenverkehr verursachte Lärmbelastung ist in dieser Arbeit nicht mit berücksichtigt.

8.1 Daten der Lärmkartierung

Dortmund gehört zu den Ballungsräumen, für die gemäß der ersten Stufe der 34. Verordnung zum Bundes-Immissionsschutzgesetz (34.BImSchV) eine Lärmkartierung durchgeführt wurde.

So liegen für Dortmund für das Erhebungsjahr 2006 Daten der Kartierung des Umgebungslärms in 5dB-Schritten vor. Dabei wurden Bereiche erfasst, bei denen der L_{den} -Wert von 55dB(A) oder der L_{night} -Wert von 50db(A) überschritten wird.

Mit Hilfe von Isophonenbändern (Flächen, begrenzt von Linien gleichen Schallpegels) sind diese Bereiche ausgewiesen.

Zur Konkretisierung der der EU-Richtlinie 2002/39/EG über die Bewertung und Bekämpfung von Umgebungslärm trat im Jahr 2006 in Deutschland als 34. Verordnung zum Bundes-Immissionsschutzgesetz (34.BImSchV) die Verordnung über die Lärmkartierung in Kraft.

In ihr sind die Verfahren und Methoden festgelegt, die zur Berechnung der Lärmpegel für Straßen-, Schienen- und Fluglärm, sowie dem, durch Industrieanlagen und Häfen verursachten Umgebungslärm, verwendet werden. Es handelt sich um vorläufige Berechnungsverfahren, die endgültigen Berechnungsverfahren sind noch in Arbeit und werden erst für 2010 erwartet. (Die Ergebnisse sind mit den in Deutschland in anderen Zusammenhängen erhobenen Lärmberechnungsdaten nicht direkt vergleichbar, da andere Kenngrößen und Berechnungsmethoden angewendet wurden.)

2007 wurde für NRW eine Lärmkartierung veröffentlicht, die, wie in der Verordnung festgelegt, in der ersten Stufe die Erfassung von Ballungsräumen mit mehr als 250 000 Einwohnern und die Aufnahme der Hauptverkehrswege beinhaltet.

Die Lärmbelastungen wurden anhand von verschiedenen Lärmindizes berechnet und kartiert.

Vorrangig finden die beiden Indizes L_{den} (d(day)/e(evening)/n(night) Tag-Abend-Nacht-Pegel) sowie L_{night} (Nachtlärmpegel) Verwendung, die auch in den Karten der Lärmkartierung visualisiert werden.

Als Bezugsjahr liegt das Jahr 2006 zugrunde, die ausgewiesenen Lärmpegel beziehen sich auf eine Bezugshöhe von 4m über dem Gelände.

Der Schalldruckpegel wird in Dezibel (dB) gemessen. Da das natürliche Hören durch unterschiedliche Frequenzen beeinflusst wird, wird zur exakteren Bestimmung des Schalldruckes dem Messinstrument ein Filter (A) vorgeschaltet.

Daher wird der ermittelte Lärmpegel in der Einheit dB(A) angegeben.

Zur Berechnung des L_{den} wurde der ermittelte Lärmpegel des Bezugsjahres gemittelt, wobei der Lärm in den Abendstunden (zwischen 18.00Uhr und 22.00Uhr) mit 5dB Zuschlag und der Lärm in den Nachtstunden (zwischen 22.00Uhr und 6.00Uhr) mit 10dB Zuschlag gewichtet wurde.

Zur Bestimmung des Nachtpegels L_{night} wurde der Lärmpegel der Nachtstunden (Zeitraum zwischen 22.00Uhr und 6.00Uhr, mindestens jedoch 8 Stunden) gemittelt.

Bei der Kartierung des Umgebungslärms werden in 5dB-Schritten mit Hilfe von Isophonenbändern alle Bereiche ausgewiesen, bei denen der L_{den} -Wert von 55dB(A) oder der L_{night} -Wert von 50db(A) überschritten wird.

Umgebungslärm unterhalb der genannten Grenzwerte wird nicht erfasst.

Untersucht wurden dabei folgende Lärmquellen:

- Straßenverkehr auf Hauptverkehrsstraßen mit mehr als 6 Millionen PKW pro Jahr, in Ballungsräumen auch weitere Straßen, die Umgebungslärm verursachen.
- Schienenverkehr auf Haupteisenbahnstrecken mit mehr als 60 Tausend Zügen pro Jahr. Hier wurden die Untersuchungen nach Zuständigkeit getrennt.
Für die Untersuchung der Schienenwege des Bundes ist das Eisenbahnbundesamt (EBA) zuständig, die Schienenwege innerhalb der Ballungsräume, vorrangig Schienenwege des Öffentlichen Personen Nahverkehr (ÖPNV), wurden von der jeweiligen Ballungsraumgemeinde untersucht.
- Flughäfen mit mehr als 50 Tausend Starts oder Landungen pro Jahr.
In Ballungszentren wurden auch Flughäfen mit aufgenommen, die weniger Starts und Landungen aufweisen, aber relevanten Umgebungslärm verursachen.

- Industrieanlagen und Häfen mit einer Gesamtumschlagsleistung von mehr als 1,5 Tonnen pro Jahr

Die Anzahl der Personen, die in den jeweiligen lärmbelasteten Bereichen wohnen, wird nach bestimmten Berechnungsverfahren ermittelt. Ebenfalls ermittelt wird die Anzahl der Wohnungen, Schulen und Krankenhäusern, die in Bereichen mit 55dB, 65dB oder 75dB Lärmbelastung liegen. Die lärmbelasteten Flächen werden berechnet.

(Die vorangestellte Beschreibung ist den „Erläuterungen zur Kartierung von Geräuschen im Rahmen der EU-RL Umgebungslärm in NRW“, Stand 24.06.2008, entnommen.)

Bei Belastungen von mehr als 60 dB(A) L_{night} und/oder 70 dB(A) L_{den} besteht entsprechend der 34. Verordnung zum Bundes-Immissionsschutzgesetz der Handlungsbedarf, Lärmschutzmaßnahmen durchzuführen.

Wobei eine Reihe unterschiedlicher Lärmschutz oder Lärmvermeidungsmaßnahmen denkbar sind.

Beispielsweise kann eine Umleitung des Verkehrs, das Errichten von Lärmschutzwänden oder -wällen, aber auch die Verwendung von sogenanntem „Flüsterasphalt“ (offenporiger Asphalt, der Fahrgeräusche mindert) den Lärm verringern.

Ziel der Lärmkartierung ist, die Belastung durch Umgebungslärm aufzuzeigen, um so der Öffentlichkeit Umweltinformationen zur Verfügung stellen zu können.

Sie soll darüberhinaus die Grundlage bilden für Aktionspläne zur Erarbeitung und Umsetzung von Maßnahmen zur Vermeidung oder Verringerung der gesundheitsschädlichen Lärmbelastung.

Um die Wirkung, der in den Aktionsplänen vorgeschlagenen Maßnahmen zur Lärminderung prüfen zu können, ist vorgesehen, die Lärmkartierung spätestens alle 5 Jahre zu wiederholen.

Wobei der Erfolg der durchgeführten Maßnahmen anhand von Indikatoren beurteilt / bewertet werden kann.

Ein Indikator kann z.B. die Anzahl der von Lärmbelastung betroffenen Bevölkerung sein.

Ob es sich beim Umgebungslärm lediglich um eine Lärmbelästigung handelt, oder ob die Einwirkungen des Lärms Gesundheitsschäden hervorrufen, hängt nicht vom subjektiven Empfinden ab, sondern kann anhand des Lärmpegels benannt werden.

So gilt ein Lärmpegel von mehr als 55 dB(A) als Lärmbelästigung, während Personen, die längerfristig einem Lärmpegel von 65 bis 75 dB (A) ausgesetzt sind, vermehrt unter Stress leiden und ein höheres Risiko haben, an Herz- oder Kreislauferkrankungen zu erkranken.

Personen, die über längere Zeit einem Lärm von mehr als 85 dB(A) ausgesetzt sind, können schwerhörig werden.

In den erstellten Lärmkarten, die im Internet unter der Adresse <http://www.umgebungs-laerm-kartierung.nrw.de/> bzw. <http://laermkartierung.eisenbahnbundesamt.de/> präsentiert werden, werden die ermittelten L_{den} - und L_{night} -Werte in 5dB-Schritten mit Hilfe von Isophonenbändern dargestellt. Die Anzahl der vom Lärm betroffenen Personen wird in den Internetanwendungen in Tabellenform zur Verfügung gestellt.

Aber:

"Für die Prioritätensetzung von Maßnahmen ist es sinnvoll, die Betroffenenzahlen nicht nur -wie von der Richtlinie gefordert- tabellarisch, sondern auch kartenmäßig darzustellen."(Schwedler 2008, S.11)

In der vorliegenden Arbeit wurden die Daten der Lärmkartierung bezogen auf den Straßenverkehrslärm (Tagespegel) verwendet.

8.2 Daten der Einwohnerstatistik

In Dortmund wird das gesamte Stadtgebiet aus administrativen und stadtplanerischen Gründen in 12 Stadtbezirke, 62 statistische Bezirke sowie 170 statistische Unterbezirke unterteilt.

Die Größe der Unterbezirke variiert dabei zwischen 13,12 ha und 616,79 ha.

Als Einwohnerdaten wurden die Zahlen der Hauptwohnbevölkerung bezogen auf den Stichtag 31.12. 2007 verwendet.

Unter Hauptwohnbevölkerung versteht man alle Personen, deren alleiniger Wohnsitz oder, falls jemand mehrere Wohnsitze hat, der Hauptwohnsitz Dortmund ist.

Diese Daten wurden bis 31.12.1998 monatlich ermittelt, indem, ausgehend von der zuletzt durchgeführten Volkszählung, Bewegungsdaten (Geburten, Sterbefälle, Zuzüge, Fortzüge, Umzüge) hinzugezählt wurden.

Seit 01.01.1999 werden als Bewegungsdaten die Daten des Einwohnerregisters der Stadt Dortmund hinzugezählt.

Von der Stadt Dortmund wurden für das Beispielgebiet folgende Daten zur Verfügung gestellt:

- Anzahl der Hauptwohnbevölkerung pro Unterbezirk
- Anzahl der Hauptwohnbevölkerung pro Gebäudeblock
- Anzahl der Hauptwohnbevölkerung bezogen auf ein 100 m x 100 m Raster

Zur Erzeugung der Rasterdaten wurde ein Gitternetz mit einer Maschenweite von 100m x 100m über das Beispielgebiet gelegt.

Der Ursprung des Rasters wurde so gewählt, dass das gesamte zu untersuchende Gebiet optimal abgedeckt wurde.

Zur Geokodierung der adressbezogenen Angaben der Einwohnerstatistik wurde diesen Daten mit Hilfe einer Tabelle, in der jeder Hausadresse/Hauskoordinate eine geographische Koordinate zugewiesen ist, die entsprechende Koordinate zugeordnet.

Anschließend wurde die Anzahl der Einwohner mit Hauptwohnsitz pro Rasterzelle ermittelt.

Diese Zahl wurde der jeweiligen Rasterzelle als Wert zugewiesen, wobei kein Wert unterdrückt wurde.

So sind 8,17% der Rasterzellen mit Werten von 1 bis 10 Einwohnern pro Zelle belegt.

Bei dem so erzeugten Raster handelt es sich um ein Vektorraster.

Die einzelnen Zellen sind gleich große, regelmäßig angeordnete (quadratische) Polygone, die in ArcGIS als Shape-File gespeichert werden.

Für die durchgeführten Analysen wurden Daten aus zwei verschiedenen Erhebungszeiträumen verwendet:

Die Einwohnerzahlen beziehen sich auf das Jahr 2007 (Stichtag 31.12.2007), die Daten der Lärmkartierung auf das Jahr 2006.

Lärmkartierungsdaten wurden entsprechend der 34. Verordnung zum Bundes-Immissionsschutzgesetz (34.BImSchV) erstmalig im Jahr 2006 erhoben und im Jahr 2007 veröffentlicht. Daher liegen für das Erhebungsjahr 2007 keine Daten vor.

Die Zahl der Einwohner mit Hauptwohnsitz in Mengede ist von 2006 (Stichtag 31.12.2006) bis 2007 (Stichtag 31.12.2007) um 107 Personen gesunken.

Das entspricht einer Abnahme von 0,279%.

(Zum Vergleich: Im gesamten Dortmunder Stadtgebiet ist die Anzahl der Hauptwohnbevölkerung in diesem Zeitraum um 0,188% gesunken.)

Da in dieser Arbeit die Verwendung der zur Verfügung gestellten statistischen Daten für eine GIS-gestützte räumliche Analyse untersucht werden, die Arbeit somit einen methodischen Schwerpunkt hat, wurden die Einwohnerdaten des Jahres 2007 aufgrund der geringen Abweichung der Gesamtpersonenzahl der beiden Erhebungsjahre für die Betroffenenanalyse verwendet.

8.3 Verwendetes Bezugssystem

Für den Beispieldatensatz wurde als Koordinatensystem DHDN_3_Degree_Gauss_Zone_2 verwendet.

Diesem System liegt eine Gauß Krüger Projektion zugrunde mit folgenden Einstellungen:

False_Easting: 2500000,000000

False_Northing: 0,000000

Central_Meridian: 6,000000

Scale_Factor: 1,000000

Latitude_Of_Origin: 0,000000

Linear Unit: Meter

GCS_Deutsches_Hauptdreiecksnetz

Datum: D_Deutsches_Hauptdreiecksnetz

Bei einer Transformation des vorhandenen Rastergitters in ein anderes Koordinatensystem würde sich die Form der quadratischen Zellen verändern.

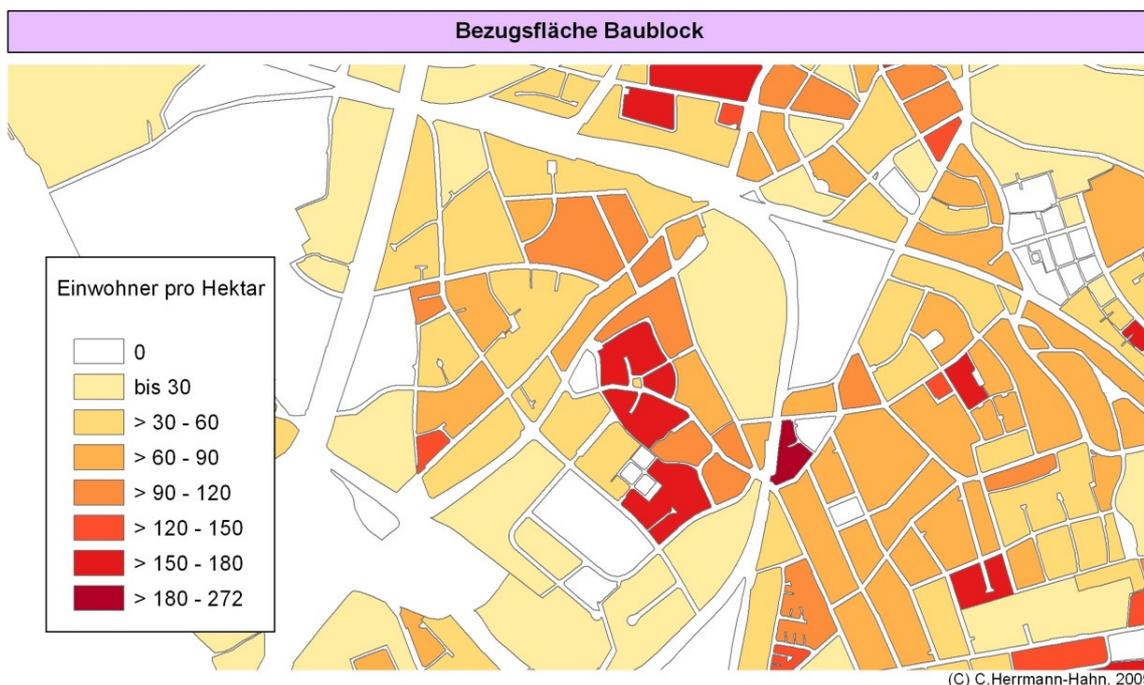
In diesem Fall ist ein Neuerzeugen der Daten aus den Ursprungsdaten angebracht.

8.4 Ausgangsdaten / Vektordaten

Bei den zur Verfügung stehenden Vektordaten handelt es sich um die Unterbezirksgrenzen und Baublockgrenzen des Dortmunder Stadtteils Mengede.

Verwendet werden die Polygone der Baublöcke, an die, über eine Verknüpfung mit der Baublocknummer, die Anzahl der Haupteinwohner pro Baublock als Attribut angehängt werden.

Abb. 7: Bevölkerungsverteilung - Bezugsfläche Baublock



Die Abbildung 7 zeigt einen Ausschnitt der, auf Baublöcke bezogenen, Einwohnerzahlen.

Berechnet wurden hier aus der absoluten Zahl der Einwohner pro Baublock sowie der Fläche des Baublocks die „Einwohner pro Hektar“.

