

### 3 Modellierung von Geodaten

Um Informationen der realen Welt rechnergestützt verarbeiten und speichern zu können, ist es erforderlich, diese Informationen in einem *Modell* abzubilden. Dabei muss die Komplexität der realen Welt geeignet vereinfacht und verallgemeinert werden. Es gibt eine Reihe von Gründen für eine solche *Abstraktion*:

- Rechnergestützte Informationssysteme unterliegen Beschränkungen hinsichtlich des verfügbaren Speicherplatzes und der möglichen Verarbeitungsgeschwindigkeit, sodass ein Modell hinsichtlich seiner Komplexität eingeschränkt werden muss, um überhaupt von einem Rechner verwendet werden zu können.
- Die Einfachheit der Handhabung und die Überschaubarkeit von Daten nehmen mit zunehmender Komplexität des Modells ab. So ist zum Beispiel eine  $n$ -zu- $m$ -Beziehung mächtiger als eine  $1$ -zu- $n$ -Beziehung, aber auch aufwendiger zu erstellen und zu pflegen.
- Bei der Zielsetzung, die Wirklichkeit exakt zu repräsentieren, stellt sich die Frage, ab welchem Modellierungsgrad man tatsächlich exakt ist. Dabei treten üblicherweise Definitionsprobleme auf. Wie kann ein Küstenverlauf und damit zum Beispiel ein Längenmaß für Küsten definiert werden? Ein anderes gerne betrachtetes Beispiel ist die Länge der Grenze zwischen Spanien und Portugal. In dem statistischen Jahrbuch von Portugal soll (angeblich) ein größerer Wert angegeben sein als im spanischen Werk. In Portugal hat man also offenkundig ein anderes, vermutlich genaueres Modell verwendet als im flächenmäßig wesentlich größeren Spanien.
- Ein anderer Aspekt der Modellbildung betrifft deren Zielsetzung. Ein mit Text bedrucktes Blatt Papier kann durch eine Größenangabe des Blattes sowie einer zeichenorientierten Codierung des gedruckten Textes modelliert werden. Eine andere Repräsentation dieses Blattes erhält man, wenn man das Blatt wie bei einem Fax einscannet. Optisch spiegeln die gescannten Daten das Blatt sicher besser wider als das zeichenorientierte Modell. Eine inhaltliche Analyse des Textes ist allerdings in dieser Repräsentation ungleich schwieriger als in der ersten Variante. Aus chemischer Sicht sind die beiden genannten Modelle ungeeignet. Hier müsste man die Molekülstruktur des Papiers und der Farbe beschreiben. Aus Sicht der Physik könnte man sich auf die Ebene der Atome, auf die Ebene der Elementarteilchen oder auf die Ebene von Quarks zur Beschreibung von Elementarteilchen begeben. Die inhaltliche Bedeutung des Textes wird man durch eine solche Modellierung wohl kaum ableiten können.

Ein exaktes Abbild der realen Welt ist offenkundig sowohl ein unerreichbares als auch ein nicht erstrebenswertes Ziel. Daher muss man sich immer im Klaren sein, dass die in einer Datenbank gespeicherten Daten ein vereinfachtes *Datenmodell* der realen Welt darstellen. Durch die Modellbildung gehen – ebenso wie bei der Erfassung der Daten – Informationen verloren bzw. werden verfälscht.

Für die Speicherung und Verwaltung von Geodaten in einem Geoinformationssystem benötigen wir also ein geeignetes Datenmodell. Welche Eigenschaften von Geodaten in einem solchen Modell berücksichtigt werden sollten, ist das Thema des nächsten Abschnitts. Nachdem Abschnitt 3.2 die Standardisierung von Geodaten behandelt, wird im dritten Abschnitt die ISO-Norm 19107 „Spatial Schema“ vorgestellt. Dieses Feature-Geometry-Modell ist Basis für konkrete Datenbankschemata, wie das Simple-Feature-Modell und SQL/MM/

Spatial, die in den nachfolgenden Abschnitten eingeführt werden. Die Behandlung von räumlichen und linearen Bezugssystemen bildet den Abschluss dieses Kapitels.

