

Paris Lodron-Universität Salzburg

Universitätslehrgang

Geographical Information Science & Systems

(UNIGIS MSc Jahrgang 2002)

**Entwicklung und Implementierung
eines Datenmodells für geologische
Flächendaten des Niedersächsischen
Landesamtes für Bodenforschung**

von Sven Brinkmann

31.03.2004

Ich versichere, dass ich diese Master Thesis ohne fremde Hilfe und ohne Verwendung anderer als der angeführten Quellen angefertigt habe 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.

Datum, Unterschrift

Danksagung

Ich möchte mich für die Übernahme des Referats, für konstruktive Kritik und Vorschläge bei Herrn Professor Dr. J. Strobl bedanken.

Mein Dank gilt insbesondere Herrn Dr. H. Bombien, Frau Dipl.-Ing. S. Dieler, Herrn Dipl.-Ing. P. Köhler, Frau Dipl.-Ing. U. Knorr, Herrn Dipl.-Ing. R. Reimann und Frau Dipl.-Ing. Agr. E. Schneider vom Referat *Fachinformationssystem Flächen- und Raumdaten* des Niedersächsischen Landesamtes für Bodenforschung. Ohne ihre stetige Diskussions- und Hilfsbereitschaft wäre diese Arbeit nicht möglich gewesen.

Für die Bereitstellung von Ressourcen und für organisatorische Hilfe sei dem Referat *Kohlenwasserstoffgeologie* des Niedersächsischen Landesamtes für Bodenforschung unter Leitung von Herrn Dipl.-Ing. R. Sedlacek gedankt.

Weiterhin möchte ich mich für Hilfeleistungen bei unterschiedlichen Arbeitsschritten bei Herrn Dipl.-Phys. H. Köhncke, Herrn Dr. J. Sbresny und Herrn Dr. H. Brauner bedanken.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Zielsetzung	2
1.2	Wissenschaftlicher Kenntnisstand	3
1.3	Vorgehensweise.....	6
2	Vor- und Nachteile der Speicherung von Geometriedaten in Datenbankmanagementsystem	7
3	Analyse der Ausgangssituation (Definitionsphase)	12
3.1	Ein Überblick über Funktionen und Inhalte geologischer Karten.....	12
3.2	Die Geologische Karte 1:25.000 des Niedersächsischen Landesamtes für Bodenforschung	13
3.2.1	Allgemeines.....	13
3.2.2	Das Datenmodell der GK 25.....	15
3.2.2.1	Exkurs: Das Coverage-Format von Arc/Info	15
3.2.2.2	Geometriedaten	18
3.2.2.3	Attributdaten.....	20
3.2.3	Einbindung der Flächendaten in die Aufgaben und Strukturen des Niedersächsischen Landesamtes für Bodenforschung.....	21
3.3	Die Datenmodelle der weiteren Kartenserien des Niedersächsischen Landesamtes für Bodenforschung	22
3.4	Fazit: Anforderungen an das Datenmodell und die Datenmigration	23
4.	Konzeption und Implementierung des Datenmodells für die Geologische Karte 1:25.000	26
4.1	Exkurs: Objektorientierung und die <i>Universal Modeling Language</i>	27
4.2	Konzeptuelles Datenmodell.....	29
4.2.1	Geologische Entitäten	29
4.2.2	Topologie	33
4.2.3	Kartenrandobjekte.....	40
4.3	Logisches und physisches Modell	41
4.3.1	Exkurs: Die Geodatabase von ArcGis 8.3	43
4.3.2	Geometriedaten	46
4.3.3	Topologie	51
4.3.4	Attributdaten ohne Raumbezug	53
4.4	Implementierung des Datenmodells und Datenmigration	54

4.4.1 Implementierung des Datenmodells mit Hilfe von CASE-Tools	54
4.4.2 Datenmigration	55
4.4.2.1 Geometriedaten	55
4.4.2.2 Attributdaten ohne Raumbezug	56
4.4.3 Blattrandproblematik	56
5 Diskussion und Schlussfolgerungen	64
5.1 Schlussfolgerungen für die Geologische Karte 1:25.000	64
5.1.1 Diskussion der Arbeitsergebnisse	64
5.1.2 Weiterführende Arbeitsschritte	68
5.3 Allgemeine Schlussfolgerungen	70
6 Zusammenfassung	71
7 Summary	73
8 Literaturverzeichnis	76

Abbildungsverzeichnis

Abb. 3.1: Schematische Darstellung des Arbeitsprogramms.....	7
Abb. 3.2: Geometrische Primitive von Arc/Info (nach RIGAUX, SCHOLL & VOISARD 2002, verändert).....	15
Abb. 3.3: Geometrische Dateien von Arc/Info (nach RIGAUX, SCHOLL & VOISARD 2002, verändert).....	16
Abb. 3.4: PAT von Arc/Info am Beispiel eines Coverages der GK 25.	19
Abb. 3.5: AAT von Arc/Info am Beispiel eines Coverages der GK 25.....	19
Abb. 4.1: Schema eines UML-Klassendiagramms und UML-Notation.....	28
Abb. 4.2: Topologische Beziehungen zwischen zwei Polygonen im zweidimensionalen Raum (nach EGENHOFER 1994, verändert).	34
Abb. 4.3: Graphische Repräsentation von topologischen Beziehungen im UML-Klassendiagramm (Erläuterung im Text).....	38
Abb. 4.4: Blick auf eine Geodatabase in ArcCatalog.	43
Abb 4.5: Kodierung der Attribute der Feature Class Stoerung im Attribut Code. Das Attribut Code kann zur Visualisierung der Features verwendet werden (vereinfachtes Beispiel; Erläuterung im Text).....	50
Abb. 4.6: Must-not-have-gaps Fehler nach der Validierung der Topologie.	57
Abb. 4.7: Must-not-overlap Fehler nach der Validierung der Topologie. ...	57
Abb. 4.8: Must-not-have-gaps (grün) und Must-not-overlap (rot) Fehler am ehemaligen Blattrand zwischen zwei Coverages. (Anm: Die Darstellung der Topologie-Fehler korrespondiert bei dieser Zoom-Stufe nicht mit den tatsächlichen Abmessungen der Fehler).....	58
Abb. 4.9: Durch Validierung mit zu großer Cluster Toleranz entstandene Lücken in der Feature Class Kartiereinheit (Gebiet liegt nicht an einem ehemaligen Blattrand).	61
Abb. 4.10: Schema zur Verwendung des ArcGIS-Werkzeugs Zipper Task zur Korrektur der Topologie-Fehler der Feature Class Kartiereinheit.	62
Abb. 4.11: Topologie-Fehler der Feature Class Kartiereinheit nach der Behandlung mit dem Zipper-Task-Werkzeug und einer Zipper-Toleranz von 1,0 m (Reduktion der Fehleranzahl um ca. 60 %; vgl Abb. 4.6 und 4.7).....	63

Abb. 4.12: Schematische Darstellung eines geologischen Linienobjekts der Feature Class Stoerung (Erläuterung im Text).....	65
Abb. 4.13: Schematische Darstellung von Attribut-bezogenen Fehlern am ehemaligen Blattrand der Coverages (Erläuterung im Text).	66