Die Flächengröße und Form der Baublöcke variiert dabei sehr stark.

Während der kleinste Baublock in Mengede eine Fläche von 0,0156 ha beträgt, hat der größte eine Ausdehnung von 322,159 ha.

Die Bevölkerungsdichte schwankt dabei zwischen 0 Einwohnern bzw. als kleinster Wert 0,077 Einwohnern pro ha und 72,89 Einwohnern pro ha.

Obschon man augenscheinlich einen Eindruck über die vorherrschende Bevölkerungsverteilung erhält, sollte man nicht vergessen, dass hier die Einwohnerzahl bezogen auf die gesamte Fläche des Baublocks berechnet wird.

Unberücksichtigt bleiben dabei bebaute und nicht bebaute Gebiete.

Die Daten der Lärmkartierung liegen als Polygone der Isophonenbänder vor.

Diesen Polygonen ist als Attribut der jeweilige Lärmpegel in dB(A) zugewiesen.

8.5 Ausgangsdaten (Vektor-)Rasterdaten

Die Erstellung von Rasterdaten aus den Daten der Statistik ist nur möglich, wenn diesen Daten eine geographische Koordinate zugeordnet ist oder ihnen ein Koordinatenwert zugewiesen werden kann (vgl. Wonka 2008, S.14).

Im verwendeten Datensatz ist der Raumbezug über die Gebäudekoordinate hergestellt. Über eine Aggregation der Adresspunkte wird den jeweiligen Rasterzellen ein Wert zugewiesen.

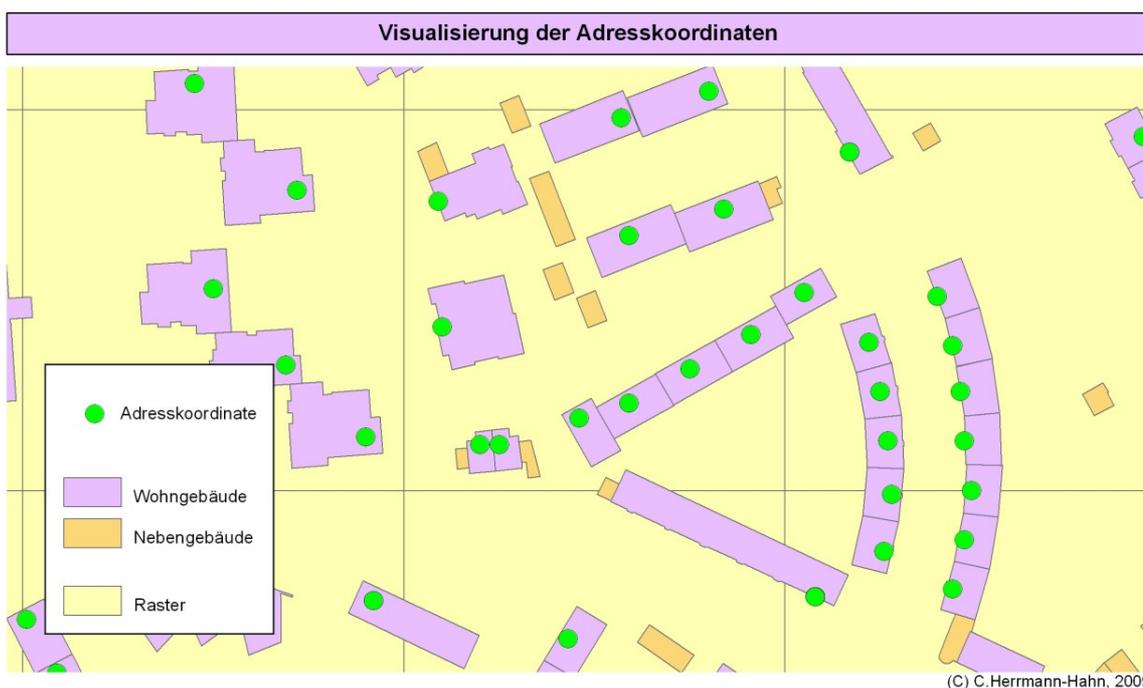
Dieser Wert ergibt sich entweder als Anzahl der Punkte, die innerhalb einer Rasterzelle liegen („count“), oder es wird, wie im verwendeten Datensatz, ein, der Adresskoordinate zugeordneter Wert (hier: Einwohnerzahl) addiert („sum“).

So entsteht ein Raster, in dem jeder Zelle die Anzahl der Einwohner pro Rasterfläche zugewiesen ist. Da sich dieser Wert auf eine einheitliche Flächengröße bezieht, kann er als Dichtewert verwendet werden.

Bei dem verwendeten 100m x 100m Raster geben diese Werte Auskunft über die Bevölkerungsdichte: Einwohner pro ha.

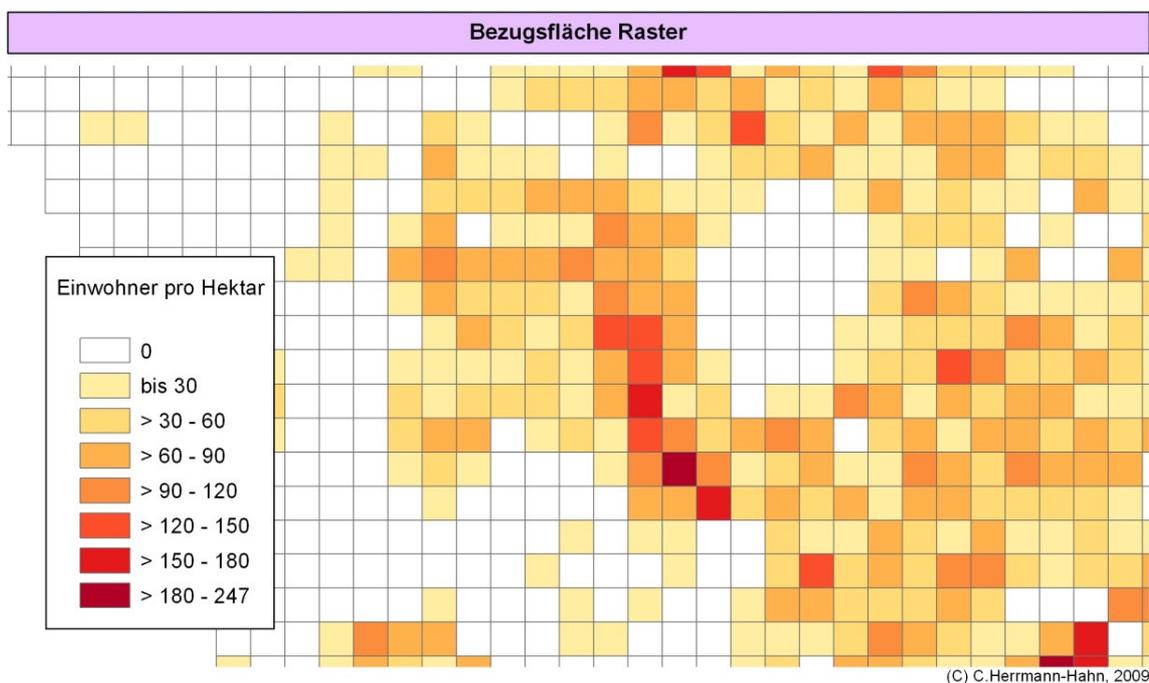
Wie die Abbildung 8 zeigt, liegen die Adresskoordinaten des verwendeten Datensatzes nicht in der Mitte der Gebäudepolygone.

Abb. 8 : Adresskoordinaten



Hier werden als Gebäudekoordinate die Koordinaten der in der ALK (Automatisiertes Liegenschaftskataster) platzierten Hausnummern verwendet (Hausnummernkoordinate).

Abb. 9: Bevölkerungsverteilung – Bezugsfläche Raster



Die Rasterkarte Abb.9 zeigt die Bevölkerungsdichte bezogen auf 1 ha große Rasterzellen.

Im Gegensatz zur Darstellung der Bevölkerungsdichte bezogen auf administrative Grenzen, sind hier nur Gebiete /Zellen, in denen Wohngebäude vorhanden sind, mit einem Wert belegt.

So kann die Wohnbevölkerung entsprechend ihrer „wahren“ Verteilung visualisiert werden.

Egal wie klein die Rasterzelle gewählt wird: Die reale Verteilung der Bevölkerung kann nur durch Vergleich mit den Ursprungsdaten kontrolliert werden.

Auch bei einer Verfeinerung des Rasternetzes sind die Werte der Rasterzellen mit Hilfe der Ursprungsdaten neu zu berechnen.

Liegen, wie im hier verwendeten Beispieldatensatz, keine auf Adresskoordinaten bezogene statistische Informationen vor, kann über eine Disaggregation eine Verfeinerung der Daten durchgeführt werden.

Während eine hierarchische Aggregation zur nächstgrößeren Rasterzelle ohne Probleme möglich ist, ist eine Disaggregation zur Verfeinerung der Daten nicht ohne Zuordnungsfehler durchzuführen.

Je kleiner die Rasterzelle, desto größer ist der Zuordnungsfehler (vgl. Steinnocher et.al. 2005).

Anmerkung:

Bei den verwendeten Daten liegt eine Differenz zwischen der Gesamtbevölkerungszahl, ermittelt aus der Summe der Einwohner pro Baublock, gegenüber der Bevölkerungszahl, ermittelt als Summe der Einwohner pro Rasterzelle, vor.

So werden 37.722 Einwohner über die Baublöcke ermittelt, während die Summe der Einwohner pro Rasterzelle nur 33.746 Einwohner ergibt.

Eine gewisse Abweichung der Absolut-Zahlen der beiden Datensätze ist durch die unterschiedliche Abdeckung des Untersuchungsgebietes durch die verschiedenen Bezugsflächen begründet.

Aus datenschutzrechtlichen Gründen stehen für die vorliegende Arbeit die Ursprungsdaten, aus denen die Einwohnerdichte der Rasterzellen berechnet wurde, nicht zur Verfügung.

Um zu Prüfen, ob eine weitere Ursache eine fehlerhafte Zuordnung der Gebäudemittelpunkte zu den Rasterzellen sein könnte, wird ein fiktiver Datensatz erstellt.

In diesem Datensatz werden Gebäudemittelpunkten, die über die Geometrie der Gebäudeumringe erzeugt wurden, willkürliche Einwohnerzahlen zugewiesen.

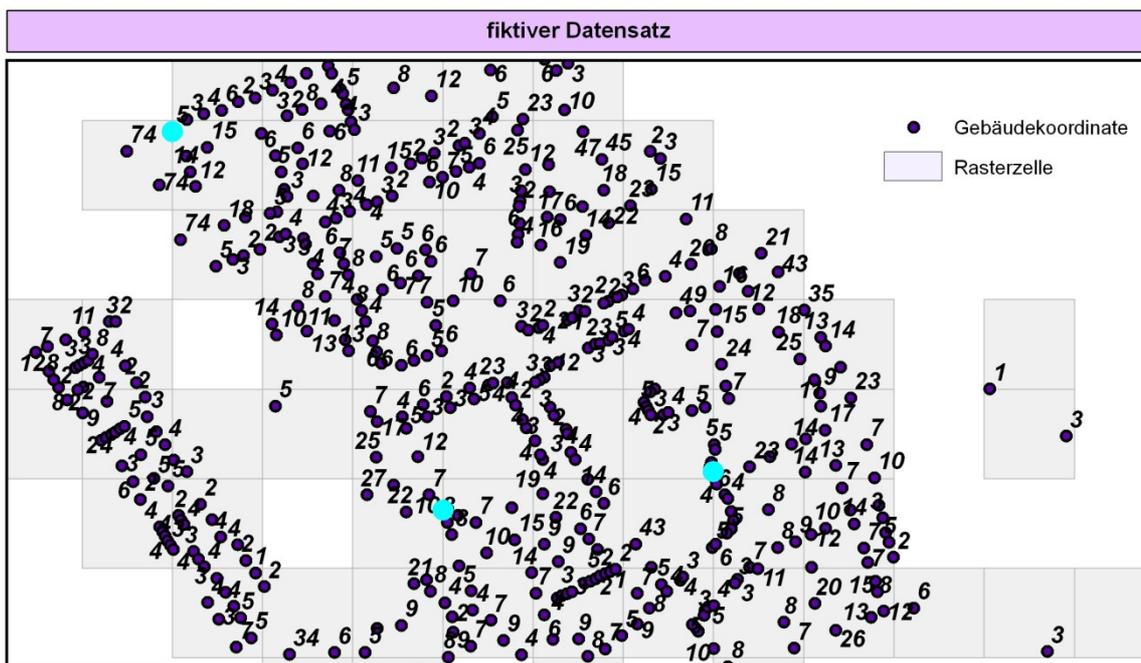
Die Summe der fiktiven Einwohner beträgt 3.734.

Anschließend wird die Summe der Einwohner pro Rasterzelle berechnet.

Die Summe der Einwohner bezogen auf die Rasterzellen wird mit 3.716 errechnet.

Die Überprüfung des Datensatzes zeigt: Die in der Abbildung blau markierten Gebäudemittelpunkte, die sehr nah an den Polygongrenzen liegen, werden keiner Zelle zugewiesen.

Abb. 10: fiktiver Datensatz: Zuordnung Gebäudemittelpunkt - Zelle



(C) C. Herrmann-Hahn, 2009

Parallel dazu werden den Gebäudemittelpunkten die Attribute der Zelle zugeordnet, in der sie liegen und anschließend die Einwohnerzahlen der Gebäude mit gleicher Zellen-ID addiert.

Die Summe der Einwohnerzahl wird hierbei richtig ermittelt.

Die deutlich geringere Einwohnerzahl der vorliegenden Rasterdaten könnte somit beim Aufsummieren der Werte innerhalb einer Zelle entstanden sein.

9 Darstellung der Bevölkerungsdichte

Die Daten der Einwohnerstatistik werden meist auf administrative Flächeneinheiten bezogen und mit Hilfe von Choroplethenkarten dargestellt.

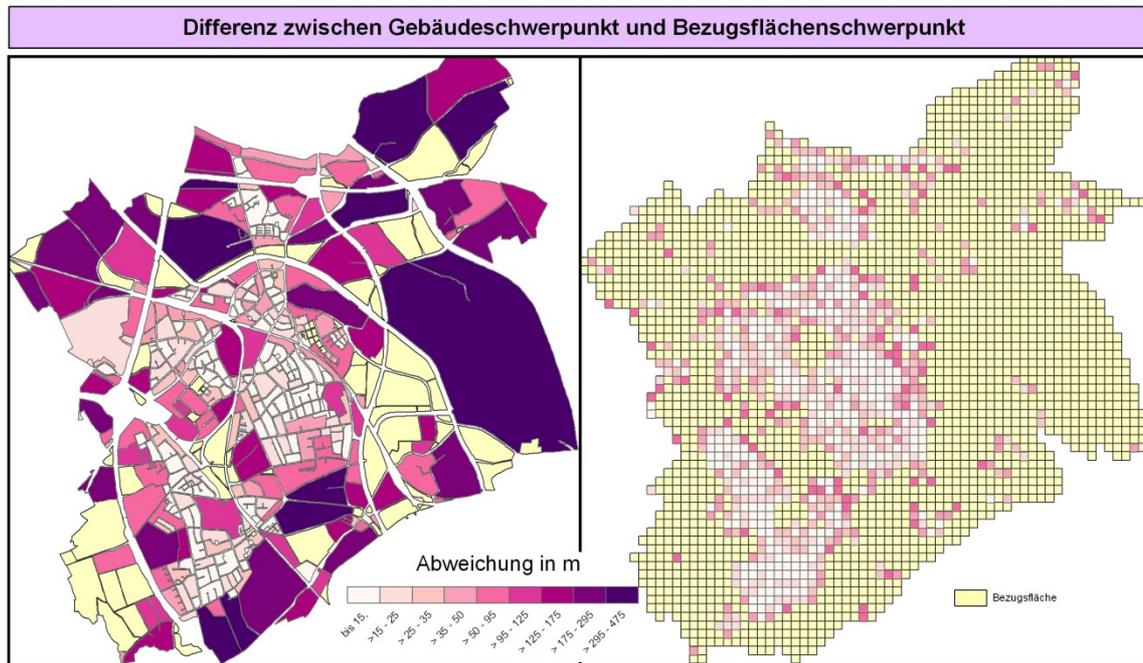
Die Interpretation der Daten auf Grundlage dieser Darstellungen wird dabei nicht nur durch die unterschiedlichen Größen und Formen der Bezugsflächen erschwert.

Auch die Verwendung von Dichtewerten, bezogen auf die gesamte Bezugsfläche, kann zu Fehlinterpretationen führen, da die Verteilung des untersuchten räumlichen Phänomens nicht ersichtlich ist.