## 3.1 Eigenschaften von Geodaten

Die zu modellierenden Eigenschaften von Geodaten lassen sich in vier Kategorien einteilen [97]:

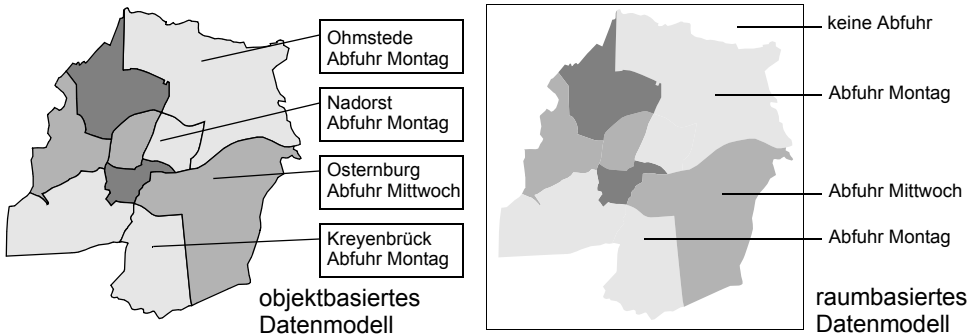
- *Geometrische Eigenschaften* beschreiben die Lage und Ausdehnung von Objekten im Raum. Über *Koordinaten* kann zum Beispiel die *Lage* eines Punktes (z.B. des Ortsmittelpunktes) oder Lage und *Ausdehnung* (bzw. Form) eines Gebietes (z.B. des Gemeindegebietes) repräsentiert werden.
- *Topologische Eigenschaften* dienen der Beschreibung der relativen räumlichen Beziehungen von Objekten zueinander, wobei von der Geometrie abstrahiert wird. Typische topologische Beziehungen betreffen die Nachbarschaft oder die Überschneidung. Die Topologie einer Karte wird durch Transformationen wie das Verschieben, Drehen oder Skalieren aller Kartenelemente bezüglich eines Transformationspunktes nicht verändert. Für die Veranschaulichung dieses Sachverhaltes wird gerne eine auf einem Luftballon aufgetragene Karte herangezogen. Durch Aufpumpen oder Ablassen von Luft verändert sich die Topologie dieser Karte nicht.
- *Thematische Eigenschaften* entsprechen Sachattributen. *Nominale Eigenschaften* sind beispielsweise Bezeichnungen wie Ortsnamen oder Postleitzahlen. *Qualitative Eigenschaften* können der Status einer Gemeinde oder der Wochentag der Müllabfuhr sein. *Quantitative Eigenschaften* wären die Niederschlagsmenge oder die Einwohnerzahl der Gemeinde.
- *Temporale Eigenschaften* beschreiben den Zeitpunkt oder den Zeitraum, für den die übrigen Eigenschaften gelten. Liegen die Geometrie und die übrigen Eigenschaften eines Objektes für mehrere (aufeinanderfolgende) Zeitpunkte bzw. -räume vor, so kann die *Dynamik* des Objektes beschrieben werden. Verändert sich über die Zeit im Wesentlichen nur die Lage, aber nicht die Form der Geometrie, so spricht man auch von sich *bewegenden Geoobjekten*. Die Unterstützung von *spatio-temporalen Datenmodellen* durch ein GIS oder Geodatenbanksystem ist derzeit noch Thema der Forschung. Daher wird im Folgenden zunächst nicht auf die temporalen Eigenschaften von Geodaten eingegangen; diese sind Thema von Abschnitt 14.2 im letzten Kapitel dieses Buchs.

### 3.1.1 Thematische Eigenschaften

Die thematischen Eigenschaften lassen sich über die normalen Mechanismen eines relationalen Datenbanksystems beschreiben. So können qualitative Eigenschaften durch alphanumerische Attribute und quantitative Eigenschaften durch numerische Attribute repräsentiert werden.

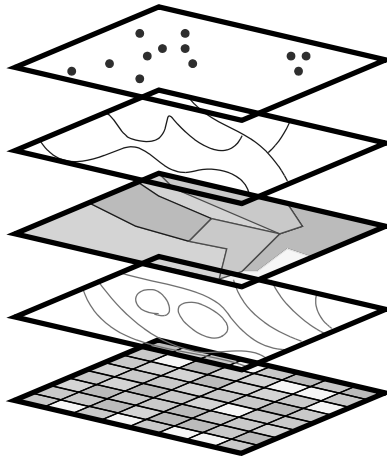
Thematische Eigenschaften von Geodaten müssen geeignet dem (zwei- oder mehrdimensionalen) Raum zugeordnet werden. Dabei lassen sich zwei wesentliche Prinzipien voneinander unterscheiden. Bei *objektbasierten Datenmodellen* ist der Ausgangspunkt der Betrachtung das *Geoobjekt* (engl. *Feature*), das ein ausgezeichnetes Geometrieattribut aufweist. Die übr-