Tabellenverzeichnis

Tab. 4.1: Geologische Entitäten im konzeptuellen Modell der GK 25 (vgl. Anhang B1).....	30
Tab. 4.2: Im konzeptuellen Modell wegfallende Linientypen des Arc/Info-Linientypenkataloges.....	30
Tab. 4.3: Anzahl topologischer Beziehungen per Gruppierung von geometrischen Primitiven (nach UBEDA & EGENHOFER 1997; Erläuterung im Text).	35
Tab. 4.4: Wichtige topologische Beziehungen im konzeptuellen Datenmodell der GK 25.	39
Tab. 4.5: Verwendete Topologie-Regeln im physischen Datenmodell der GK 25 (Variante 1).	52
Tab. 4.6: Topologie-Fehler nach der Validierung der Topologie.	55
Tab. 4.7: Vergleich der Topologie-Fehler der Feature Class Kartiereinheit bei Verwendung verschiedener Cluster Toleranzen.....	59

Anhang

Anhang A1: Bearbeitungsstand der digitalen GK 25 (Februar 2004)	
Anhang A2: Datenbankstruktur der Attributdaten der GK 25	
Anhang A3: Linientypenkatalog der GK 25	
Anhang A4: Plotausgabe der GK 25 Blatt 4027 Lutter am Barenberge	
Anhang B1: Konzeptuelles Datenmodell der GK 25	
Anhang B2: Physisches Datenmodell der GK 25, Geometriedaten Variante 1	
Anhang B3: Physisches Datenmodell der GK 25, Geometriedaten Variante 2	
Anhang B4: Physisches Datenmodell der GK 25, Fachdaten	
Anhang B5: Repräsentation der Arc/Info-Linientypen im Datenmodell der Geodatabase	

1 Einleitung

Eine wesentliche Aufgabe des Niedersächsischen Landesamtes für Bodenforschung (NLfB) ist die Erhebung und Verwaltung von raumbezogenen, geologischen und bodenkundlichen Daten. Diese Daten müssen sowohl für die fachliche Arbeit der Mitarbeiter als auch zur Produktion veröffentlichter Kartenserien, wie z. B. der Geologischen Karte von Niedersachsen (Maßstäbe 1: 25.000 und 1:50.000), der Bodenkarte (1:25.000) oder der Bodenübersichtskarte (1:50.000) von Niedersachsen, bereitgestellt werden. Zur effizienten Wahrnehmung dieser Aufgaben werden am NLfB seit vielen Jahren Geographische Informationssysteme (GIS) und Datenbankmanagementsysteme (DBMS) eingesetzt.

Die Speicherung der Raumdaten und ihrer Attributinformationen ist am NLfB derzeit als sogenanntes duales oder hybrides System organisiert (vgl. z. B. HEALEY 1991, WORBOYS 1999, RIGAUX, SCHOLL & VOISARD 2002). Dies bedeutet, dass die Raumdaten als Datei-basiertes System gespeichert werden, welches auf dem proprietären Datenformat *Coverage* der Firma *Environmental Systems Research Institute* (ESRI) beruht. Die in einem Coverage abgelegten geologischen Informationen beziehen sich dabei auf die Fläche eines Blattschnittes der Topographischen Karte 1:25.000. Die raumbezogenen Daten sind über Schlüsselfelder logisch, jedoch nicht System-inhärent, mit Attributdaten verknüpft, welche in einem DBMS abgelegt sind. Insbesondere bei großen Datenbeständen birgt die duale Datenspeicherung einige unüberwindbare Probleme in sich (z. B. CHRISMAN 1990, HARDY 2001), welche jedoch durch neuere technische Entwicklungen im Bereich der GIS-Software und der einhergehenden Steigerung von Rechnerleistungen prinzipiell gelöst worden sind (vgl. HEALEY 1991, BATTY & NEWELL 1996, WORBOYS 1999). So ist es seit einigen Jahren mit kommerziellen Softwareprodukten technisch realisierbar, raumbezogene Daten auch in (objekt-) relationalen Datenbankmanagementsystemen zu halten. Systeme, die diesem Ansatz folgen, werden als integrierte Systeme bezeichnet. Dieses Konzept bietet gegenüber dualen Systemen einige wesentliche Vorteile, die dem GIS-Nutzer zu einer Effizienz- und Produktivitätssteigerung verhelfen können. Zu diesen Vorteilen gehören u. a. die Nutzung von DBMS-Funktionalitäten zur Kontrolle der Datenintegrität und -sicherheit, die Vermeidung von Redundanzen durch Multi-User-Zugriff, Zusammenführung von Geometrie- und Attributdaten in einem DBMS, sowie die blattschnittfreie Datenhaltung (vgl. HEALEY 1991, ESRI 2000, ESRI 2003b sowie Kapitel 2).

Um von diesen Vorteilen profitieren zu können, sollen die umfassenden raumbezogenen Datenbestände des NLfB mittelfristig von dem dualen Speichersystem in eine geeignete integrierte Architektur überführt werden. Begründet aus vergangenen Entscheidungen arbeitet das NLfB derzeit mit Softwareprodukten der Firmen ESRI und Microsoft, mit denen das Vorhaben umgesetzt werden soll. Im Einzelnen sind dies auf Seiten der GIS-Software ArcGIS und ArcSDE in den Versionen 8.3 sowie SQL-Server 2000 als DBMS. Für die Speicherung von raumbezogenen Daten in einem relationalen DBMS verwendet ArcGIS 8.x das generische Datenmodell *Geodatabase*, welches grundsätzlich anders konzipiert ist als das derzeitige benutzte Coverage-Modell. Vor der Migration der Daten in das DBMS steht also der Schritt, ein für die Geodatabase geeignetes Datenmodell für die Daten des NLfB zu konzipieren und zu implementieren.

Ein erfolgreiches Datenmodell beinhaltet, dass die Natur der Daten selbst, ihre Repräsentation in der digitalen Umgebung und die Anforderungen des Datennutzers im Umgang mit dem Datenbestand, in Einklang gebracht werden müssen (CONNOLLY & BEGG 2002). Eine Forderung, die eine eingehende Analyse der Daten und ihrer verschiedenen Verwendungen voraussetzt.

Nach der Migration des Datenbestandes in das DBMS zeigt sich in der Praxis oft (vgl. CHRISMAN 1990, HARDY 2001), dass es bei Blattschnitt-orientiert erhobenen und gespeicherten Daten zu Positions- und Attributfehlern im Bereich der ehemaligen Blattränder kommt. Positionsfehler äußern sich durch Überlappungen oder Lücken zwischen den entsprechenden Raumdaten. Um eine topologisch fehlerfreie Datenbasis zu erhalten, muss dieses Problem, für das es keine allgemeingültige Lösung gibt, gelöst werden. Unter Attributfehlern sind Fehler oder Inkonsistenzen zu verstehen, die sich daraus ergeben, dass z. B. benachbarte Flächen, die ehemals durch die Blattschnittgrenze getrennt waren, nun divergierende thematische Informationen aufweisen, die offensichtlich nicht die realen Verhältnisse beschreiben (z. B. eine geologische Kartiereinheit, die schnurgerade am ehemaligen Blattschnitttrand an ein Gewässer grenzt). Diese Art von Fehler ist oft nur durch fachliche Neubewertungen zu beheben.