Die Daten der Wohnbevölkerung sind unzweifelhaft mit der Lage der Wohnbebauung verbunden. Werden bei der Berechnung der Bevölkerungsdichte allerdings große Gebiete unbebauten Raums mit einbezogen, wird der Dichtewert für die bebauten Bereiche zu niedrig und für die nicht bebauten Bereiche zu hoch berechnet.

Der Fehler der korrekten Zuordnung steigt dabei mit zunehmender Flächengröße an.

Abb.:11: Differenz zwischen Gebäudeschwerpunkt und Bezugsflächenschwerpunkt



(C) C. Herrmann-Hahn, 2009

Abbildung 11 zeigt die Abweichung zwischen dem berechneten Gebäudeschwerpunkt und dem Schwerpunkt der jeweiligen Bezugsfläche.

Für beide Darstellungen werden eigene Gebäudeschwerpunkte ermittelt:

Für die Darstellung bezogen auf Baublöcke wird der Gebäudeschwerpunkt aller Wohngebäude innerhalb des jeweiligen Baublocks bestimmt, für die Darstellung,

bezogen auf Rasterzellen, der Schwerpunkt aller Wohngebäude innerhalb der einzelnen Zelle.

Anschließend wird der Abstand dieses Gebäudeschwerpunktes zum jeweiligen Mittelpunkt des Blockes bzw. der Rasterzelle errechnet (= Standardabweichung).

Damit die errechneten Abstände den Flächeneinheiten eindeutig zuzuordnen sind, erhalten diese die Farbe, die der Abweichung in der Legende zugewiesen ist.

Während diese Abweichung bei den Rasterzellen durchschnittlich 24,12 m (maximal 68,04 m) beträgt, beträgt die Abweichung bei den Baublöcken im Durchschnitt 55,65 m (maximal 472,74 m). Hierbei handelt es sich um gemittelte Werte, die einzelnen Abweichungen können deutlich höher sein.

Wobei die Abweichung maximal die größte Ausdehnung der Bezugsfläche betragen kann.

9.1 Bevölkerungsdichte- Bezugsfläche Baublock

Zur Visualisierung der räumlichen Verteilung der Bevölkerung ist die Darstellung der Bevölkerungsdichte unter Verwendung der Wohngebäude geeignet.

Diese Methode kann angewendet werden für größere Kartenmaßstäbe (ca. bis Maßstab 1 : 10.000), in denen eine Einzelhausdarstellung gegeben ist.

Für kleinere Maßstäbe (Maßstab kleiner als 1 : 10.000) ist zur Visualisierung der bebauten Flächen die Verwendung einer Siedlungsmaske angebracht.

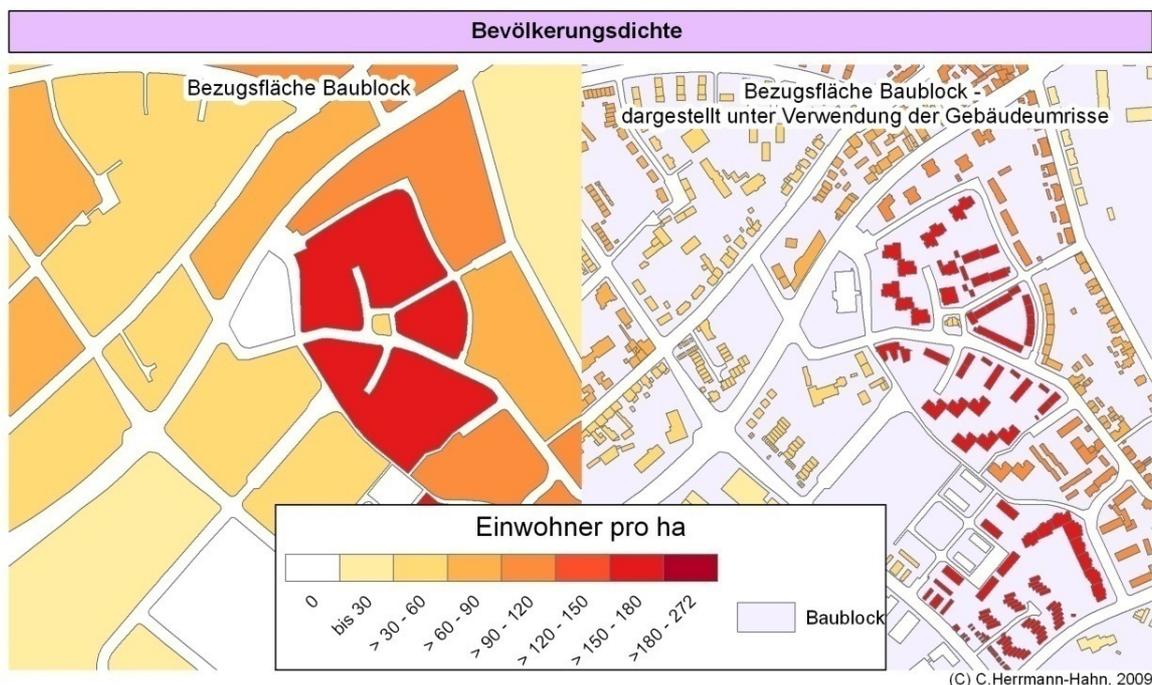
So ist für die Darstellung der Bevölkerungsdichte in der zweiten Karte der Abbildung 12 die Bevölkerungsdichte des Baublocks den Gebäudeumriss Polygonen als Attribut zugewiesen.

So können Bebauungsfreiräume gut erkannt werden.

Industriegebäude, Öffentliche Gebäude und Wohnnebengebäude werden bei dieser Berechnung nicht berücksichtigt.

Die erste Karte der Abbildung 12 zeigt die Bevölkerungsdichte = Einwohner pro Baublock.

Abb. 12: Bevölkerungsdichte – Bezugsfläche Baublock

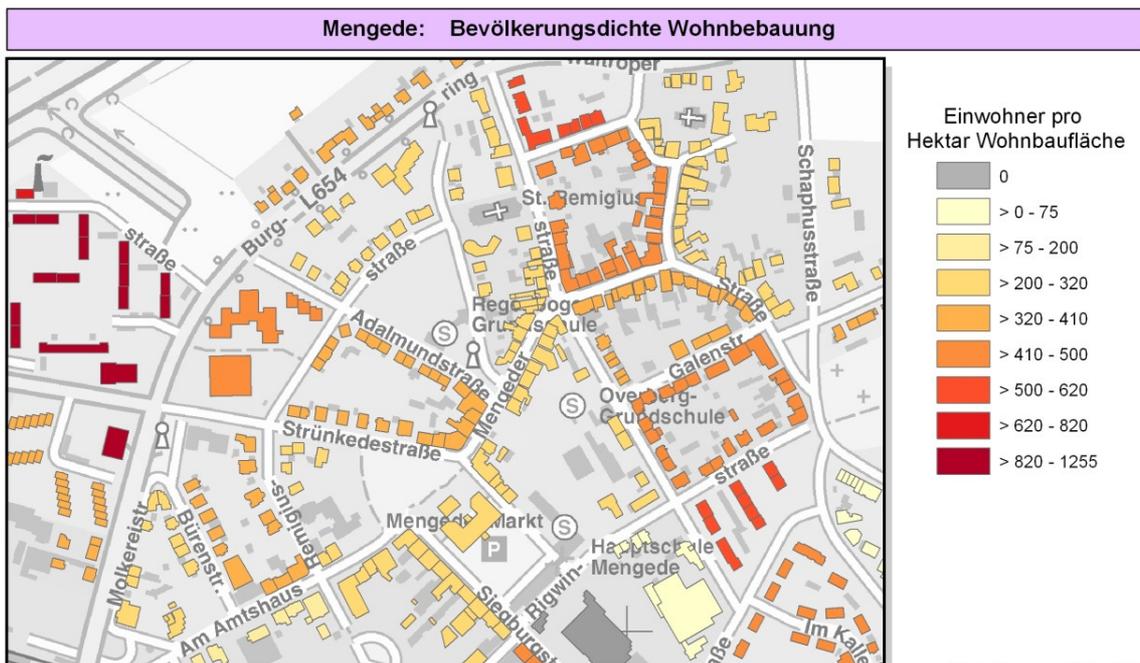


Da bei der Berechnung der hier dargestellten Einwohnerdichte auch die nicht bewohnten Bereiche mit einbezogen werden, ergibt sich für nicht besiedelte Gebiete mit großer Wahrscheinlichkeit ein zu hoher Einwohnerdichtewert, während für die bebauten Bereiche, der Dichtewert zu niedrig angegeben wird.

Aus diesem Grund ist für die folgende Darstellung der Bevölkerungsdichte die Fläche der Wohnbebauung innerhalb eines Blockes ermittelt und mit Hilfe dieses Wertes die Bevölkerungsdichte pro Wohnbaufläche errechnet.

Die Abbildung 13 zeigt eine Darstellung der Bevölkerungsdichte je Baublock, bei der die Einwohnerzahl pro bebaute Wohnfläche berechnet ist.

Abb. 13: Mengede: Bevölkerungsdichte Wohnbebauung



(Bei Verwendung der Bevölkerungsdichte der Wohnbebauung werden rund 42,98% der Gesamtbevölkerung als von Lärmbelastung betroffen errechnet.)

Unberücksichtigt bleibt bei dieser Beschreibung der Bevölkerungsdichte allerdings die Geschosshöhe der Gebäude. Somit wird auch hier, rein rechnerisch, von einer homogenen Verteilung der Bevölkerung auf die vorhandene Wohnbaufläche ausgegangen.

Aber auch Wohngebäude, die nicht im verwendeten Datensatz erfasst sind oder fälschlicherweise als Wohngebäude ausgewiesen wurden, verfälschen die Berechnung: Es wird u.U. eine zu niedrige (oder zu hohe) Gesamtwohnbaufläche errechnet.

9.2 Bevölkerungsdichte- Bezugsfläche Rasterzelle

Bei der Verwendung von Rasterzellen als Bezugsgröße für die Darstellung der Bevölkerungsdichte kann die räumliche Tiefe dem darzustellenden Thema angepasst werden. Je kleiner die Rasterzelle gewählt wird, desto präziser kann die räumliche Verteilung visualisiert werden.

Größere Rasterzellen bedeuten zwar immer eine Vergrößerung der Darstellung, auf der anderen Seite steigt die Wahrscheinlichkeit, dass die Zellen keine Werte unterhalb der Mindestfallzahl aufweisen.

Räumliche Unterschiede, wie die Abgrenzung von Wohnbebauung gegenüber unbebauter Fläche, bleiben erhalten.

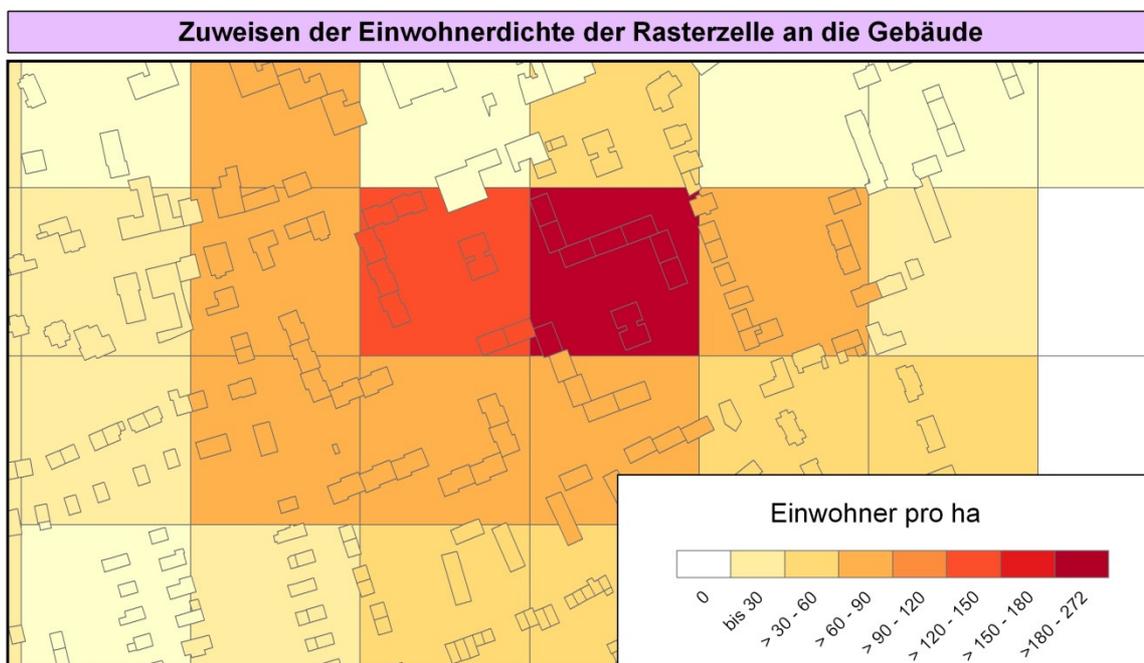
Zur detaillierten Visualisierung der Bevölkerungsverteilung kann man die Informationen der Rasterzellen mit den Umrissen der Wohngebäude verknüpfen.

Um die Informationen des Rastergitters den Gebäuden zuweisen zu können, werden zunächst für die Wohngebäude die Gebäudemittelpunkte erzeugt.

Diesen Gebäudepunkten wird die Einwohnerdichte der Rasterzelle zugewiesen, in der sie liegen.

Zuordnungsfehler treten bei dieser Methode u.U. im Randbereich der Rasterzellen auf: Den Gebäuden wird der Dichtewert über den Gebäudemittelpunkt zugewiesen, ohne zu berücksichtigen, in welcher Rasterzelle der flächenmäßig größte Teil des Gebäudes liegt.

Abb.14: Zuweisen der Einwohnerdichte der Rasterzellen an die Gebäude

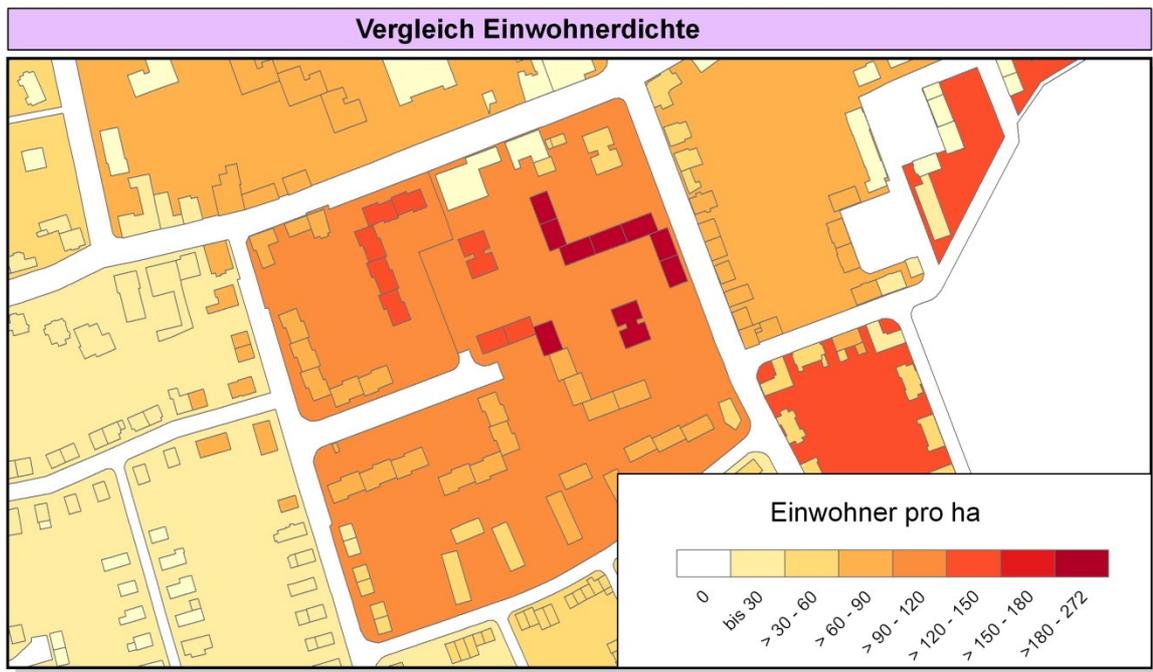


(C) C.Herrmann-Hahn, 2009

Abbildung 15 zeigt die Einwohnerdichte der Baublöcke, überlagert von den Wohngebäudepolygonen, die mit dem Dichtewert der Rasterzellen attributisiert sind.

Wie die Abbildung zeigt, können durch den Gebäudebezug die Bereiche mit hoher Einwohnerdichte gut veranschaulicht werden.

Abb. 15: Vergleich Einwohnerdichte



(C) C.Herrmann-Hahn, 2009

Da für den Bereich Mengede die Einwohnerzahlen bezogen auf die Hauskoordinaten nicht zur Verfügung stehen, kann hier keine detailliertere Zuweisung der Bevölkerungsdichte vorgenommen werden.