gen Attributwerte lassen sich dieser Geometrie zuordnen. Besitzt eine Gemeinde zum Beispiel die thematischen Attribute „Gemeindestatus“ und „Einwohnerzahl“ sowie das geometrische Attribut „Gemeindegebiet“, so gelten die genannten thematischen Attributwerte für das Gemeindegebiet. Neben dem ausgezeichneten Geometrieattribut kann ein Geoobjekt durchaus weitere Geometrieattribute aufweisen; eine Gemeinde kann z.B. neben der Gemeindefläche als ausgezeichnetem Geometrieattribut zusätzlich das Gemeindezentrum als punktförmiges Attribut besitzen.



**Abb. 3.1:** Beispiel für ein objekt- und ein raumbasiertes Datenmodell

Bei *raumbasierten Datenmodellen* ist der Datenraum der Ausgangspunkt der Betrachtung. Mit jedem Punkt im Datenraum ist ein Attributwert verknüpft. Anders formuliert (unter Annahme eines zweidimensionalen Datenraums): Es existiert eine kontinuierliche Funktion  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ . Ein raumbasiertes Datenmodell wird auch als *Coverage* bezeichnet. Abbildung 3.1 illustriert an einem Beispiel das objektbasierte und das raumbasierte Modell.



**Abb. 3.2:** Beschreibung unterschiedlicher Thematiken in mehreren Ebenen

Zu einem Geoobjekt kann es mehrere thematische Attribute geben. Dann bilden alle Geoobjekte zu jeweils einem dieser Attribute eine *thematische Ebene* (engl. *Layer*). Gleiches gilt, wenn bei einem raumbasierten Datenmodell mehrere verschiedene Funktionen vorliegen.

Dann beschreibt jede dieser Funktionen eine thematische Ebene. So kann es zum Beispiel die thematischen Eigenschaften „Niederschlagsmenge“ und „Müllabfuhrtermine“ geben, die zwei verschiedene Layer bilden. In Abbildung 3.2 ist ein Modell dargestellt, das mehrere thematische Ebenen umfasst. Geoinformationssysteme erlauben typischerweise das Ein- und Ausblenden einzelner Layer sowie das Verschneiden von mehreren Ebenen zu einem neuen Layer.

### 3.1.2 Geometrische Eigenschaften

Bei der Repräsentation von geometrischen Eigenschaften lassen sich zwei grundsätzliche Modellformen voneinander unterscheiden: das *Vektormodell* und das *Rastermodell*. Das Rastermodell zerlegt den Datenraum in gleichförmige Teilflächen. Typischerweise erfolgt die Aufteilung in quadratische oder rechteckige *Rasterzellen*, die in der Bildverarbeitung auch *Pixel* genannt werden. Die thematischen Eigenschaften eines Gebietes sind für eine Rasterzelle einheitlich; dadurch kann es offenkundig je nach Größe der Rasterzellen zu einem Verlust an Genauigkeit kommen. Vektormodelle bauen auf *Punkten* und *Linien* auf. Die Lage der Punkte wird über Koordinaten bezüglich eines Koordinatensystems beschrieben. Auf dieser Basis werden komplexere Einheiten wie *Streckenzüge*, *Polygone* (ohne und mit Löchern) oder *Multipolygone* zur Beschreibung von Linien und Flächen gebildet. Über Linien werden die thematischen Eigenschaften unterschiedlicher Gebiete scharf voneinander getrennt, was, insbesondere bei kontinuierlichen Übergängen, z.B. zwischen Gebieten mit hoher und niedriger Luftverschmutzung oder bei seichten Küsten, zu einer nicht sachgerechten Modellierung führen kann. Abbildung 3.3 zeigt ein Beispiel, in dem eine Fläche durch ein Vektor- und durch ein Rastermodell repräsentiert wird.