Es zeigt sich also, dass eine Migration von raumbezogenen Daten in ein DBMS generell ein umfassendes Projekt darstellt, welches für den Datenhalter einen hohen Einarbeitungsaufwand bedeutet und mit hohen Zeit- und Geldkosten verbunden ist.

Anliegen dieser Arbeit ist es, den Weg einer Migration von einem dualen Speichersystem zu einer topologisch fehlerfreien, integrierten Datenhaltung in einem relationalen DBMS zu beleuchten. Dabei liegt der Schwerpunkt auf der Konzeption eines für die Geodatabase von ArcGIS 8.3 geeigneten Datenmodells für geologische Flächendaten des NLfB.

1.1 Zielsetzung

Das Ziel dieser Arbeit ist es, die Migration eines Bestandes von geologischen Flächendaten, welche in einem dualen, Coverage-basierten System gespeichert sind, zu einer integrierten Speicherung in einem relationalen DBMS durchzuführen. Um dieses Ziel zu erreichen, müssen folgende, aufeinander aufbauende Teilschritte erfolgreich abgeschlossen werden:

Zunächst muss ein Datenmodell für das integrierte System konzipiert werden, dessen Software-Architektur aus ArcGIS 8.3 als GIS-Frontend, ArcSDE 8.3 als Middleware und SQL-Server als relationales DBMS besteht. Danach erfolgen die Implementierung des Datenmodells und die Migration des Datenbestandes selbst. Bei diesem Schritt ist davon auszugehen, dass der Datenbestand – zumindest an den ehemaligen Blatträndern – nicht topologisch fehlerfrei in die Datenbank geladen wird. Für diese Problematik soll eine praktikable Lösung gefunden werden, so dass das Ziel der Arbeit bei Vorliegen eines topologisch fehlerfreien Datenbestandes in der Datenbank erreicht ist.

Die Aufgabe soll exemplarisch am Datenbestand der Geologischen Karte 1:25.000 für Niedersachsen (GK 25) erfolgen. Das Datenmodell soll jedoch grundsätzlich auch auf die weiteren Kartenserien des NLFb übertragbar sein. Der Datenbestand der GK 25 wurde ausgewählt, da er von den am NLFb vorliegenden Flächendaten die größte Datenvielfalt und somit die höchste Komplexität des Datenmodells aufweist (vgl. Kapitel 3). Weiterhin erfolgte an diesem Datenbestand bisher noch kein fachlicher oder technischer Abgleich der Blattränder. Es ist also sehr wahrscheinlich, dass mit diesem Datenbestand alle Eigenschaften und möglichen Probleme der weiteren Flächendatensätze des NLFb abgedeckt werden, wodurch die grundsätzliche Übertragbarkeit der Ergebnisse auf diese Datensätze unterstützt wird.

1.2 Wissenschaftlicher Kenntnisstand

Der Kernteil dieser Arbeit beschäftigt sich mit Datenmodellierung für geologische Flächendaten. Datenmodellierung im Allgemeinen ist Gegenstand der Forschung seit es Datenbankmanagementsysteme (DBMS) gibt. Ein breites Anwendungsspektrum und hohe Forschungsintensität führten und führen zu zahlreichen Ansätzen und ausgereiften Konzepten, mit denen der Modellierungsprozess für eine Vielzahl von Daten und Anwendungen robust und erfolgreich durchgeführt werden kann.

Datenmodellierung ist ein Abstraktionsprozess, bei dem Informationen über Objekte der realen Welt diskretisiert werden, um sie in eine computergerechte Struktur zu bringen. Der Natur eines Modells entsprechend, geschieht dies immer vor dem Hintergrund einer bestimmten Sichtweise bzw. eines bestimmten Anwendungsfalles. Vor dem eigentlichen Modellierungsprozess muss eine Phase erfolgen, in der die Anforderungen an die zu erstellende Datenbank bzw. Datenbankanwendung detailliert formuliert werden (z. B. CONNOLLY & BEGG 2002). In dieser Definitionsphase sollten z. B. die strategischen Ziele und Pläne der Datenbank-Nutzer analysiert oder Stärken und Schwächen bestehender Systeme beurteilt werden. Weiterhin ist die Sichtweise des Nutzers auf die Daten (auch externes Modell genannt) zu ermitteln und zu verstehen. Der dann folgende Modellierungsprozess wird üblicherweise in mehrere Schritte eingeteilt. Nach ADAM & GANGOPADHYAY (1997) sind in den meisten Projekten folgende drei Schritte realisiert:

- 1) Konzeptuelles (oder konzeptionelles) Datenmodell: In diesem Schritt wird ein Modell entwickelt, welches entsprechend der Sicht und den Anforderungen des Nutzers die reale Welt abstrahiert. In diesem Schritt wird u. a. festgelegt, welche Objekte, Eigenschaften und Beziehungen der realen Welt für die gegebene Anwendung notwendig sind. Wichtig ist weiterhin, dass dieses Modell unabhängig von der zur Implementierung verwendeten Software-Plattform entwickelt werden sollte. Zur konzeptuellen Modellierung wird häufig das *Entity-Relationship Model* (ER Modell) angewendet (z. B. CHEN 1976). Die grundlegenden Bausteine des ER Modells sind Entitäten, Attribute und Beziehungen. Eine Entität repräsentiert ein Objekt oder Konzept der realen Welt, das aus Sicht des Nutzers von

wesentlicher Bedeutung ist und eine unabhängige Existenz aufweist (CONNOLLY & BEGG 2002). Attribute beschreiben die Eigenschaften einer Entität und Beziehungen stellen bestimmte Assoziationen zwischen Entitäten dar. Verschiedene Autoren (z. B. CONNOLLY, BEGG & STRACHAN 1996 oder ELMASARI & NAVATHE 1994) entwickelten Ansätze, um das ER Model für komplexere Modellierungserfordernisse, wie z. B. übergeordnete/untergeordnete Entitäten, zu erweitern (siehe auch unten).

- 2) Logisches Datenmodell: In diesem Schritt erfolgt die Übersetzung des konzeptuellen Modells in eine Datenstruktur, welche mit der zur Implementierung ausgewählten Software-Plattform konsistent ist. Im vorliegenden Fall handelt es sich hierbei um ein relationales Modell, da ein relationales DBMS verwendet werden soll. Neben dem relationalen Modell, welches hier als bekannt vorausgesetzt wird, existieren hierarchische, Netzwerk- oder Objekt-orientierte Modelle. Über diese Modelle kann zahlreich erschienene Literatur Aufschluss geben, wie z. B. DATE 1986; CODD 1990; WORBOYS, HEARNshaw & MAGUIRE 1990; SAUER 1992; CONNOLLY & BEGG 2002; Kleiner 2003.
- 3) Physisches Datenmodell: Dieses Modell, z. T. auch internes Modell genannt, ist direkt auf die Anforderungen seitens der zur Implementierung ausgewählten Softwareprodukte zugeschnitten. Hier sind u. a. Belange zu beachten, die wichtig für die Performance einer Datenbank(anwendung), für die Datensicherheit und –integrität oder für den Speicherplatzbedarf sind. Für dieses Projekt müssen demnach in diesem Schritt die Möglichkeiten von SQL-Server 2000 und der Geodatabase Beachtung finden.