10 Ermittlung der von Lärmbelastung betroffenen Bevölkerung

Erst die Verknüpfung der statistischen Daten mit anderen Thematiken ermöglicht es, raumbezogene Analysen durchzuführen. Wobei hier die Analyseergebnisse genauer ausfallen, je kleiner die räumliche Bezugseinheit der Daten ist. Je genauer also das untersuchte Phänomen seiner wahren Lage im Raum zugeordnet wird.

Für die Analyse wird die Software ArcGIS der Fa. ESRI verwendet.

Die Software bietet neben der Möglichkeit den erforderlichen Raumbezug herzustellen, verschiedene Funktionen für die Aufbereitung der Daten, der Datenanalyse und der Visualisierung der Daten. Für die Bearbeitung steht die Version ArcGIS 9.2 zur Verfügung.

Die Umgebungslärmrichtlinie sieht vor, dass die Zahlen der von Lärm betroffenen Wohnbevölkerung nach bestimmten, komplexen Berechnungsmethoden erfolgen soll (VBEB – Vorläufige Berechnungsmethode zur Ermittlung der Belastetenzahlen durch Umgebungslärm).

Bei dieser Berechnungsmethode werden an den Gebäudeseiten Fassadenpegel bestimmt. D.h. für jede Gebäudeseite wird an mehreren Punkten errechnet, wie hoch der Lärmpegel ist.

Nach einem bundeseinheitlichen Berechnungsverfahren werden dann die im Haus lebenden Personen gleichmäßig den Fassadenpegeln zugewiesen.

Pro Pegelstufe wird anschließend die Zahl der Betroffenen aufsummiert.

Zur Ermittlung der von Straßenlärm betroffenen Personen stehen für die vorliegende Arbeit die Dichtewerte der Wohnbevölkerung, bezogen auf unterschiedliche Bezugsflächen zur Verfügung.

Die vom Straßenlärm betroffenen Bereiche oberhalb einer Lärmpegelgrenze von 55dB(A) liegen als Isophonenbänder (Polygonflächen, deren Begrenzung Linien gleichen Schallpegels darstellen) vor.

Zur Ermittlung der Betroffenenzahlen wird eine Verschneidung der Daten durchgeführt. (Da die Absolut-Zahlen der Gesamtbevölkerung der verwendeten Datensätze nicht identisch sind, werden die ermittelten Belastetenzahlen im Folgenden als Anteil an der Gesamtbevölkerung in % angegeben.)

10.1 Bezugsfläche Baublock

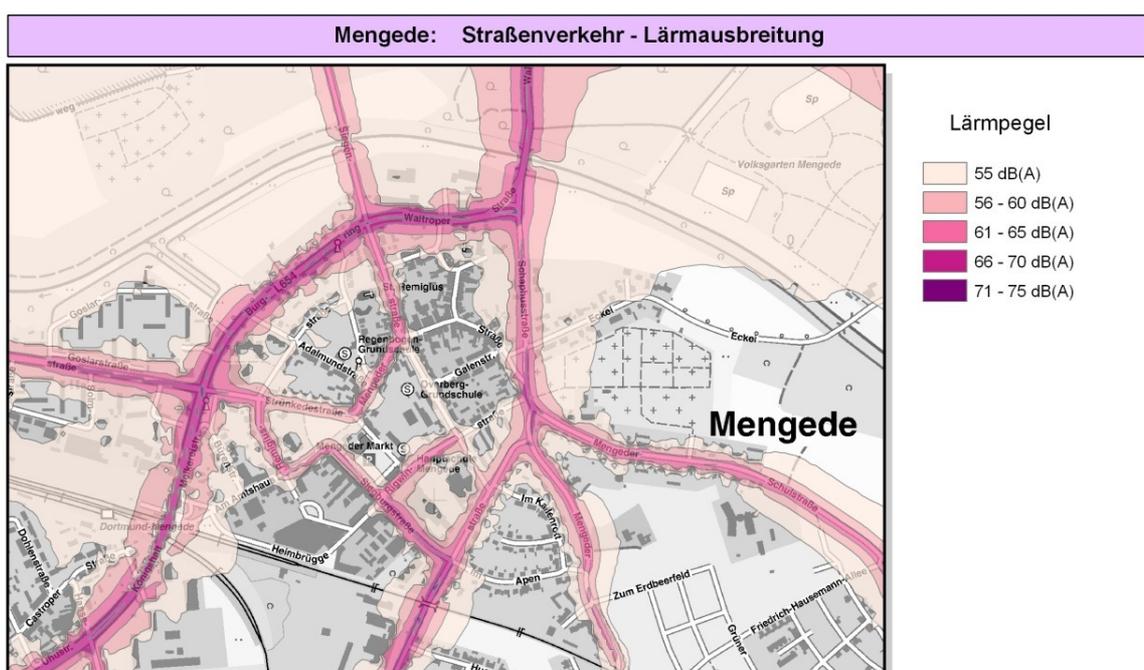
Die Ermittlung der von Lärmbelastung betroffenen Bevölkerung erfolgt über ein Polygon Overlay.

Mit Hilfe eines Polygon Overlay können die Informationen zweier Layer (Informationsebenen) miteinander verknüpft werden.

Zur Bestimmung der Betroffenzahlen werden die Isophonenbänder der Lärmkartierung zusammengefügt und anschließend mit den Polygonen der Baublöcke (Einwohner pro ha) verschnitten.

Abbildung 16 zeigt die Isophonenbänder des durch Straßenverkehr verursachten Lärms.

Abb. 16: Bezugsfläche Baublock: Mengede: Straßenverkehr - Lärmausbreitung



(C) C.Herrmann-Hahn, 2009

Abbildung 17 zeigt die Bevölkerungsdichte in Einwohner pro ha

Abb. 17: Bezugsfläche Baublock: Mengede - Bevölkerungsdichte

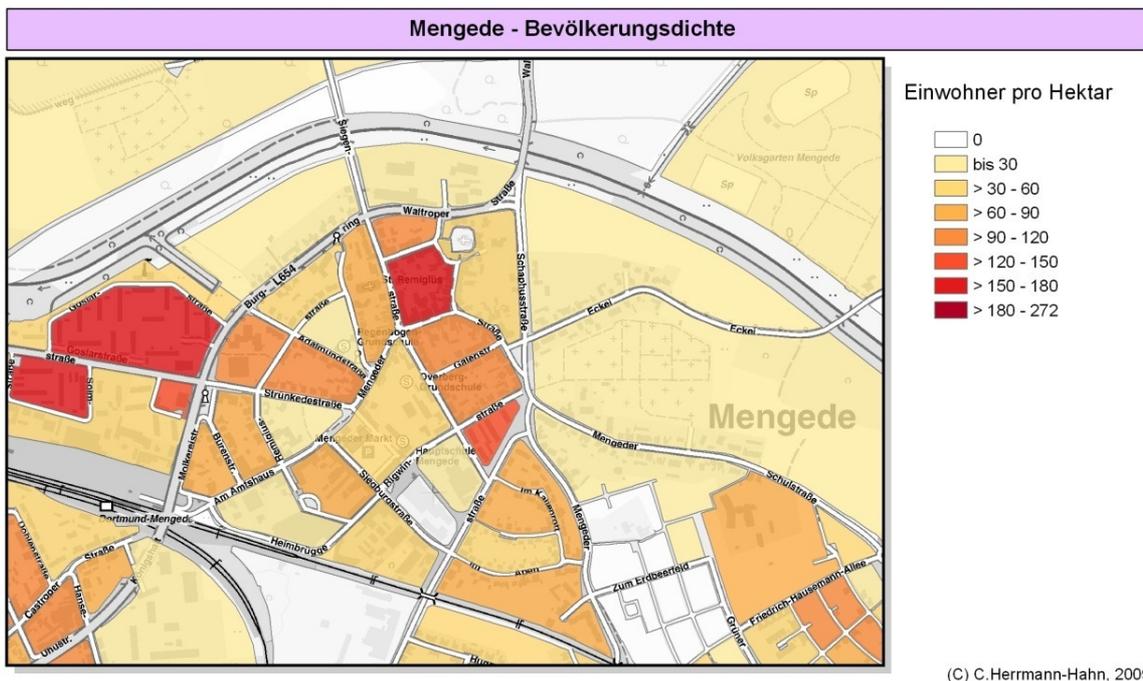
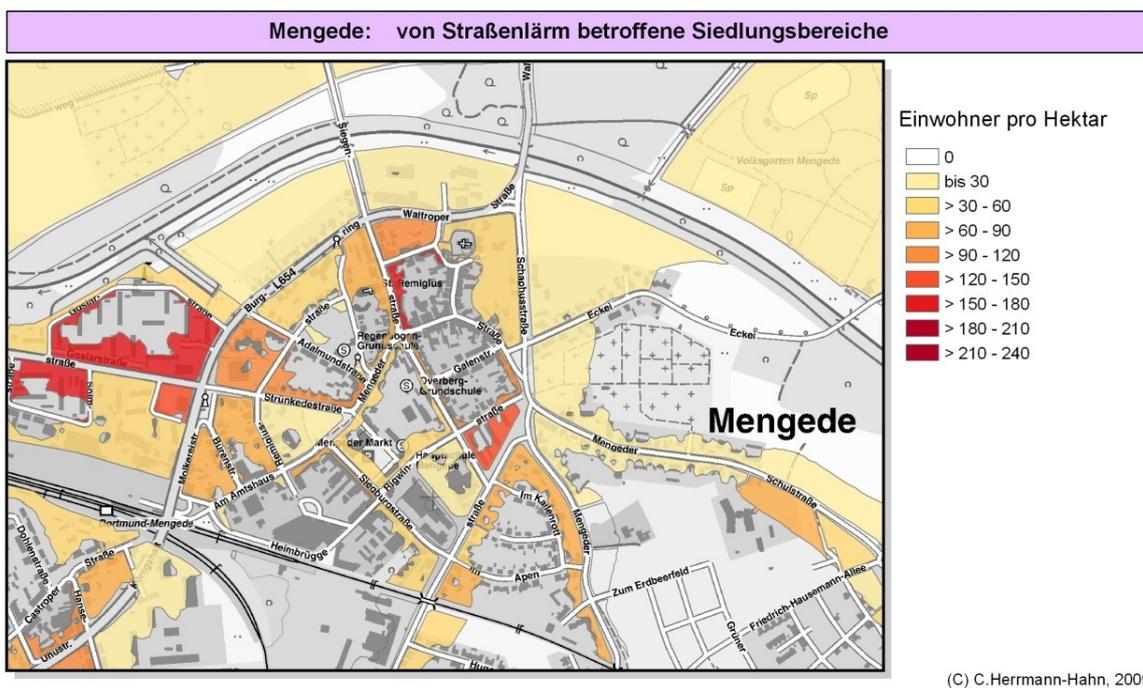


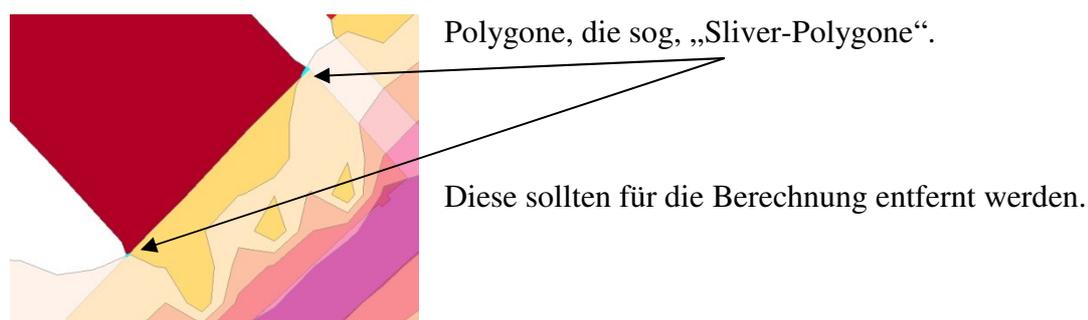
Abbildung 18 zeigt das Ergebnis des Polygon Verschneidung: Die von Straßenlärm betroffenen Siedlungsbereiche.

Abb. 18: Bezugsfläche Baublock: Mengede: von Straßenlärm betroffene Siedlungsbereiche



Über die so entstehenden Polygonflächen kann die Betroffenenzahl berechnet werden. Neben der Berechnung der betroffenen Personen pro Baublock, kann die Gesamtfläche, der von Lärmbelastung beeinträchtigten Siedlungsbereiche ermittelt werden. (Bei dieser Methode wird ein Anteil von 43,83% der Bevölkerung als von Lärmbelastung betroffen errechnet.)

Bei der Verschneidung entstehen allerdings oft neben den gewünschten Teilpolygonen auch kleine Restflächen. Dabei handelt es sich um sehr kleine und/oder langgezogene

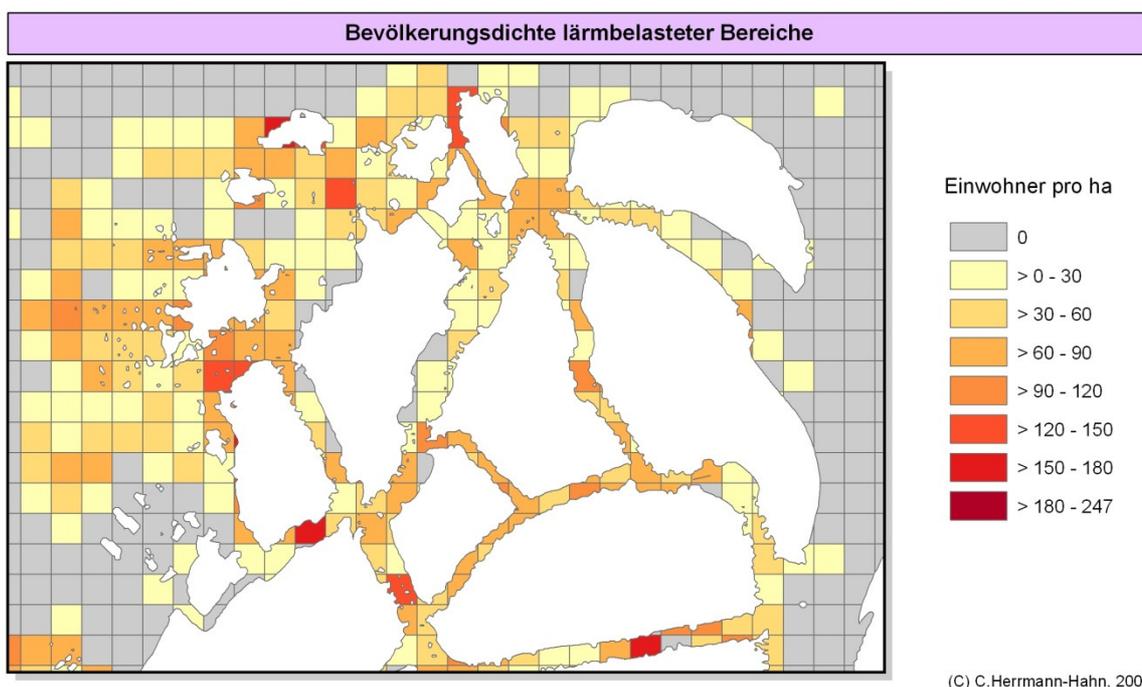


Die räumliche Zuordnung der Einwohnerdichte zu einer administrativen Einheit, wie hier den Baublöcken, kann allerdings nicht die reale Verteilung der Bevölkerung im Raum wiedergeben. So wird bei der o.g. Berechnung der Betroffenenzahlen von der (falschen) Annahme einer homogenen Verteilung der Bevölkerung innerhalb der Baublöcke ausgegangen.

10.2 Bezugsfläche Rasterzelle

Im Vektorrastraster liegen die einzelnen Zellen als Polygone vor. Durch Verschneidung mit dem Polygon der Lärmbelastung können hier die von Lärmbelastung betroffenen Bereiche abgegrenzt werden.

Abb. 19: Bezugsfläche Rasterzelle: Bevölkerungsdichte lärmbelasteter Bereiche



Zur Berechnung der Betroffenzahlen werden die Größen der beschnittenen Polygone berechnet und anschließend über diese Flächen mit Hilfe der Einwohnerdichte die absolute Zahl der betroffenen Personen berechnet.

(Als von Lärmbelastung betroffen werden hier rund 47,29% der Bevölkerung berechnet.)

Die Interpretation der kartographischen Darstellung wird dadurch erleichtert, dass nur die Bereiche innerhalb der Zelle eingefärbt dargestellt werden, die lärmbelastet sind.

10.3 Bezugsfläche Raster Pixel

Die als Vektorraster vorliegenden statistischen Daten können problemlos zu Rasterdaten / Griddaten konvertiert werden.