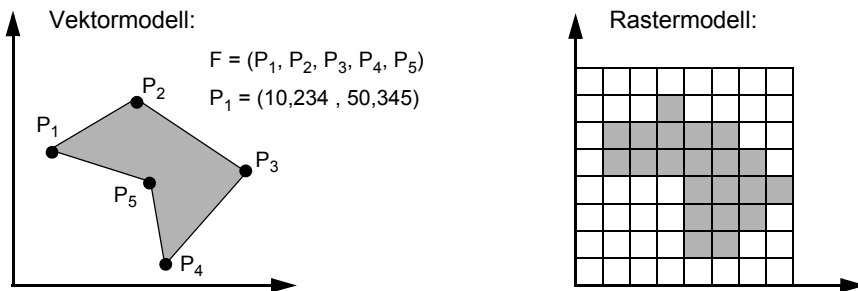


Abb. 3.3: Vektor- und Rastermodell

Raster- und Vektormodelle weisen eine Reihe von Vor- und Nachteilen für die Modellierung von Geodaten auf. Gegenüberstellungen finden sich z.B. in [7] und [97]. Wenn wir diese beiden Modellierungsformen mit der objekt- und raumbasierten Modellierung verbinden, ergeben sich vier Kombinationsmöglichkeiten, die in der Tabelle 3.1 dargestellt sind.

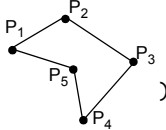
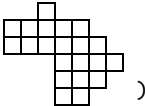
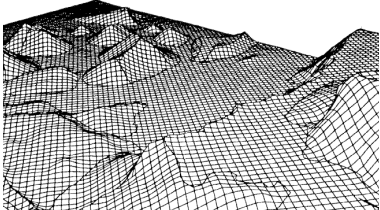
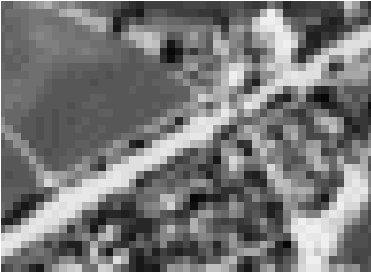
	<i>Vektormodell</i>	<i>Rastermodell</i>
<i>Objektbasiertes Datenmodell</i>	<p>Über ein geometrisches Attribut eines Geoobjektes wird eine Vektorgeometrie, zum Beispiel ein Polygon oder eine Linie, gespeichert. Die Kombination eines objektbasierten Vektormodells ist sehr typisch für die Modellierung von Geoobjekten in Geodatenbanksystemen.</p> <pre>gemeinde1 = ('Oldenburg', 'Stadt',</pre>  <pre>)</pre>	<p>Über ein geometrisches Attribut eines Geoobjektes wird eine Rastergeometrie gespeichert. Die Verwendung von objektbasierten Rastermodellen ist eher unüblich. Eine Beschreibung eines solchen Ansatzes findet sich z.B. in [195].</p> <pre>gemeinde1 = ('Oldenburg', 'Stadt',</pre>  <pre>)</pre>
<i>Raumbasiertes Datenmodell</i>	<p>Die räumliche Funktion wird über eine vektorielle Beschreibung einer Ebene im dreidimensionalen Raum repräsentiert. Dies erfolgt zum Beispiel bei der Modellierung von Geländeformen.</p> 	<p>Rasterbilder, die aus Satellitenbildern oder anderen Luftaufnahmen gewonnen wurden, stellen ein raumbasiertes Rastermodell dar.</p> 

Tabelle 3.1: Kombinationsmöglichkeiten der Modelle

### 3.1.3 Topologische Eigenschaften

Die Topologie kann entweder durch Angaben entsprechender Beziehungen explizit in einem *topologischen Datenmodell* repräsentiert werden oder implizit aus der Geometrie abgeleitet werden. So kann zwischen allen Flächen, die aneinander stoßen, eine entsprechende n-zu-m-Beziehung definiert werden. Das Problem der expliziten Modellierung ist, dass bei Änderungen der Geometrie auch alle topologischen Beziehungen angepasst werden müssen. Dies kann recht aufwendig sein. Bei der impliziten Modellierung wird eine topologische Eigenschaft bei Bedarf aus der Geometrie abgeleitet. So wird z.B. anhand von Koordinaten bestimmt, welche Flächen eine Überlappung aufweisen. Dies ist die (bislang) übliche Vorgehensweise in Geodatenbanksystemen. Allerdings hat dieser Ansatz auch Nachteile: Zum einen ist ein solches Vorgehen recht rechenaufwendig und zum anderen kann es aufgrund fehlerhafter oder ungenauer Geometriedaten zu falschen Schlussfolgerungen hinsichtlich der Topologie führen.