Ein Werkzeug zur Datenmodellierung, welches sich in den letzten Jahren als ein De-facto-Standard etabliert hat (ERLER 2002, KLEINER 2003), ist die *Unified Modeling Language* (UML). Hierbei handelt es sich um eine Sprache zur Spezifikation, Visualisierung, Konstruktion und Dokumentation von Softwaresystemen. Die UML verwendet dabei das Konzept der Objektorientierung (vgl. Abschnitt 4.1), welches in der Softwareentwicklung durch objektorientierte Programmiersprachen, wie Java oder C++, eine große Bedeutung erlangt hat. Viele Softwareprodukte unterstützen bereits die UML (so auch ArcGIS 8.3), deren Diagramme dann direkt zur Implementierung der Modelle verwendet werden können (vgl. Abschnitt 4.4). UML bleibt jedoch ein Werkzeug zur Modellierung und ersetzt nicht einen sinnvollen Modellierungsprozess, wie er oben beschrieben wurde. Auch zur UML und zum Konzept der Objektorientierung ist zahlreiche Literatur erschienen (z. B. ERLER 2002; RUMBAUGH, JACOBSON & BOOCH 1999; WORBOYS, HEARNshaw & MAGUIRE 1990).

Bei den Daten, die in dieser Arbeit modelliert werden, handelt es sich um raumbezogene Daten. Diese Daten sind in vielen Belangen anders – komplexer - als die Daten, die üblicherweise in DBMS abgelegt werden. Einen interessanten Überblick über die Andersartigkeit von raumbezogenen Daten geben GOODCHILD (1992a) oder BATTY & NEWELL (1996). Diese Andersartigkeit spiegelt sich nicht nur in der Entwicklung von spezieller Software wie z. B. GIS wider, sondern auch durch Ansätze, die zum Ziel haben, die Möglichkeiten zur Modellierung von raumbezogenen Daten zu verbessern. Beispielsweise entwickelten LAURINI & THOMPSON

(1992) eine Erweiterung zum ER Model, GADIA & CHOPRA (1993) sowie HADZILACOS & TRYFONA (1996) beschreiben Erweiterungen des relationalen Datenmodells und MILNE et al. (1993), DAVID et al. (1993) oder WORBOYS (1994) berichten über objektorientierte Datenmodelle. PARENT, SPACCAPIETRA & ZIMÁNYI (1999) stellen jedoch heraus, dass diese Ansätze verschiedene Defizite aufweisen und dass in der Praxis keine Unterstützung durch entsprechende Software-Werkzeuge verfügbar ist. Auch zum Zeitpunkt dieser Arbeit existieren keine konzeptuellen oder logischen Datenmodelle für Raumdaten, die eine allgemeine Akzeptanz erreicht haben.

Über die Modellierung von digitalen geologischen Daten im Speziellen ist in der Literatur nur wenig zu finden. Auch VOISARD (1999) kommt zu diesem Ergebnis und zieht u. a. das Fazit, dass derzeit nur wenige komplette geologische Kartenwerke in digitaler Form vorliegen und dass der Prozess der Digitalisierung bei vielen Karteneigentümern noch nicht abgeschlossen ist. Mitarbeiter des *British Geological Surveys* (BGS) berichten in BAIN & GILES (1997) über die Entwicklung eines konzeptuellen und logischen Modells für britische geologische Flächendaten und in LAXTON & BECKEN (1996) wird die Implementierung in einem dualen System (bestehend aus einem GIS von der Firma Intergraph und einem relationalen DBMS der Firma Oracle) kommentiert. Viele der Schlussfolgerungen, die aus diesen Arbeiten gezogen wurden, sind auch heute noch aktuell (vgl. Abschnitt 5.1). Wie die Entwicklung der digitalen geologischen Karten des BGS nach 1997 weiter verlaufen ist, lässt sich der Literatur oder der Webseite (<http://www.bgs.ac.uk/home.html>, Stand 24.02.04) leider nicht entnehmen. Der *U.S. Geological Survey* (USGS) arbeitet seit längerer Zeit an einem umfangreichen konzeptuellen Datenmodell für geologische Karteninformationen, welches zur Standardisierung von geologischen Karten in den USA beitragen soll (z. B. JOHNSON et al. (1999), SOLLER et al. 2002, NADM 2003; zahlreiche Dokumente unter <http://ncgmp.usgs.gov/ngmdbproject/reports>, Stand 24.02.04). Allerdings kommt der USGS auch im Jahr 2003 noch zu dem Fazit, dass die geologische Datenmodellierung und die Standardisierung von geologischen Informationen keine ausgereifte Disziplin darstellt und dass hier noch viel Arbeit zu leisten ist. Weiterhin bietet ESRI ein *Geoscience Data Model* für ArcGIS 8.x zum Download an, welches eine sehr umfangreiche Tabellenstruktur aufweist, die primär auf die Modellierung geowissenschaftlicher Attributdaten ausgerichtet ist (ESRI 2003d). Strukturelle Aspekte, wie z. B. topologische Beziehungen zwischen geographischen Entitäten, werden in diesem Prototyp zurzeit nicht abgebildet, so dass sich für ihn in dieser Arbeit keine Verwendungsmöglichkeiten eröffnen.

Dieser kurze Überblick über den wissenschaftlichen Kenntnisstand zeigt, dass die Datenmodellierung für vergleichsweise einfach strukturierte Daten, wie sie in vielen Datenbanken von Wirtschaftsunternehmen zu finden sind, als eine viel diskutierte und ausgereifte Disziplin bezeichnet werden kann. Für komplexere, häufig mit Unsicherheiten behaftete und/oder raum- oder zeitbezogene Daten, wie sie z. B. in vielen umweltwissenschaftlichen Fachgebieten vorliegen, ist durchaus noch viel Forschungsbedarf vorhanden.

1.3 Vorgehensweise

Die vorliegende Arbeit richtet sich prinzipiell nach dem in Abschnitt 1.1 beschriebenen Datenmodellierungs-Schema. Es ist jedoch zu beachten, dass in diesem Projekt nicht ein System von Grund auf neu entwickelt wird, sondern dass es sich vielmehr um eine Änderung eines bestehenden Systems handelt. Demzufolge existieren bereits ein konzeptuelles, ein logisches und ein physisches Modell für das bestehende System. Es gilt also zunächst, das bestehende System zu analysieren, um festzustellen wie dieses aufgebaut ist, welche Anforderungen an das neue System gestellt werden und ob Teile des alten Systems weiterhin für das neue verwendet werden können. Insbesondere ist dabei das bestehende konzeptuelle Modell herauszuarbeiten, da es Software-unabhängig ist und da es die zu modellierenden Entitäten, Attribute und Beziehungen festlegt. Auf das bestehende konzeptuelle Modell muss also auch die weitere Modellierung des neuen Systems aufsetzen. Da z. T. das logische und besonders das physische Modell auf die verwendete Software zugeschnitten werden müssen, ist in diesen Bereichen der Hauptteil der Modellierungsarbeit zu leisten. Das Arbeitsprogramm dieser Arbeit ist in Abbildung 3.1 graphisch dargestellt. Die Analyse des bestehenden Systems erfolgt in Kapitel 3, welches mit einem Anforderungskatalog (Abschnitt 3.4) an das neue System endet. Nach der Analyse kann das neue System entworfen und implementiert werden, welches in Kapitel 4 beschrieben wird. Als Werkzeug wird hierfür die UML verwendet (vgl. Abschnitt 1.2 und 4.2). Auch die Aspekte zur Datenmigration finden sich in diesem Kapitel (Abschnitt 4.4). Eine abschließende Betrachtung der Ergebnisse findet sich in Kapitel 5.