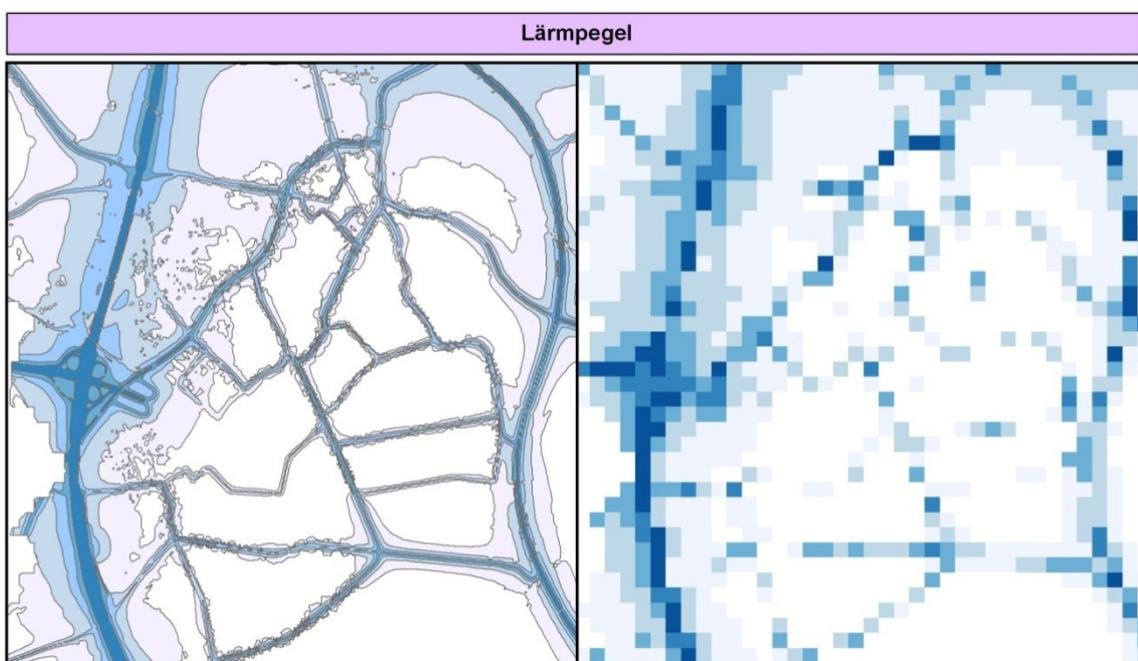
So sind Verknüpfungen dieser Daten mit digitalen Luftbildern oder gescannten Daten möglich. Eine Verschneidung von Rasterdaten erfolgt durch Berechnen eines neuen Output-Rasters unter Verwendung von Inputrastern. Dazu wird aus den Daten der Lärmkartierung zuerst ein einzelner Rasterdatensatz erzeugt.

Mit Hilfe dieses Datensatzes können diejenigen Rasterzellen ermittelt werden, die innerhalb des von Lärm betroffenen Bereiches liegen.

Hierfür werden die Rasterdaten der Lärmkartierung, die aus den Vektordaten erzeugt wurden, als „Maske“ verwendet.

Wie in der Abbildung 14 unschwer zu erkennen, sind die Isophonenbänder durch ein Raster der Auflösung 100m x100m nur lückenhaft wiederzugeben.

Abb. 20: Bezugsfläche Pixel: Lärmpegel



Bei der Konvertierung der Vektordaten in Rasterdaten wird die Sachdateninformation, die die Rasterzelle des Output-Datensatzes erhält, durch den Zellmittelpunkt bestimmt.

D.h. die Mittelpunkte der Rasterzellen werden mit den Vektordaten verschnitten.

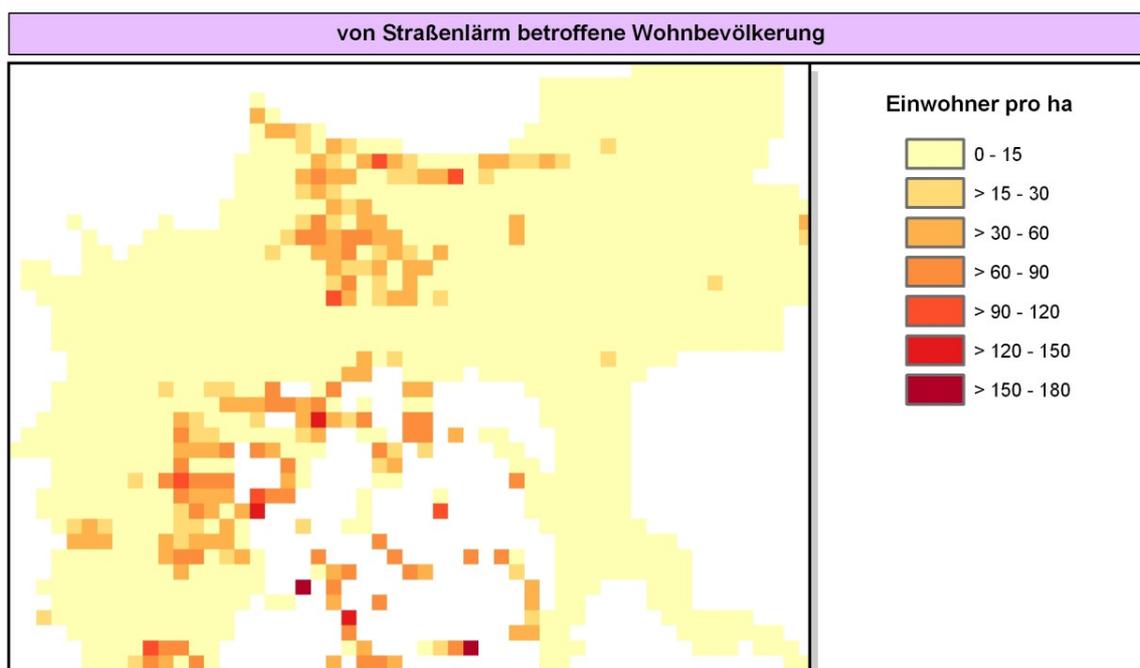
Liegt der Mittelpunkt einer Rasterzelle innerhalb einer Fläche des Vektordatensatzes, so erhält die Zelle als Attributwert diesen Wert. (vgl. de Lange 2002, S.348)

Befindet sich im Rastermittelpunkt keine Fläche, so wird keine Zelle berechnet, kein Wert vergeben.

Somit beeinflusst die gewählte Zellgröße die Darstellung der Form eines Objektes.

Durch die Wahl kleinerer Zellgrößen können die Isophonenbänder des Lärmpegel ihrer realen Lage entsprechend besser modelliert werden, das Zuordnungsproblem wird allerdings nicht durch eine kleinere Zellgröße behoben (vgl. de Lange 2002, S.347).

Abb. 21: Bezugsfläche Pixel: von Straßenlärm betroffene Wohnbevölkerung



(C) C. Herrmann-Hahn, 2009

(Als von Lärm betroffene Bevölkerung werden hier 47,79% der Gesamtbevölkerung ermittelt.)

Unter Verwendung der „Lärmpegelmaske“ werden von den 2911 Zellen, die das Beispielgebiet Mengede abdecken, 1666 Zellen ermittelt, die in Lärm belasteten Bereichen liegen. Für diese Berechnung werden keine Zellen unterdrückt.

Wäre hier eine Mindestfallzahl von 15 Einwohnern pro Zelle einzuhalten, hätten 137 Zellen unterdrückt werden müssen, die hier einer Einwohnerzahl von 621 entsprechen.

Für die kartographische Visualisierung der Ergebnisse, werden die Klassen so gebildet, dass eine Lokalisierung von Zellen, die die gewählte Mindestfallzahl von 15 Einwohnern pro ha nicht erfüllen, unmöglich ist.

11 Methoden zur Wahrung des Datenschutzes – angewendet auf das Beispielgebiet

Für die kartographische Darstellung von Analyseergebnissen ist es ausreichend, die Klassen der Werteverteilung so zu definieren, dass ein Erkennen und somit eine Zuordnung von unterbesetzten Zellen nicht möglich ist.

Bei einer Weitergabe der Daten zur weiteren Bearbeitung allerdings sind datenschutzrechtliche Bestimmungen einzuhalten.

Die Aufbereitung der Daten zur Weitergabe erfolgt innerhalb der Fachbereiche der amtlichen Statistik.

11.1 Löschen von unterbesetzten Zellen

Zur Wahrung des Datenschutzes können, vor der Weitergabe der Einwohnerdaten an Dritte, Zellen, die die Mindestfallzahl nicht aufweisen, gelöscht werden.

Das, den Stadtteil Mengede abdeckende, Raster besteht aus insgesamt 2911 Zellender Größe 100m x 100m, wobei nur rund 28,6 % der Zellen einen Wert aufweisen. Bei den verbleibenden Zellen sind rund 49,5 % mit einer Einwohnerzahl von unter 30 belegt.

Abb. 22 : Anzahl der Zellen mit unter 30 bzw. unter 10 Einwohnern pro ha

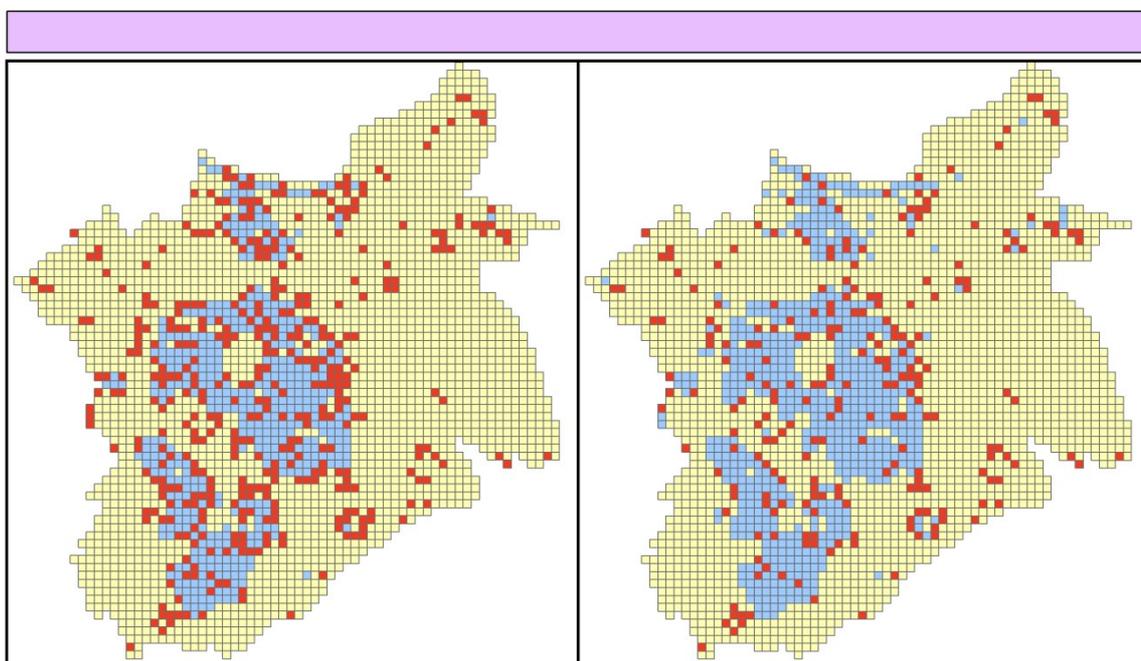


Abbildung 22 zeigt in rot die Zellen mit unter 30 Einwohnern pro ha (linke Abbildung) bzw. unter 10 Einwohnern pro ha (rechte Abbildung).

Die blau gefärbten Zellen zeigen die Einwohnerzahlen von mehr als 10 bzw. mehr als 30 Einwohnern pro ha.

Nimmt man für die weitere Verwendung der Daten eine Mindestfallzahl von 10 Einwohnern pro Zelle an, so würden 238 Zellen entfallen (= 8,17%), bei einer Mindestfallzahl von 30 Einwohnern pro Zelle wären es 412 Zellen(= 14,15%).

Tab. 3: Anteil der zu löschenden, unterbesetzten Zellen

Mindestfallzahl x	Anzahl der Zellen	Einwohnerzahl	Anteil der zu unterdrückenden Zellen an der Gesamtzellenzahl
x < 4	84	166	2,88%
x < 10	238	1.119	8,17%
x < 20	322	2.302	11,06%
x < 30	412	4.501	14,15%
x < 50	555	10.135	19,07%
x = 0	2.911	33.746	0%

Durch das Löschen oder Unterdrücken von unterbesetzten Zellen wird die Absolut-Zahl verändert.

So werden bei einer einzuhaltenden Mindestfallzahlregel von 10 Personen 1.119 Personen nicht erfasst, bei einer Mindestfallzahl von 30 Personen wären es 4.501 nicht berücksichtigte Personen.

Ebenfalls verfälschen würde man die Absolut-Zahl, durch Verwenden der Mindestfallzahl als Wert für unterbesetzte Zellen.

Bezogen auf die Zahlen der Tabelle 4 würden bei einer Mindestfallzahl von 10 insgesamt 1261 Personen zu viel ermittelt werden, bei einer Mindestfallzahl von 30 Personen sogar 7.859 Personen.

11.2 Verwendung variabler Zellgrößen

Um die Unterdrückung von Zellen zu vermeiden, bietet sich die Berechnung der Bevölkerungszahl mittels variabler Zellgrößen an.

Da für den verwendeten Datensatz Gebäude bezogene Daten nicht vorliegen, wird den Zellmittelpunkten der 100m x 100m Zellen die Werte der Zelle zugewiesen.

Anschließend werden Raster in den Kantenlängen 200m, 400m und 800m erzeugt.

Die Werte der jeweiligen Zellgrößen werden als Summe der innerhalb ihrer Umgrenzung liegenden Zellmittelpunkte des 100m x 100m Rasters ermittelt.

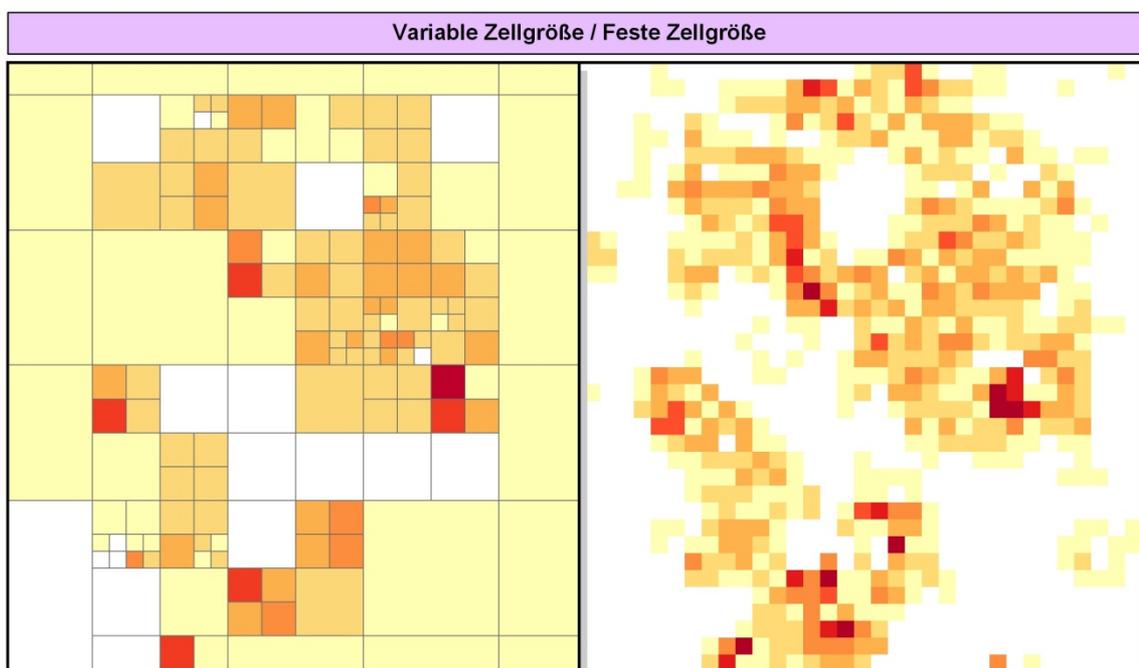
Anschließend werden in den jeweiligen Rasterdatensätzen diejenigen Zellen selektiert, die Zellen der nächstkleineren Größe enthalten, die nicht die Mindestfallzahl (hier: 15) aufweisen. Die Datensätze werden abschließend zu einem Datensatz zusammengefügt.

Abbildung 23 zeigt die Bevölkerungsdichte unter Verwendung von variablen Zellgrößen (Die Mindestfallzahl beträgt hier 15 Personen) gegenüber der Darstellung von Rasterzellen fester Größe.

Da für diese Bearbeitung keine Daten außerhalb des Beispielgebietes vorliegen, sind die Werte der Zellen im Randbereich des Gebietes kritisch zu betrachten.

Hier ragen die Flächen der großen Zellen deutlich über das Beispielgebiet hinaus, da für Bereiche außerhalb Menge keine Daten vorliegen, wird hier die Bevölkerungsdichte viel zu niedrig berechnet. So werden im Randbereich aufgrund fehlender Daten größere Zellen erzeugt, die vorhandene detaillierter Information der kleineren Zellen geht verloren.

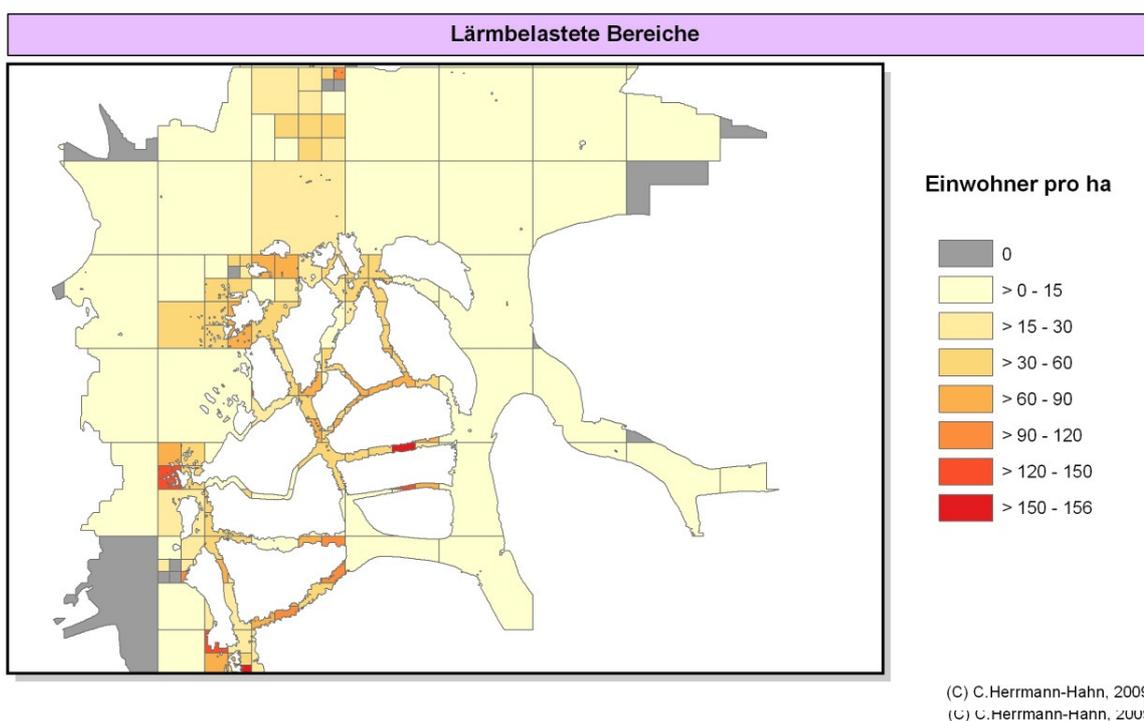
Abb.23: Variable Zellgröße / Feste Zellgröße



Wie die Abbildung 23 zeigt, werden die Siedlungsschwerpunkte bei Verwendung der variablen Zellgröße gut wiedergegeben.

Da es sich beim Datensatz der variablen Zellgrößen um einen Vektordatensatz handelt, kann hier die Anzahl der von Lärm betroffenen Personen über eine Polygon Verschneidung berechnet werden.

Abb. 24: Bezugsfläche variable Zellen: Lärmbelastete Bereiche



(Hier werden 48,12% der Gesamtbevölkerung als lärm-belastete Personen ermittelt.)

Abbildung 24 zeigt das Ergebnis der Polygonverschneidung.

Zur besseren Interpretation des Ergebnisses in der kartographischen Darstellung sind nur die Polygone eingefärbt, in denen eine Lärmbelastung vorliegt.

12 Vergleich der Analyseergebnisse

Die vorangestellten Analysen zeigen, dass bei Verwendung der Dichtewerte bezogen auf administrative Grenzen weniger Personen als lärmbelastet ermittelt werden (hier: durchschnittlich 43,4%), als bei Verwendung von Rasterzellen (hier: durchschnittlich 47,73%).

Da für diese Arbeit nicht die Einwohnerzahlen der einzelnen Adresskoordinaten vorliegen, wird für einen Vergleich von Betroffenenzahlen ein fiktiver Datensatz verwendet.

In diesem Datensatz werden Gebäudemittelpunkten, die über die Geometrie von Gebäudeumringen erzeugt wurden, willkürliche Einwohnerzahlen zugewiesen.

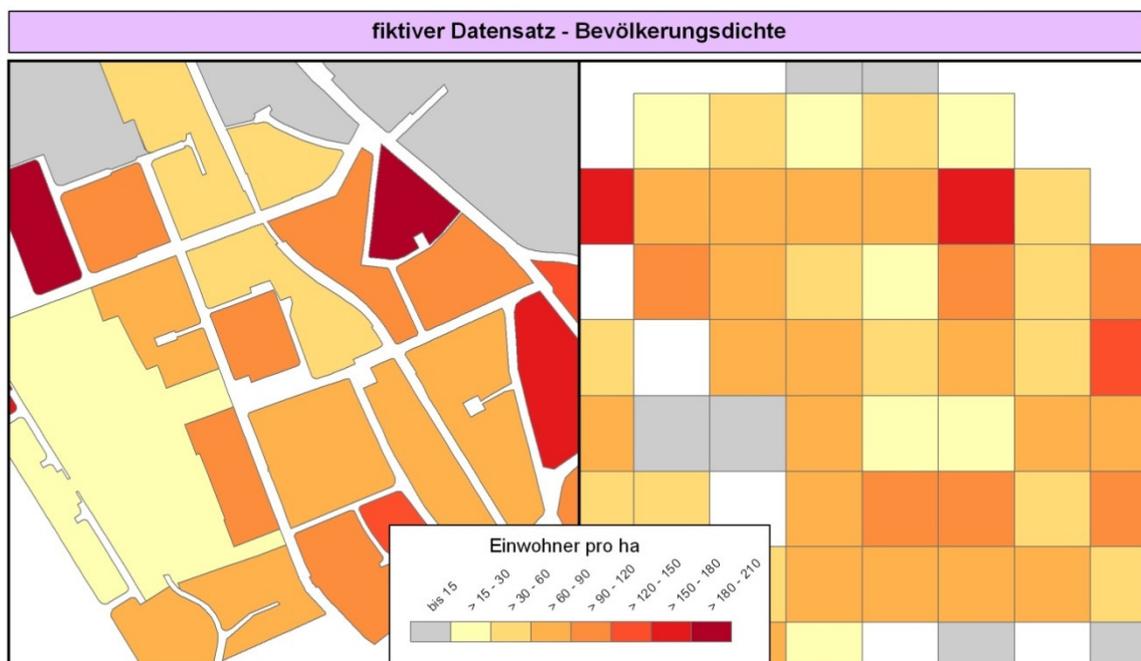
Die Summe der fiktiven Einwohner beträgt 3.734.

Diese fiktiven Einwohnerzahlen werden mit den jeweiligen Baublöcken bzw. Zellen verknüpft.

Die kartographische Darstellung der Bevölkerungsdichte zeigt die Unterschiede der beiden Datensätze:

Während im Datensatz mit Bezug auf Rasterzellen deutlich Siedlungsfreiräume zu erkennen sind, werden diese Bereiche bei dem Bezug auf Baublöcke, durch Mittelwertbildung mit den angrenzenden Bereichen mittlerer Einwohnerdichte, mit geringer Einwohnerdichte dargestellt.

Abb. 25: fiktiver Datensatz - Bevölkerungsdichte



Der Datensatz besteht aus 33 Baublöcken, deren Flächen von 0,33 ha bis 21,2 ha variieren.

Der Rasterdatensatz besteht aus insgesamt 66 Zellen der Größe 100m x 100m.

Davon sind 12 Zellen mit einer Personenanzahl von unter 15 besetzt.

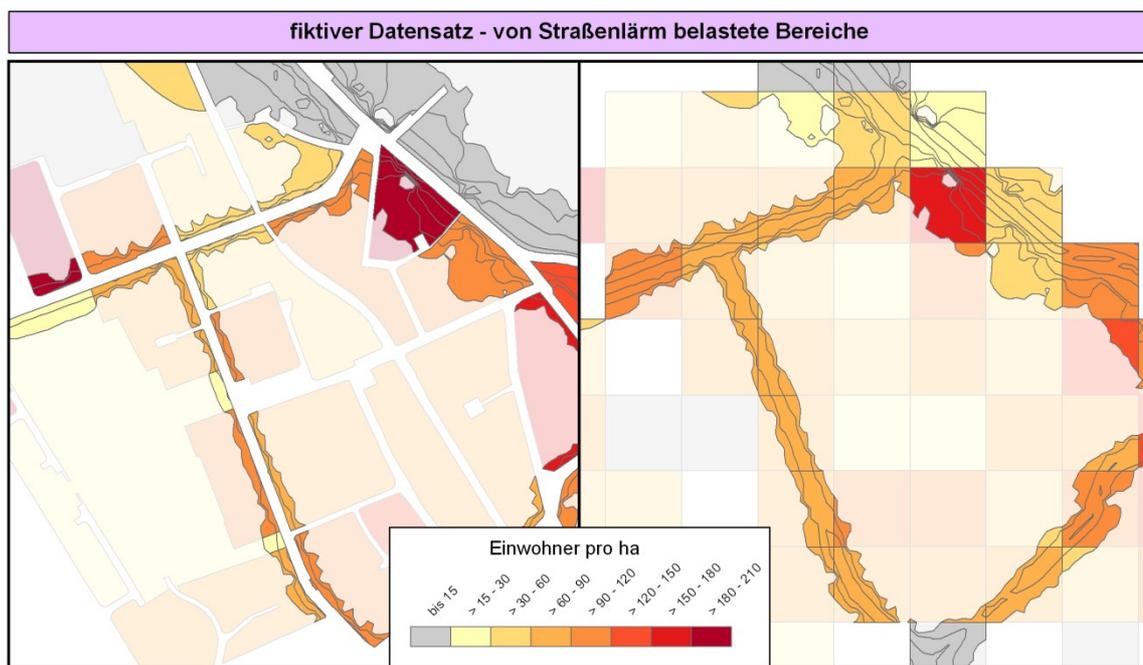
Würden diese Zellen gelöscht, würden insgesamt 61 Personen aus der Berechnung entfernt werden (ca. 1,6 %).

Für beide Datensätze werden die von lärmbelasteten Bereiche ermittelt und die Betroffenenzahlen für die jeweiligen Lärmpegelbereiche berechnet.

Abbildung 26 zeigt die für beide Datensätze ermittelten lärmbelasteten Bereiche.

Im Datensatz mit Bezug auf die Baublöcke werden die Straßenzüge ausgespart, während die Dichtewerte, bezogen auf die Rasterzellen, flächendeckend erfasst werden.

Abb. 26: fiktiver Datensatz – von Straßenlärm belastete Bereiche (Bezugsfläche Baublock / Rasterzelle)



(C) C.Herrmann-Hahn, 2009

Des weiteren werden die Betroffenenzahlen ermittelt für einen Datensatz, für den variable Zellgrößen berechnet wurden.

Problematisch ist hier, dass die größeren Zellen weit über das eigentliche Bearbeitungsgebiet hinausragen.

Für diese Bereiche liegen keine Daten vor, so dass hier die Bevölkerungsdichte, bezogen auf die Zellgröße, mit einem sehr geringen Wert berechnet wird.

Da die Absolut-Zahlen über die Fläche und die Einwohnerzahl pro Hektar berechnet werden, ist dieser Datensatz in den Grenzen der 100m x 100m Zellgröße beschnitten, um ihn mit den Ergebnissen der anderen Berechnungen vergleichen zu können

Abb. 27: fiktiver Datensatz – von Straßenlärm belastete Bereiche (Bezugsfläche variable Zellen)

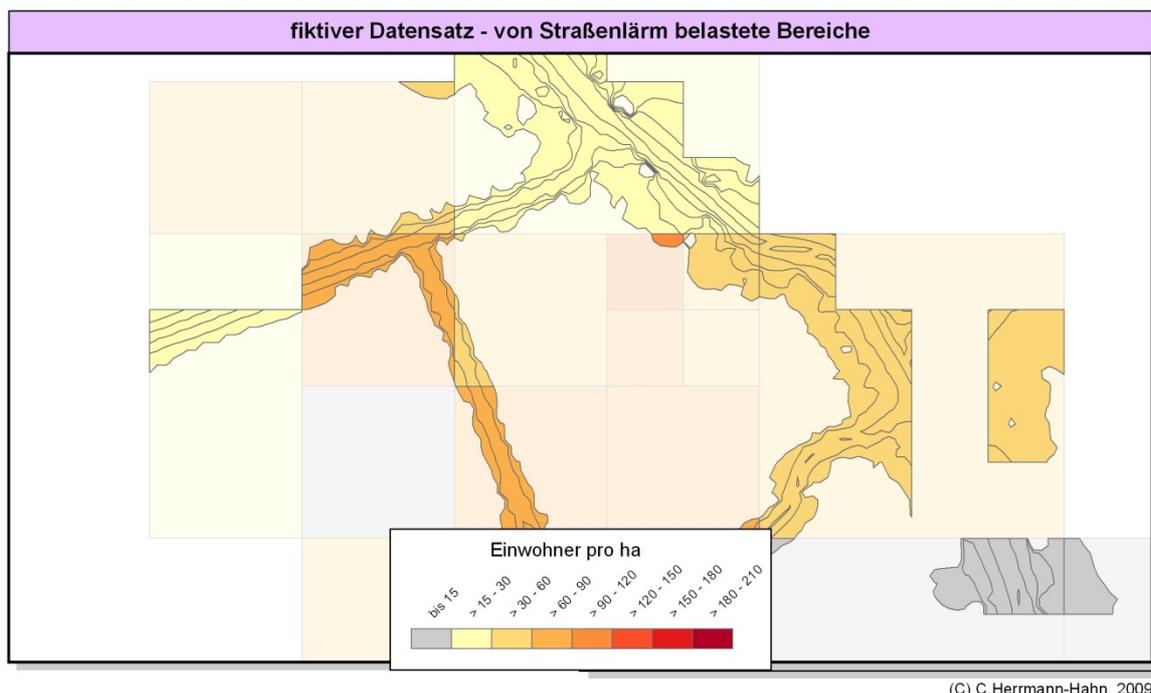


Tabelle 4 zeigt die ermittelten Belastetenzahlen.

Als Kontrollgröße ist die Anzahl der lärmbelasteten Personen pro Gebäude gewählt.

Hier ist entsprechend der von Lärm betroffenen Gebäudefläche anteilig die von Lärm betroffene Personenanzahl berechnet.

Tab. 4: Von Lärm belastete Personen

Lärmpegel in dB(A)	Bezugsfläche Baublock	Bezugsfläche Rasterzelle 100 x 100m	Bezugsfläche Variable Zellegröße	Bezugsfläche Gebäude
55	476,26 (51,48 %)	495,27 (41,99 %)	372,67 (45,75 %)	682,00 (58,38 %)

56-60	270,21 (29,21 %)	360,60 (30,57 %)	239,97 (29,46 %)	351,71 (30,10 %)
61-65	116,86 (12,63 %)	194,20 (16,46 %)	119,60 (14,68 %)	97,44 (8,34 %)
66-70	57,77 (6,25 %)	90,11 (7,64 %)	54,80 (6,73 %)	33,65 (2,88 %)
71-75	3,94 (0,43 %)	39,29 (3,34 %)	27,55 (3,38 %)	3,47 (0,30 %)
insgesamt	925,04 (100 %)	1179,47 (100 %)	814,59 (100 %)	1168,27 (100 %)
Anteil der betroffenen Bevölkerung an der Gesamtbevölkerungszahl	24,77 %	31,59 %	21,81 %	31,29 %

Wie die Tabelle zeigt, liegt die Absolut-Zahl der Betroffenen, ermittelt über die Bezugsfläche Rasterzelle, am nächsten an dem Wert, ermittelt über die Gebäudefläche, der als realistisch angenommen wird.

Die Personenzahlen der einzelnen Pegelbereiche variieren allerdings.

Die Bezugsflächen Baublock und variable Zelle sind bei dieser Berechnung nahezu gleich zu setzen.

Da sich die Dichtewerte auf die gesamte Baublockfläche bzw. Fläche, der Zellen variabler Zellgröße, beziehen, sind sie insgesamt niedriger berechnet, als die Dichtewerte der Zellen mit 100m x 100m Größe. Damit fällt hier auch die Summe der Belastetenzahlen geringer aus, als die angenommene realistische Zahl.