Eine Umstellung von einem dualen zu einem integrierten Speichersystem von räumlichen Daten stellt einen tiefgreifenden Einschnitt in die Wertschöpfungskette des GIS-Nutzers dar und ist mit hohem Zeit- und Kostenaufwand verbunden. Es stellt sich also stets die Frage, ob eine solche Umstellung überhaupt sinnvoll ist. Um zur Beantwortung dieser Frage einen Beitrag zu leisten, wird daher in Kapitel 2 zunächst ein Überblick über die Vor- und Nachteile eines integrierten Speichersystems gegeben.

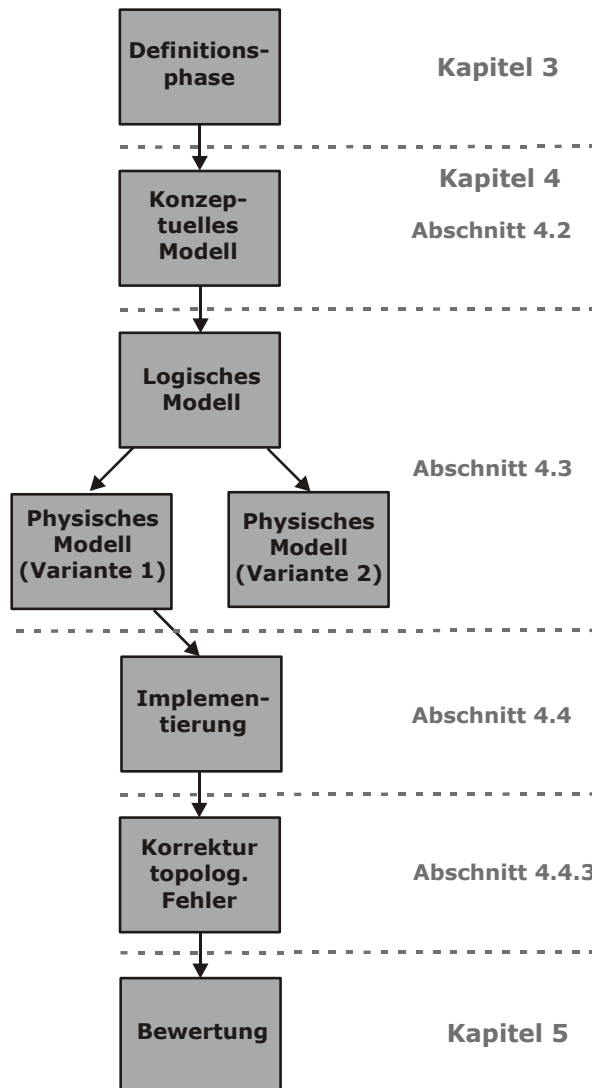


Abb. 3.1: Schematische Darstellung des Arbeitsprogramms.

2 Vor- und Nachteile der Speicherung von Geometriedaten in Datenbankmanagementsystemen

Warum sollte ein bewährtes, funktionierendes System umgestellt werden? Diese oder eine ähnliche Frage werden sich viele GIS-Nutzer von dualen Systemen stellen, wenn sie sich gedanklich mit einer möglichen Umstellung auf ein integriertes System beschäftigen. Da eine Umstellung mit hohem Zeit- und Kostenaufwand verbunden ist, wird der GIS-Nutzer sie nur durchführen, wenn er sich entsprechend große Vorteile erhofft. Die wesentlichen Vorteile der Speicherung von Geometriedaten in einem DBMS (im folgenden auch GeoDBMS genannt) werden im folgenden aufgelistet. Näheres zu den einzelnen Punkten kann der Fachliteratur entnommen

werden (z. B. HEALEY 1991, BATTY 1992, BATTY & NEWELL 1996, ESRI 1998, WORBOYS 1999, RIGAUX, SCHOLL & VOISARD 2002, ESRI 2004a, ESRI 2004b).

Blattschnittfreie Datenbasis

Dass räumliche Datenbestände in vielen Fällen Blattschnitt-orientiert abgelegt werden, ist eine Hinterlassenschaft der manuellen Kartographie. Zu Zeiten der manuellen Kartographie war es schlicht eine Notwendigkeit, Karten in einer handhabbaren Größe herzustellen und weiterzugeben (z. B. GOODCHILD 1992a). Die manuell erstellten Karten wurden und werden in vielen Fällen digitalisiert und bilden somit auch die Grundlage vieler digitaler Datenbestände, in denen sich dann die Blattschnitt-orientierte Speicherung fortgesetzt hat. Ein zwingender Bedarf an Blattschnitt-orientierter Speicherung ergibt sich – insbesondere vor dem Hintergrund immer kostengünstiger werdender Speicherplatzkapazitäten – in der digitalen Welt jedoch nicht, so dass dieses Konzept schon seit längerer Zeit kritisiert wird (z. B. FRANK 1988, CHRISMAN 1990). Einige Nachteile der Blattschnitt-orientierten Speicherung sind offensichtlich: Jeder Kartenbenutzer kennt das Problem, dass das interessierende Gebiet auf mehrere Blattschnitte verteilt ist, was Unannehmlichkeiten oder echte Probleme verursachen kann. In einem GIS kann eine Blattschnitt-orientierte Speicherung bedeuten, dass Arbeitsschritte oder das Verfassen von Programmcode unnötig erschwert werden, wenn z. B. identische Bearbeitungsschritte für mehrere Blattschnitte durchgeführt werden müssen. Andere Nachteile sind weniger offensichtlich, aber unter Umständen schwerwiegender: CHRISMAN (1990) weist darauf hin, dass es auch bei sorgfältig angelegten Blattschnitt-orientierten Datenbasen häufig zu Positionsfehlern kommt, welche sich dann an den Blatträndern durch Überlappungen oder Lücken äußern. Weiterhin belegt CHRISMAN (1990), dass es zu Ungenauigkeiten und Fehlern kommen kann, wenn geometrische Eigenschaften, wie Fläche, Umfang oder Länge, von Objekten innerhalb eines mehrere Blattschnitte umfassenden Gebietes abgefragt werden. Ursache hierfür sind Objekte, welche sich über die Blattschnittgrenzen hinaus fortsetzen, aber durch eben diese Grenzen zerschnitten werden, ohne dass das GIS sie als zusammengehörig behandelt. Dieses Problem kann bei Datenbasen ohne Positionsfehler durch GIS-Funktionen wie z. B. *Merge* oder *Dissolve* (Terminologie nach KENNEDY 2001) gelöst werden. Neben den Positionsfehlern können auch Attribut-bezogene Fehler auftreten, die sich z. B. dadurch äußern, dass die Attribute von Objekten an den Blattschnittgrenzen keinen konsistenten Übergang aufweisen.