Tab. 5: Betroffenenanzahl bei Anwendung einer Mindestfallregel

Lärmpegel in dB(A)	Bezugsfläche	Bezugsfläche	Bezugsfläche	Bezugsfläche
	Rasterzelle Ohne Mindestfallzahl- beschränkung	Rasterzelle Mindestfallzahl 10 Personen	Rasterzelle Mindestfallzahl 30 Personen	Gebäude
55	495,27 (41,99 %)	484,72 (41,78 %)	474,78 (41,71 %)	682,00 (58,38 %)
56-60	360,60 (30,57 %)	356,02 (30,68 %)	352,27 (30,95 %)	351,71 (30,10 %)
61-65	194,20 (16,46 %)	191,37 (16,49 %)	188,38 (16,55 %)	97,44 (8,34 %)
66-70	90,11 (7,64 %)	89,08 (7,68 %)	86,43 (7,59 %)	33,65 (2,88 %)
71-75	39,29 (3,34 %)	39,12 (3,37 %)	36,34 (3,19 %)	3,47 (0,30 %)
insgesamt	1179,47 (100 %)	1160,31 (100 %)	1138,2 (100 %)	1168,27 (100 %)
Anteil der betroffenen Bevölkerung an der Gesamt- bevölkerung	31,59 %	31,40 %	32,35 %	31,29 %

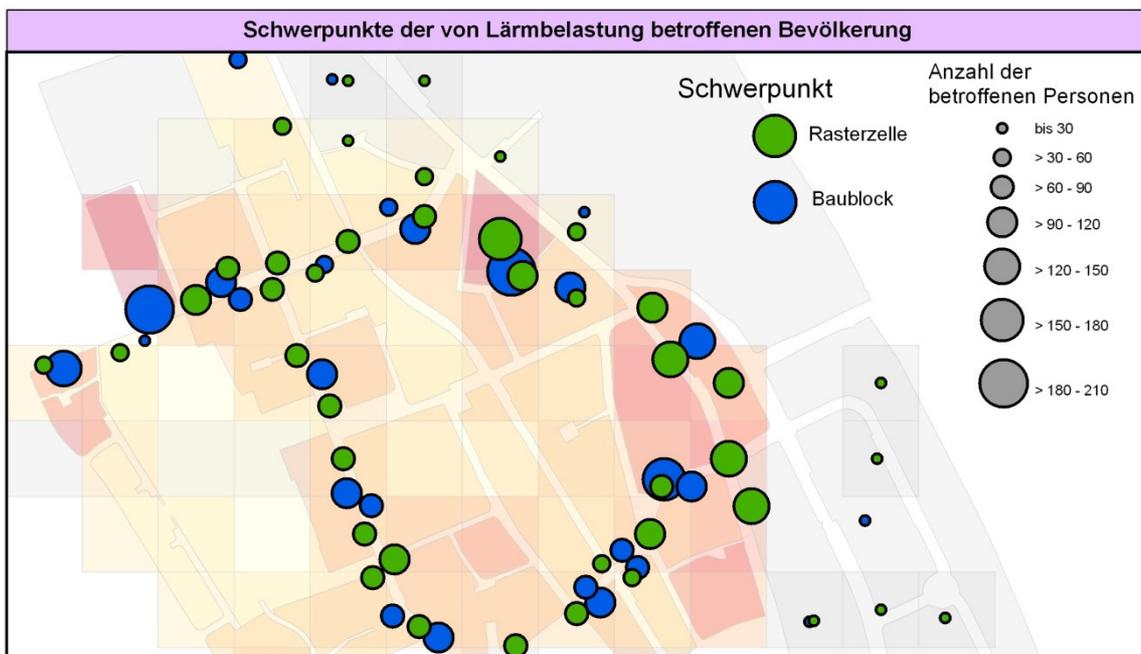
Tabelle 5 zeigt den Unterschied der ermittelten lärmbelasteten Personen bei Anwendung einer Mindestfallzahl von 10 Personen bzw. 30 Personen.

Hierbei beträgt die Gesamtzahl der Wohnbevölkerung 3.695 Personen bei Unterdrückung von Zellen mit einem Wert unter 10, bei einer Unterdrückung der Werte unter 30 werden 3.518 Personen errechnet.

Der prozentuale Anteil der Belasteten bleibt dabei annähernd gleich.

Unterschiede zwischen der Verwendung der administrativen Grenzen als Bezugsfläche und der Verwendung von Rasterzellen sind auch bei der Ermittlung der Schwerpunkte der von Lärmbelastung betroffenen Bevölkerung erkennbar:

Abb. 28: Schwerpunkte der von Lärmbelastung betroffenen Bevölkerung



(C) C.Herrmann-Hahn, 2009

Hier sind die Centroiden der jeweiligen lärmbelasteten Bereiche erzeugt.

Die Kreise zeigen die absoluten Zahlen der lärmbelasteten Bevölkerung, wobei der Kreisdurchmesser entsprechend der Anzahl in abgestuften Größen dargestellt ist.

Die größten Unterschiede zeigen sich in den Bereichen mit geringer Einwohnerdichte. Während bei der Rasterdarstellung der Schwerpunkt der belasteten Bevölkerung pro ha visualisiert wird, und somit hier 44 Schwerpunkte beschrieben werden, verteilen sich die insgesamt 28 Schwerpunkte der Baublöcke weiter über das Untersuchungsgebiet.

13 Ergebnisdiskussion

Die Beurteilung der aufgestellten Hypothesen erfolgt durch die Beantwortung der dazu eingangs definierten Fragen bzgl. geometrischer Genauigkeit, Handhabung der Daten sowie Interoperabilität der Daten.

- Die geometrische Genauigkeit bezieht sich hierbei auf die durchschnittliche Lagetreue der Eingangsdaten (hier: Gebäudekoordinaten) bezogen auf die jeweilige Bezugsfläche.

Dazu ist die Differenz zwischen den Gebäudeschwerpunkten und den Mittelpunkten der 100m x 100m Rasterzellen bzw. Schwerpunkten der Baublöcke ermittelt.

Sie zeigt, dass die maximale Lageabweichung bezogen auf die Bezugsfläche Rasterzelle im Voraus über die Zellengröße bestimmbar ist. Die Lageabweichung zum Zellenmittelpunkt kann maximal die Länge vom Mittelpunkt zu einer Zellenecke betragen. Bezogen auf die gesamte Zelle kann die Abweichung maximal den Wert der Länge von zwei gegenüberliegenden Eckpunkten erreichen.

Da die Baublöcke als Flächen, begrenzt durch administrative Grenzen, sehr unterschiedlich in Größe und Form sein können, variieren die Lageabweichungen innerhalb eines Datensatzes sehr stark.

Auch hier kann die Abweichung maximal den Wert der größten Ausdehnung des Baublocks annehmen.

- Bei einem Vektorraaster können die vielfältigen Analysefunktionen, die jedes GIS für Vektordaten zur Verfügung stellt, angewendet werden.

Die Rasterquadrate können darüberhinaus ohne Informationsverlust zu Grid-Daten umgerechnet werden. Damit sind sie als Grundlage für Analysen mit anderen Grid-Daten (Rasterdaten) aus Orthophotos oder gescannten Vorlagen verwendbar.

Die Umrechnung von Polygonen, begrenzt durch administrative Grenzen, ist dagegen mit einem Verlust der Form- und Lagetreue der Objekte verbunden.

Wobei die Güte der Form- und Lagebeschreibung durch die Pixelgröße beeinflusst wird.

- Die Interoperabilität der Daten ist nicht nur festzumachen an technischen Standards. Diese werden über OGC Standards definiert, um den reibungslosen Austausch von Daten zu gewährleisten.

Für die Verwendung von rasterbasierten statistischen Daten ist die Definition eines einheitlichen Referenzsystems für das Rastergitter ausschlaggebend.

Flächenhafte Phänomene machen nicht an der Landesgrenze hat.

Durch das Verwenden eines einheitlichen Referenzsystems sind der Vergleich und das Zusammenführen von Daten verschiedener Länder möglich.

Die Einteilung der administrativen Bezugseinheiten in NUTS- bzw. LAU-Regionen dient vorrangig der eindeutigen Identifizierung der Gebietseinheiten.

Da die weitere Untergliederung unterhalb der Gemeinden innerhalb der Bundesrepublik Deutschland aber auch in den anderen EU-Ländern unterschiedlich gehandhabt wird, ist hier ein Vergleich der Regionen erschwert.

Die Verwendung von Rasterzellen ermöglicht eine präzisere Darstellung der Bevölkerungsverteilung, als die ungleich großen administrativen Einheiten.

Schwerpunkte der Einwohnerdichte können besser erkannt werden.

Durch Verwenden der Gebäudeumringe als Bezugsfläche für die Einwohnerdichte kann die Bevölkerungsverteilung noch aussagekräftiger visualisiert werden.

Eine Gewährleistung des Datenschutzes bei Verwendung von Rasterzellen als Bezugsflächen ist nur mit Einschränkungen bzgl. der Genauigkeit der zur Verfügung stehenden Daten möglich.

Die hier verwendeten statistischen Daten beziehen sich auf die Gesamtbevölkerungszahl.

Bei selektierten Daten, die eine kleinere Grundgesamtheit darstellen, ist die Mindestfallzahl schnell erreicht.

Variable Zellgrößen bieten hier die Möglichkeit, ohne Unterdrückung von Werten, Phänomene darzustellen, sind aber für zeitliche Vergleiche ungeeignet.

Da die Lärmkartierung zur Beurteilung der durchgeführten Lärminderungsmaßnahmen in regelmäßigen Abständen (alle 5 Jahre) wiederholt werden soll, sind für den Vergleich und zur Beobachtung von Entwicklungen, der Anzahl der von Lärmbelastung betroffenen Personen, Zellen fixer Größe besser geeignet.

Für die hier gewählte Rastergröße von 100m x 100m blieben allerdings bei Anwendung einer Mindestfallzahl von 30 Personen rund 6% der Bevölkerung unberücksichtigt.

Diese Zahl ist nur durch die Wahl größerer Zellen zu verringern.

Bei der Ermittlung der Belastetenzahlen mit Bezug Baublock, sowie der Bezugsfläche Rasterzelle werden Dichtewerte verwendet. Damit kann die Belastetenzahl nicht exakt ermittelt, sondern nur näherungsweise bestimmt werden.

Aufgrund der homogenen Grundfläche der Rasterdaten kann aber die hier ermittelte Belastetenzahl als realistischer angenommen werden.

14 Zusammenfassung / Ausblick

In der amtlichen Statistik wird eine Vielzahl von raumbezogenen Daten erhoben. Durch Kombination dieser Daten mit anderen Geodaten oder durch Durchführen raumbezogene Analysen erfahren die Daten einen Mehrwert: sie werden anwendungsspezifisch aufbereitet und/oder es entstehen neue, hochwertigere Informationen.

Für die Nutzung dieses Potentials an Informationen ist die Zugänglichkeit und Verfügbarkeit der Geodaten entscheidend.

So bestehen bei der Verwendung von Daten der Statistik, insbesondere bei Daten der Bevölkerungsstatistik, datenschutzrechtliche Bestimmungen, die einzuhalten sind, um das Recht des Einzelnen auf informelle Selbstbestimmung zu gewährleisten.

Ziel muss es daher sein, für die Verwendung dieser Daten einen Kompromiss zu finden, zwischen der Notwendigkeit präziser Datengrundlagen und der Einhaltung von Datenschutzbestimmungen.

Die Raumplanung braucht für eine stetige Beobachtung des Raumes geeignete Messgrößen.

So ist, um eine nachhaltige Raumplanung zu sichern, die Definition von Indikatoren erforderlich.

Mit Hilfe des Indikators, der einen messbaren Sachverhalt darstellt, kann die Veränderung des untersuchten Phänomens über einen längeren Zeitraum hinweg beobachtet werden.

So kann, nach Erfassung und Beschreibung des Ist-Zustands, die Wirksamkeit planerischer Maßnahmen zur Erreichung des Soll-Zustandes beurteilt werden.

Bei Verwendung geeigneter Messgrößen können auch kleinräumige Trendentwicklungen dargestellt werden.

Darüberhinaus sollte die Messgröße für unterschiedliche räumliche Auflösungen geeignet sein, so dass vertikale Vergleiche möglich sind.

Die Güte, der von den Indikatoren abzuleitenden Erkenntnisse, ist dabei abhängig von der Aussagekraft der verwendeten Input-Daten.

Je exakter in einem Datenmodell die reale Lage raumbezogener Phänomene wiedergegeben wird, desto aussagekräftiger sind die Analyseergebnisse.

Idealerweise wären hier Daten, bezogen auf die Wohngebäudekoordinate als Input-Daten zu verwenden.

Diese Daten liegen aber nicht flächendeckend vor und/oder sind aus datenschutzrechtlichen Gründen nicht zugänglich.

Die Verwendung rasterbasierter statistischer Daten für räumliche Analysen stellt eine Alternative dar, zur Verwendung statistischer Daten auf der Basis von administrativen Grenzen.

Hier kann die Lageabweichung zwischen der Position des Phänomens in der realen Welt und seiner Abbildung im Modell maximal die Größe der gewählten Rasterzelle betragen.

Im Gegensatz zur Verwendung von administrative Grenzen als Bezugsflächen für statistische Daten.

Administrative Grenzen sind historisch gewachsen und teilen den Untersuchungsraum in Gebiete unterschiedlicher Größe und Form ein. Die Lageabweichung kann hier mehrere Kilometer betragen.

Durch Verwendung von Rastern kann die Bevölkerungsverteilung modelliert werden, ohne administrative Grenzen berücksichtigen zu müssen.

Administrative Grenzen liegen meist auf Straßenzügen, sie zerschneiden die Bebauungsflächen und verfälschen dadurch im ungünstigsten Fall die wahre Verteilung der Werte.

Rasterzellen können für unterschiedliche Analysemaßstäbe hierarchisch aufgebaut werden und ihre Größe ist, entsprechend dem angestrebten Detaillierungsgrad der Analyse, frei zu wählen.

Die Größe der Rasterzelle erfährt allerdings durch datenschutzrechtliche Bestimmungen Einschränkungen. So darf es anhand der verwendeten statistischen Daten nicht möglich sein, Informationen einzelnen Personen zuzuordnen.

Dies kann durch die Anwendung einer Mindestfallzahl-Regel erreicht werden.

Zur Sicherung der Qualität der Analyse sollte ein Unterdrücken oder Verfälschen der Zellen, die nicht die Mindestfallzahl aufweisen, vermieden werden.

Andernfalls wird die Absolut-Zahl so stark verändert, dass die, bei der Analyse ermittelten Werte, in Frage gestellt werden müssen.

Alternativ kann der Anwender die Zellgröße so groß zu wählen, dass keine Zelle unterbesetzt ist, oder er verwendet variable Zellgrößen. Wobei Datensätze, in denen variable Zellgrößen verwendet werden, nicht geeignet sind, um Entwicklungen über einen längeren Zeitraum zu beobachten, da sich die Zellgrößen, je nach Berechnung, verändern.

Entscheidende Rahmenbedingungen für die Verwendung von rasterbezogenen statistischen Daten sind schon vorhanden:

Zur Harmonisierung der Geodaten sind die Länder der europäischen Union aufgefordert, entsprechend der INSPIRE Direktive ein harmonisiertes Raster mit multipler Zellgröße mit festem Ursprung in einer europaweit anzuwendenden Projektion zu definieren.

Darüberhinaus stellt das amtliche deutsche Vermessungswesen flächendeckend Hauskoordinaten und Hausumringe bereit.

Damit können statistische Informationen gebäudescharf erfasst werden.

Und anschließend zu Rasterzellenwerten aggregiert werden.

Auch OGC-konforme Standards für den Aufbau von Web Coverage Processing Services (WCPS) stehen zur Verfügung.

Aber:

In der Bundesrepublik Deutschland wird die Verwendung statistischer Raster zwar seit einiger Zeit diskutiert, für die Umsetzung fehlen aber derzeit noch die entsprechenden gesetzlichen Regelungen.

Wobei es allein mit der Festlegung einer Mindestfallzahl-Regel nicht getan ist.

Wie das Gutachten von Forgó et.al. (2008) beschreibt, ist in den verschiedenen Gesetzestexten die Handhabung personenbezogener Daten nicht eindeutig geregelt.

Die bisherigen Festlegungen bieten zu viel Interpretationsspielraum bzgl. der Frage, wann ein Datenschutz relevanter Personenbezug vorliegt.

Literaturverzeichnis

- Artikel-29-Datenschutzgruppe (2007). *Stellungnahme 4/2007 zum Begriff "personenbezogene Daten"*. WP 136. URL:
http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_de.pdf
[Stand 13.07.2009]
- Babisch, W. et.al. (April 2007). *Lärm das unterschätzte Risiko*. Dessau. URL:
<http://www.umweltdaten.de/publikationen/fpdf-l/3203.pdf> [Stand 13.07.2009]
- Baumann, Peter (2008). Standardisierte Mehrwertdienste für GMES-, In Situ- und Simulationsdaten: *zfv: Wißner*. (Heft 4), S. 262 - 268.
- Bartelme, Norbert (1995). *Geoinformatik: Modelle, Strukturen, Funktionen*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg
- Bartelme, Norbert (2005). *Geoinformatik: Modelle, Strukturen, Funktionen*. 4., vollständig überarbeitete Auflage. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. (Springer-11777 /Dig. Serial).
- Berry, Joseph K. (1996 bis 2007). *Beyond Mapping III - online book: Procedures and Applications in GIS Modeling*. URL:
<http://www.innovativegis.com/basis/MapAnalysis/> [Stand 13.07.2009]
- Bill, Ralf (Hg.) (1994). *Hardware, Software und Daten*. 2. Aufl. Heidelberg: Herbert Wichmann Verlag GmbH.
(Grundlagen der Geo-Informationssysteme / R. Bill, D. Fritsch, Bd. 1).
- Bill, Ralf & Fritsch, Dieter (Hrsg.) (1996). *Analysen, Anwendungen und neue Entwicklungen*. Karlsruhe: Wichmann (Grundlagen der Geo-Informationssysteme / R Bill, D. Fritsch, Bd. 2).
- Bill, Ralf & Zehner, Marco L. (Hrsg.) (2001). *Lexikon der Geoinformatik*. Heidelberg: Wichmann
- Bundesamt, Statistisches (1987). Gesetz über die Statistik für Bundeszwecke: Bundesstatistikgesetz - BStatG_Online im Internet: URL:
http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/SharedContent/Oeffentlich/AZ/ZD/Rechtsgrundlagen/Statistikbereiche/AllgemeineBestimmungen/010__BStatG,property=file.pdf [Stand 13.07.2009]
- Bundesdatenschutzgesetz: BDSG: 05.02.2009. URL:
http://bundesrecht.juris.de/bundesrecht/bdsg_1990/gesamt.pdf [Stand 27.07.2009]
- Burrough, P.A.(1986). *Principals of Geographical Information Systems for Land Resource Assessment*. Oxford: Clarendon Press

- commission of the european communities (2004). Directive of the european parliament and of the council: INSPIRE - establishing an infrastructure for spatial information in the Community_Online im Internet: URL:
<http://inspire.jrc.ec.europa.eu/proposal/EN.pdf> [Stand 13.07.2009]
- Dickmann, Frank & Sohst, Angela (2008). Vom Baublock zum Grid - Die Visualisierung künstlicher Raumeinheiten im Geomarketing. *Kartographische Nachrichten* 58.Jahrgang (Heft 5), S. 227-234.
- Eschwege, Arndt von & Heidrich-Riske, Holger (2006). Nutzung des Raumbezuges in der amtlichen Statistik, in Statistisches Bundesamt, Wiesbaden (Hg.): *Auszug aus Wirtschaft und Statistik*, S. 118-135.
- ESRI ArcGIS Desktop Help 9.2
- ESRI GIS Dictionary (2006)
URL:<http://support.esri.com/index.cfm?fa=knowledgebase.gisDictionary.search&searchTerm=GIS> [Stand 14.07.2009]
- EU 2002. RICHTLINIE 2002/49/EG DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 25.Juni 2002 über die Bewertung und Bekämpfung von Umgebungslärm: *Amtsblatt der Europäischen Gemeinschaften*. URL:
<http://www.umgebungslaerm.nrw.de/Dokumente/Gesetze/EU-Richtlinie.pdf> [Stand 13.07.2009]
- Fitzke, Jens (1997). Entwicklung eines GIS-Prototyps zur Quantifizierung von Straßenlärmbelastung auf der Grundlage von Schallimmissions- und Einwohnerstrukturdaten, in Dollinger, Franz & Strobl, Josef (Hrsg.): *Angewandte Geographische Informationsverarbeitung IX*. Salzburg: Selbstverlag des Instituts für Geographie der Universität Salzburg
- Forgó, Nikolaus, Krügel, Tina & Reiners, Nico (2008). *Forschungs- und Entwicklungsauftrag zum Thema Geoinformation und Datenschutz (GEODAT)*. URL:
http://www.iri.uni-hannover.de/tl_files/pdf/Gutachten%20GEODAT.pdf [Stand 13.07.2009]
- Giessing, Sarah & Dittrich, Stefan (2006). Tabellengeheimhaltung im statistischen Verbund - ein Verfahrenvergleich am Beispiel der Umsatzsteuerstatistik, in Statistisches Bundesamt, Wiesbaden (Hrsg.): *Auszug aus Wirtschaft und Statistik*, S. 805 - 814.
- Gröger, Gerhard (2000). *Modellierung raumbezogener Objekte und Datenintegrität im GIS*. Heidelberg: Wichmann
- Hake, Günter (1985). *Kartographie II: Thematische Karten, Atlanten, kartenverwandte Darstellungen, Kartenredaktion und Kartentechnik, rechnergestützte Kartenherstellung, Kartenauswertung, Kartengeschichte*. 3., neubearb. Aufl., 2 Bde. Berlin, New York: Walter de Gruyter, Sammlung Göschen. (Bd.2).

- Helmcke, Thomas (2008). Regionalstatistik auf europäischer und nationaler Ebene, in Statistisches Bundesamt, Wiesbaden (Hrsg.): *Auszug aus Wirtschaft und Statistik*, S. 207 - 216
- Helmetberger, Markus & Ortner, Simon (2008). Erreichbarkeitsbasierte Raster Raumanalyse-Anwendungen in der Landesplanung, in Schrenk, Manfred, u.a. (Hrsg.): *Beiträge zum Symposium CORP 2008*. Wien, S. 255 - 263.
- Helmholtz Zentrum München 2008. *Lärm: Krach, der uns krank macht*. URL: <http://www.helmholtz-muenchen.de/fileadmin/FLUGS/PDF/Themen/Laerm/Laerm.End.pdf> [Stand 13.07.2009]
- Herter, Michael (2008). Geodaten sind nicht persönlich: Geomarketing als sichere Alternative zu persönlichen Analysen. *Geomarketing - Magazin für Markt und Raum* 9. Jahrgang (Heft 4), S. 8-9.
- Hillen, Richard (2008). *Erläuterungen zur Kartierung von Geräuschen im Rahmen der EU-RL Umgebungslärm in NRW*. URL: <http://www.umgebungslaerm-kartierung.nrw.de/laerm/erlaeuterung/erlaeuterung.pdf> [Stand 13.07.2009]
- Höferl, Karl-Michael; Kalasek, Robert; Seher, Walter; Wonka, Erich & Weber Gerlind (2007). Stadt, Land, Fluss - zur Strukturtypenbildung ländlicher Räume in Österreich, in Strobl, Josef; Blaschke, Thomas & Griesebner, Gerald (Hrsg.): *Angewandte Geoinformatik 2007: Beiträge zum 19. AGIT-Symposium Salzburg [vom 3. bis 6. Juli 2007]*. Heidelberg: Wichmann, S. 271-276.
- INSPIRE Thematic Working Group on Reference Systems 2008. *D2.8.I.1 Specifications on Coordinate Reference Systems -Draft Guidelines*. URL: http://www.gdi-de.org/de_neu/inspire/sharing_review_annex1.html [Stand 13.07.2009]
- JRC-IES-LMU-ESDI, Editor Annoni, Alessandro. (2004). *1st Workshop on European Reference Grids Ispra, 27-29 October 2003: JRC ESDI Action 2142 European Spatial Data Infrastructure*. Ispra. URL: http://www.eionet.europa.eu/gis/docs/EuroGRID_Final_Short_Proceedings.pdf [Stand 13.07.2009]
- Kaiser, Ulrich & Wagner, Joachim (2007). *Neue Möglichkeiten zur Nutzung geheimer Mikrodaten aus amtlichen Erhebungen: FDZ- Arbeitspapier Nr.20*. URL: http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz_arbeitspapier-20.pdf [Stand 13.07.2009]
- Kaminger, Ingrid & Wonka, Erich (2005). Regionalstatistische Gebietsgliederungen nach geographischen Rastern. *Statistische Nachrichten / Statistik Austria* (Heft 8), S. 753 - 760.

- Kaminger, Ingrid & Meyer, Werner (2007). Neue Raster-orientierte Statistik Europa, in Strobl, Josef; Blaschke, Thomas & Griesebner, Gerald (Hrsg.): *Angewandte Geoinformatik 2007: Beiträge zum 19. AGIT-Symposium Salzburg [vom 3. bis 6. Juli 2007]*. Heidelberg: Wichmann, S. 303 - 308.
- Karg, Moritz (2008). *Datenschutzrechtliche Rahmenbedingungen für die Bereitstellung von Geodaten für die Wirtschaft: Gutachten im Auftrag der GIW-Kommission*. Kiel.
URL:
<http://www.geobusiness.org/Geobusiness/Redaktion/PDF/Publikationen/ampelstudie-datenschutzrechtliche-rahmenbedingungen-bereitstellung-geodaten-lang.property=pdf,bereich=geobusiness,sprache=de,rwb=true.pdf>
[Stand 13.07.2009]
- Karusseit, Oliver & Simon, Marcel (2008). Statlas.NRW: Interaktive Karten statt Tabellen. *LDVZ-Nachrichten* (Heft 1/2008_8), S. 21-23.
- Knabenschuh, Martin & Westenberg, Gerfried (2009). Hauskoordinaten und Hausumringe: Angebot des Amtlichen deutschen Vermessungswesens etabliert. *fub - Flächenmanagement und Bodenordnung* 71.Jahrgang (Heft 3), S. 133 - 141.
- Krügel, Tina (2008). Datenschutz und Geomarketing - Status Quo. *Geomarketing - Magazin für Markt und Raum* Jahrgang 9 (Heft 4), S. 6-7.
- Lange, Norbert de (2002). *Geoinformatik in Theorie und Praxis: Mit 59 Tabellen*. Berlin: Springer
- Lange, Norbert de (2003). Einsatz von Geoinformationssystemen zur Entscheidungsunterstützung in der Stadtplanung: Modellierung von Rasterdaten und Oberflächen. *Münstersche Geographische Arbeiten* (Heft 46), S. 221 - 230.
- Longley, Paul A.; Goodchild, M. F.; Maguire, D. J. & Rhind, D. W. (2001). *Geographic Information Systems and Science: [GIS]*. Chichester: John Wiley & Sons, Ltd
- Lux, Felix (2007). Inspire INSPIRE! Beseelt INSPIRE!, in Strobl, Josef, Blaschke, Thomas & Griesebner, Gerald (Hrsg.): *Angewandte Geoinformatik 2007: Beiträge zum 19. AGIT-Symposium Salzburg [vom 3. bis 6. Juli 2007]*. Heidelberg: Wichmann, S. 457-462.
- Meinel, Gotthard & Siedentop, Stefan (2007). Erhebung und indikatorgestützte Bewertung der Siedlungsstruktur und ihrer Entwicklung - Konzept "Deutschlandmonitor Siedlungs- und Freiraumentwicklung", in Strobl, Josef, Blaschke, Thomas & Griesebner, Gerald (Hrsg.): *Angewandte Geoinformatik 2007: Beiträge zum 19. AGIT-Symposium Salzburg [vom 3. bis 6. Juli 2007]*. Heidelberg: Wichmann, S. 473-481.

- Meyer, Werner (2007). *Rasterkaren mit variablen Zellgrößen (v. 01)*. 6. VoGIS Fachtagung. Feldkirch. URL:
http://www.vorarlberg.at/vorarlberg/bauen_wohnen/bauen/landesvermessungsamt/weitereinformationen/fachforumvogis/fachforum2007.htm [Stand 13.07.2009]
- Müllegger, Christian (2007). *Grundlagen der Datenqualität (ISO 19113)*. URL:
http://homepage.univie.ac.at/wolfgang.kainz/Lehrveranstaltungen/Seminar/2006%20WS/Muellegger_Text.pdf [Stand 27.07.2009]
- OGC Open Geospatial Consortium Network. URL:<http://www.ogcnetwork.net/wcps>
[Stand 15.07.2009]
- Prinz, Thomas; Dollinger, Franz & Herbst, Stefan (2007). EuRegionale Raumindikatoren für die grenzübergreifende Infrastrukturplanung, in Strobl, Josef, Blaschke, Thomas & Griesebner, Gerald (Hrsg.): *Angewandte Geoinformatik 2007: Beiträge zum 19. AGIT-Symposium Salzburg [vom 3. bis 6. Juli 2007]*. Heidelberg: Wichmann, S. 583 - 588.
- Prinz, Thomas; Strobl, Josef & Wonka, Erich 2004. *Flexible Aggregation regionalstatistischer Erhebungen- Neue Produkte der Statistik Austria*. URL:
http://ispace.researchstudio.at/downloads/2004/prinz_strobl_wonka_2004_agit_regionalstatistik.pdf [Stand 13.07.2009]
- Radmacher-Nottelmann, Nils 2005. Geheimhaltung mit Makrodaten - Das Beispiel der Beherbergungsstatistik: *Statistische Analysen und Studien NRW. (Bd. 19)*, S. 29 - 37.
- Rech, Monika (2009). Interview WCPS - Neue Sprache für Geo-Rasterdaten, in Dirk Schmidbauer (Hrsg.): *GIS Business - Das Magazin für Geoinformation*. Heidelberg: abc Verlag. (Bd. 4), S. 50 - 51.
- Schwedler, Hanns-Uve (2008). *Silent City Leisere Kommunen: Informationen zur Umgebungslärmrichtlinie*. Berlin. URL:
<http://www.umweltdaten.de/publikationen/fpdf-l/3686.pdf> [Stand 13.07.2009]
- Steinnocher, Klaus; Petrini, Frederic; Tötzer, Tanja & Weichselbaum, Jürgen (2005). Räumliche Disaggregation von sozio-ökonomischen Daten, in Strobl, Josef & Zum Blaschke, Thomas &. (Hrsg.): *Angewandte geographische Informationsverarbeitung XVI: Beiträge AGIT-Symposium*. Heidelberg: Wichmann
- Steinnocher, Klaus; Weichselbaum, Jürgen & Köstl, Mario (2006). *Linking remote sensing and demographic analysis in urbanised areas. 1st EARSeL Workshop of the SIG Urban Remote Sensing, Humboldt-Universität zu Berlin*. URL:
http://www.earsel.org/workshops/SIG-URS-2006/PDF/Session1_steinnocher.pdf
[Stand 13.07.2009]

- Strobl, Josef (2005). Hierarchische Aggregation: Detailinformation versus Datenschutz am Beispiel adressbezogener georeferenzierter Datensätze. *Salzburger Geographische Arbeiten* (Heft 38), S. 163 - 171.
- Strobl, Josef; Wonka, Erich & Prinz, Thomas (2004). Flexible Aggregation regionalstatistischer Erhebungen - Neue Produkte der Statistik Austria, in Strobl, Josef, Blaschke, Thomas & Griesebner, Gerald (Hrsg.): *Angewandte Geographische Informationsverarbeitung XVI: Beiträge zum AGIT-Symposium*. Heidelberg: Wichmann, S. 556 - 561.
- Szibalski, Martin (2006). Karten in der amtlichen Statistik, in Statistisches Bundesamt, Wiesbaden (Hrsg.): *Auszug aus Wirtschaft und Statistik*, S. 205 - 211.
- Szibalski, Martin (2007). Kleinräumige Bevölkerungs- und Wirtschaftsdaten in der amtlichen Statistik Europas, in Statistisches Bundesamt, Wiesbaden (Hrsg.): *Auszug aus Wirtschaft und Statistik*, S. 137 - 143.
- Tammilehto-Luode, Marja; Backer, Lars & Rogstad, Lars (2000). Grid data and area delimitation by definition towards a better European territorial statistical system. *Statistical Journal of the United Nations Economic Commission for Europe* (Heft 17), S. 109 - 117.
- Tomlin, C. D. (1991). Cartographic Modelling, in Maguire, D. J., Goodchild, M. F. & Rhind, D. W. (Hrsg.): *Geographical Information Systems: Principles and Applications*, S. 361 - 374.
- Umweltinformationsgesetz vom 22. Dezember 2004 (BGBl. I S. 3704): UIG_Online im Internet: URL: http://www.gesetze-im-internet.de/bundesrecht/uig_2005/gesamt.pdf [Stand 13.07.2009]
- Vorläufige Berechnungsmethode zur Ermittlung der Belastetenzahlen durch Umgebungslärm: VBEB: 9.Februar 2007. *Bundesanzeiger*(Nr.75).URL: <http://www.umweltbundesamt.de/laermprobleme/ulr.html> [Stand 14.07.2009]
- Weichert, Thilo. (2007) Der Personenbezug von Geodaten: *DuD - Datenschutz und Datensicherheit*. Wiesbaden: Vieweg Verlag / GWV Fachverlage GmbH, S. 17 - 23.
- Wonka, Erich (1997). *Die Grosszählungsdaten auf der Basis von Gebäudekoordinaten als Datenquelle für die örtliche Raumplanung*. Wien: Österr. Statistisches Zentralamt. (Schriftenreihe zur Kartographie und Geoinformation, Bd.5).
- Wonka, Erich (2008). *Regionalstatistik in Österreich auf der räumlichen Bezugsbasis von regionalstatistischen Rastereinheiten*. Salzburg und Wien. URL: <http://www.oeaw-giscience.org/download/RasterStatistikWonka.pdf> [Stand 13.07.2009]