Durch die Speicherung von Geometriedaten in einem DBMS ist es möglich, blattschnittfreie Datenbasen anzulegen. Bei idealen blattschnittfreien Datenbasen treten die oben genannten Probleme definitionsgemäß nicht auf. Sie sind damit wesentlich benutzerfreundlicher als eine Blattschnitt-orientierte Datenbasis und ihre Datenmodelle sind einen Schritt weiter in Richtung der realen Welt gerückt. „Ideal“ soll sich in diesem Zusammenhang darauf beziehen, dass nicht nur die geometrischen Eigenschaften der Objekte an den ehemaligen Blatträndern fehlerfrei sind, sondern auch ihre fachlichen Attribute. Ein Zustand, der für viele Datenbestände nur schwer herzustellen ist, was im Verlauf dieser Arbeit noch thematisiert wird (vgl. Abschnitt 5.1).

Gleichzeitiger Zugriff von mehreren Nutzern auf die gesamte Datenbasis

Durch Konzepte wie *Versioning* oder *Check Out*, die das *Long Transaction Management* in einer GeoDBMS gewährleisten, wird ermöglicht, dass mehrere Benutzer gleichzeitig auf die Datenbasis zugreifen bzw. diese bearbeiten können, ohne dass große Teilbereiche der Datenbasis gesperrt sind. Somit ist einerseits sichergestellt, dass die Mitarbeiter stets auf den aktuellen Stand der Datenbasis zugreifen können. Andererseits können durch die zentrale Datenhaltung in einem DBMS auch Redundanzen vermieden werden, die durch lokale Speicherung von Daten auf Rechnern der Mitarbeiter entstehen, was bei dualen Systemen in großen Organisationen häufig zu beobachten ist (HEALEY 2001).

Nutzung von ausgereiften DBMS-Funktionalitäten zur Gewährleistung der Datenintegrität und -sicherheit

Zu den grundlegendsten Aufgaben eines DBMS gehört es, die Integrität und Sicherheit der Daten zu jedem Zeitpunkt zu gewährleisten. Auch bei Systemzusammenbrüchen muss die Datenbasis in einem konsistenten Zustand wieder herstellbar sein. Die DBMS-Branche beschäftigt sich seit langer Zeit mit der Bereitstellung entsprechender Funktionen, so dass diese einen ausgereiften Entwicklungsstand auf einem hohen Niveau aufweisen. Durch die Speicherung von Geometriedaten in einem DBMS können diese Funktionen auch für diese Daten genutzt werden.

Allein die standardisierte Verwaltung der Daten über die Blattschnittgrenzen hinaus stellt schon eine deutliche Verbesserung der Datenkonsistenz dar. So kann in einem Blattschnitt-orientierten System beispielsweise nicht automatisch sichergestellt werden, dass dasselbe Attribut stets mit demselben Datentyp versehen ist.

Integration von Geometrie- und Attributdaten innerhalb eines DBMS

Viele Organisationen verfügen neben Geometriedaten auch über Sachdaten, die in einem mehr oder weniger direkten – vielleicht auch einem bisher unbekanntem – Verhältnis zu den Geometriedaten stehen und die in einem DBMS verwaltet werden. Durch die Integration von Geometrie- und Sachdaten in einem DBMS stehen dem Nutzer effizientere Möglichkeiten zur Verfügung, diese Daten zu verknüpfen, sie abzufragen, zu analysieren und (in Karten) darzustellen.

Vereinfachung der Datenverwaltung

Die DBMS-Speicherung von Geometriedaten kann für die Organisation eine deutliche Vereinfachung der Datenverwaltung darstellen, da Aufgaben, wie z.B. Datenbank-Design, Datenbank-Administration oder auch die Weitergabe von Daten, z. B. über das Internet, effektiver gestaltet werden können. Dies ist insbesondere der Fall, wenn die Möglichkeit besteht, verfügbare Kapazitäten und Know-How aus dem bereits vorhandenen Datenbanksektor (Hard-, und Software sowie Personal) mit zu nutzen. Weiterhin kann das duale System redundante Datenhaltung

erfordern, die durch ein GeoDBMS vermieden werden kann (vgl. Abschnitt 3.2.2). Je nach konkreter verwendeter Software können GeoDBMS einen Schritt in Richtung eines offeneren Systems bedeuten, wodurch es einfacher ist, die Integration mit weiterer Software herzustellen oder Internet-Anwendungen aufzubauen (vgl. ALBRECHT 1999, GOODCHILD 1999, SHETH 1999). (Anmerkung: Auch das Format der Geodatabase von ESRI soll ab der Version 9.0 offengelegt werden (vgl. <http://esri.com/software/arcgis/arcgis90.html>, Stand 24.02.04).

Versionierung

Versionierung (oder *Versioning*, *Version Management*, *Historienverwaltung*) dient nicht nur zur Gewährleistung des *Long Transaction Management* und des Multi-User-Zugriffs in einem GeoDBMS (siehe oben). Das Konzept kann auch dazu genutzt werden, um Szenarien zu entwickeln, ohne dass die Original-Datenbank verändert werden muss. Beispielsweise können an einer versionierten Datenbank ein oder mehrere alternative Datenbank-Schemata getestet werden, ohne dass Auswirkungen auf die Original-Datenbank zu befürchten sind (bzw. ohne dass die Datenbank komplett kopiert werden muss). Bei negativ ausfallenden Tests können die alternativen Versionen dann einfach gelöscht werden.

Erhöhte Benutzerfreundlichkeit

Dieser Vorteil bezieht sich auf den konkreten Fall dieses Projektes, in dem ein Umstieg von Arc/Info (für Workstations) auf ArcGIS 8.3 erfolgen soll. Das GIS Arc/Info ist im wesentlichen als Kommandozeilen-Anwendung konzipiert, d. h. der Benutzer muss Kenntnisse über die Befehlssyntax haben, um seine Aufgaben effizient zu bewältigen. Weiterhin erfolgt die Programmierung von Arc/Info über eine speziell entwickelte Programmiersprache, *Arc Macro Language* (AML) genannt, die vom Programmierer nur für dieses Programm erlernt werden muss. ArcGIS 8.x ist durch seine Windows-Oberfläche und seinen allgemein hohen Grad an Interaktivität als wesentlich benutzerfreundlicher einzustufen als Arc/Info. Die Programmierung von ArcGIS 8.x kann mit jeder zum *Component Object Model* (COM; ein Software-Konzept) von Microsoft kompatiblen Programmiersprache erfolgen (z. B. Visual Basic for Applications, Visual Basic 6, Visual Basic .NET, C++, Delphi u. a.).

Der von ESRI vollzogene Wechsel der Programmiersprachen kann sich für den Nutzer, zumindest kurzfristig, auch als Nachteil erweisen, denn in ArcGIS 8.x sind AML-Skripte (oder auch Avenue Skripte bei Nutzung von ArcView 3.x) nicht lauffähig. Für Organisationen mit umfangreichen AML-Skripten kann dies einen sehr hohen Aufwand an Neuprogrammierung bedeuten.

Aus dieser Liste von Vorteilen ergibt sich, dass die Umstellung auf ein integriertes System für viele GIS-Nutzer Möglichkeiten bietet, durch die eine wesentliche Effizienz- und Produktivitätssteigerung erzielt werden kann.

Die Nachteile eines integrierten Systems lassen sich folgende Punkte nennen:

Hoher Kostenaufwand für die Umstellung

Wenn ein duales System auf ein integriertes System umgestellt werden soll, treten Kosten auf. Je nach konkretem Fall können sich diese aus Hard- und Softwarekosten, Kosten für neues Fachpersonal, Schulungen der Mitarbeiter, Neuprogrammierung von Code oder Auftragsvergabe u. a. zusammensetzen. Auch der Zeitablauf muss vorsichtig geplant werden, damit es nicht zu unproduktiven Phasen während der Umstellung kommt.

Performance

Die Datenmenge, welche in einem GeoDBMS gespeichert wird, kann ein beträchtliches Ausmaß erreichen. Je nach Art und Anzahl der an den Server gerichteten Anfragen kann dieser und der Netzwerkverkehr sehr stark belastet werden. Durch die immer noch rasante technologische Entwicklung von Prozessorleistungen und Speicherkapazitäten sowie durch Weiterentwicklungen der Software, wie z. B. geeignete Indexstrukturen für räumliche Daten, kann heute jedoch eine akzeptable Performance von GeoDBMS erreicht werden. Eine auf die Organisation abgestimmte System-Architektur und das Tuning des DBMS sind für eine optimale Performance sehr wichtig, erfordern aber auch profunde Fachkenntnisse und Erfahrung des System-Designers und -Administrators.

Für einige bestimmte Abfragen bzw. Anwendungen gilt wohl nach wie vor, dass die Flexibilität eines GeoDBMS gegenüber der eines dualen Systems durch Abstriche in der Performance erkauft werden muss. In der gesichteten Literatur konnten jedoch keine konkreten Vergleichsstudien gefunden werden.

Reifegrad der Software

Gegenüber den dualen Systemen, die z. T. schon mehrere Jahrzehnte auf dem Markt sind, handelt es sich bei der Software für integrierte Systeme um vergleichsweise junge Entwicklungen, bei denen mit nicht zufriedenstellend ausgereiften Funktionalitäten und einem weniger umfassenden Funktionalitätsumfang gerechnet werden muss.

Beispielsweise ist die in dieser Arbeit verwendete Software ArcGIS schon einige Jahre auf dem Markt, jedoch ist es erst ab der Version 8.3 (ausgeliefert Mitte 2003) möglich, topologische Beziehungen zwischen den Geometriedaten zu implementieren. Bei den praktischen Arbeiten mit den Topologie-Funktionen von ArcGIS 8.3 trat in dieser Arbeit ein Problem auf, dass durch ungenaue Dokumentation der betreffenden Funktion hervorgerufen wurde und erst durch Kontaktaufnahme zur Herstellerfirma geklärt werden konnte (vgl. Abschnitt 4.3.3).

Die Beantwortung der eingangs gestellten Frage kann nicht allgemeingültig geschehen, da sie stets von den individuellen Gegebenheiten des konkreten Anwendungsfalles abhängig ist. Grundsätzlich kann jedoch festgehalten werden, dass die Vorteile eines integrierten Systems bei einer großen Datenbasis, auf die viele Bearbeiter gleichzeitig zugreifen müssen, am wirkungsvollsten ist.

3 Analyse der Ausgangssituation (Definitionsphase)

3.1 Ein Überblick über Funktionen und Inhalte geologischer Karten

Begründet durch ihre lange Tradition hat die Geologie eine reichhaltige konzeptionelle und symbolische Vielfalt entwickelt, so dass die in geologischen Karten wiedergegebenen Informationen vergleichsweise komplex sind (SOLLER et al. 2002). Die Geologie, insbesondere die geologische Landeserkundung, ist aber auch zum großen Teil eine beschreibende und keine exakte Wissenschaft, was sich auch in den Inhalten geologischer Karten widerspiegelt. Im Vergleich zu vielen anderen Kartenwerken werden auf einer geologischen Karte weniger raumbezogene Fakten dargestellt, als vielmehr die Interpretation von Beobachtungen. Im Mittelpunkt der Beobachtungen steht dabei ein dreidimensionales Phänomen – die Geologie eines bestimmten Gebietes. Der Autor einer geologischen Karte gelangt über einen iterativen Prozess zu einem gedanklichen dreidimensionalen Modell der Geologie des Gebietes, welches dann zweidimensional auf der Karte visualisiert werden muss (RASMUSSEN 1995, BAIN & GILES 1997, VOISARD 1999, JOHNSON et al. 1999). Im Allgemeinen handelt es sich um ein von einem wissenschaftlichen Personenkreis akzeptiertes Modell. Seine Hinterlegung durch Fakten (Bohrungen, Aufschlüsse, Messungen etc.) ist jedoch stark von den geologischen Verhältnissen des konkreten Gebietes abhängig. Die Inhalte geologischer Karten lassen sich in tatsächlich beobachtete Objekte, wie z. B. sichtbare Verwerfungen, Vorkommen von Gesteinen oder Erdfälle, und in interpretierte Objekte, wie z. B. vermutete Störungen oder inter- bzw. extrapolierte Gesteinsvorkommen, einteilen. Es ist prinzipiell denkbar, dass nur wenige neue Fakten das geologische Modell stark verändern würden. Die Inhalte geologischer Karten sind also stets mit einer mehr oder weniger großen Unsicherheit behaftet, die jedoch kaum quantitativ angegeben werden kann. GILES & BAIN (1995) weisen weiterhin darauf hin, dass die genauen Eigenschaften dargestellter Objekte sich z. T. erst aus dem Kontext und/oder durch weitere Hilfsmittel, wie z. B. einem Profilschnitt ergeben. Ob es sich bei einer Störung um eine Auf-, Ab- oder Seitenverschiebung handelt, kann beispielsweise zumeist nicht an der Symbolik selbst erkannt werden, sondern nur bei Betrachtung der angrenzenden Gesteinskörper und eines Profilschnittes. Demnach ist das in einer geologischen Karte dargestellte Wissen z. T. implizit und die sinn-gemäße Vermittlung des Wissens beruht auf stillschweigenden Konventionen zwischen dem

Autor und dem Benutzer der Karte (BRODARIC & GAHEGAN 2002), welche eine entsprechende fachliche Ausbildung voraussetzen.

Neben der bereits angesprochenen Visualisierung von geologischen Informationen, besitzen geologische Karten noch weitere Funktionen, die sich im Laufe der Zeit verändert haben. Zu Zeiten der manuellen Kartographie waren Karten auch zusätzlich die Speicher von geologischen Informationen (GILES & BAIN 1995). Eine Funktion, die heute Datenbanken und/oder GIS übernommen haben. Die digitale Speicherung der Daten eröffnete vielseitige Möglichkeiten zur effizienten Analyse der Daten (statistische Auswertungen, Integration mit weiteren Daten, Kartenverscheidung, Auswertungskarten usw.), so dass Datenanalyse und Visualisierung die zentralen Aufgaben eines modernen geologischen Informationssystems sind.

Die wichtigsten Inhalte geologischer Karten sind in den meisten Fällen die flächendeckend als Polygone dargestellten, geologischen Kartiereinheiten. Sie beschreiben den Ausbiss eines Gesteinskörpers an der Erdoberfläche. Jedes Polygon stellt einen Gesteinskörper dar, welcher vom Autor auf Grund seiner homogenen geologischen Eigenschaften, wie z.B. Gesteinstypus, Schichtfolge, Alter, Genese, Petrographie, Fossilieninhalt usw., von den andersartigen benachbarten Polygonen abgegrenzt wurde. Neben Gesteinskörpern zeigt die Karte auch andere Flächenbeschreibungen, wie z. B. Gewässer oder anthropogen geprägte Flächen. Die Gültigkeit der Karte in Bezug auf die dritte Dimension, die Tiefe, ist meist in Regelwerken (z. B. BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE 2002) festgelegt oder sie ist auch von den geologischen Verhältnissen abhängig. Angaben in Bezug auf die Tiefe beschränken sich oft auf das ordinale Skalenniveau; metrische Angaben sind eher selten in den Karten zu finden.

Die weiteren geologischen Objekte, die auf einer geologischen Karte zu finden sind, werden je nach ihren geometrischen Eigenschaften und in Abhängigkeit vom Maßstab als Linien- oder Punktelemente visualisiert. Zu den typischen Linienelementen zählen Störungen, Flöze, Gesteinsgänge, Terrassenkanten, Umrisse von Tagebauen, Erdfällen usw. Als Punkte können dargestellt sein: wichtige Bohrlöcher, besondere Fundstellen oder Objekte u. a. Neben dem eigentlichen Karteninhalt sind auch Metadaten für eine geologische Karte sehr wichtig. Diese finden sich meist am Kartenrand oder in Erläuterungsheften. Hierzu gehören z. B. Angaben über das Erstellungsjahr der Karte, den Autor, die topographische Kartengrundlage und die verwendeten Kürzel.

3.2 Die Geologische Karte 1:25.000 des Niedersächsischen Landesamtes für Bodenforschung

3.2.1 Allgemeines

Nach BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE (2002) ist die Geologische Karte 1:25.000 (GK 25) die wichtigste Darstellungsart der geologischen Kartier-

ergebnisse. Ihr Blattschnitt basiert auf dem der Topographischen Karte 1:25.000 (TK 25). Für die obersten 2 Meter des dargestellten Gebietes gibt die GK 25 flächendeckend die Verbreitung von geologischen Einheiten mit detaillierten Angaben zur Schichtenabfolge, Petrographie, Genese und zum Alter wieder. Mit den ergänzenden Angaben aus Profilschnitten, Auswertungskarten (z. B. Profiltypenkarte) und dem Erläuterungsband liefert die Karte auch weiterführende Modellvorstellungen über die Schichtlagerung und den tektonischen Bau des Blattgebietes. Die Erstellung der GK 25 obliegt den geologischen Landesbehörden, mit der „Geologischen Kartieranleitung“ (BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE 2002) liegt jedoch eine Arbeit vor, die das Ziel hat, zu einer länderübergreifenden Standardisierung der Kartenerstellung beizutragen. Ein weiterer Schritt in Richtung Standardisierung wird vom „Symbolschlüssel Geologie“ (PREUSS et al. 1991) geleistet. In diesem wird eine EDV-gerechte Kurzschreibweise für geologische Informationen definiert.

In Niedersachsen liegt derzeit für etwa 85 % des Landesgebietes eine GK 25 vor (vgl. Anhang A1). Die einzelnen Blätter der GK 25 weisen aus verschiedenen Gründen eine unterschiedliche Qualität auf (BOMBIEN & STOLZ 1993). Einerseits liegt dies an der über hundertjährigen Tradition der GK 25. Sehr alte Kartenblätter können durch den wissenschaftlichen Erkenntniszuwachs, der seit ihrer Erstellung erarbeitet wurde, nicht mehr direkt mit neueren Karten verglichen werden. Andererseits können auch die neueren Karten eine heterogene Datengrundlage aufweisen. So sind einige Karten aus Übersichtskartierungen oder Manuskriptkarten erstellt worden, welche naturgemäß eine weniger abgesicherte Datengrundlage aufweisen als vollständig abgearbeitete Karten. Weiterhin wird in einem Kartenblatt die Sichtweise und Interpretation eines oder mehrerer Autoren dargestellt; ein Kartenblatt ist also immer auch abhängig vom Bearbeiter.

Die Bearbeitung und Speicherung der GK 25 erfolgt am NLfB schon seit geraumer Zeit digital. Die Flächen- und Liniengeometrien der ehemals manuell angefertigten Karten wurden durch Digitalisierung oder Raster-scannen digital erfasst. Näheres über diese Prozesse kann bei BOMBIEN & LIES (1991) und BOMBIEN & STOLZ (1993) nachgelesen werden. Für die auf die Flächengeometrien bezogenen Attributdaten wurde eine relationale Datenbank aufgebaut (siehe Abschnitt 3.2.2.3). Punktdaten – typischerweise Lokalitäten von Bohrlöchern - sind in der GK 25 von Niedersachsen nicht enthalten. Für die Verwaltung von Bohrlokalitäten existiert in Niedersachsen traditionell eine separate Datenbank, die „Bohrdatenbank von Niedersachsen“ (PREUSS & VOSS 1983). Inhalte dieser Datenbank werden standardmäßig nicht in der GK 25 dargestellt.

Da in der längeren Prozedur der digitalen Erfassung der Karten verschiedene Methoden und verschiedene Softwareprodukte von verschiedenen Mitarbeitern bzw. Firmen angewendet worden sind, hat dies zur Folge, dass auch die Eigenschaften der digitalen Karten heute eine deutliche Heterogenität aufweisen. Beispielsweise ist die Stützpunktdichte der Geometrien oder die Genauigkeit der Georeferenzierung der Kartenblätter bei verschiedenen Karten sehr unterschiedlich (vgl. Abschnitt 4.3.2).

3.2.2 Das Datenmodell der GK 25

Im folgenden wird in Kürze das derzeitig am NLFb implementierte Datenmodell der GK 25 vorgestellt. Wie in Kapitel 1 bereits dargelegt, handelt es sich um ein duales System, welches zur Speicherung der Geometrien das Coverage-Format von Arc/Info (für Workstations) von der Firma ESRI benutzt. Die Attributdaten werden in einer relationalen Datenbank, implementiert in SQL-Server 2000 von Microsoft, verwaltet. Soweit es zum besseren Verständnis erforderlich erscheint, werden im nächsten Abschnitt Aspekte des Coverage-Formats kurz vorgestellt. Das relationale Datenbankmodell wird als bekannt vorausgesetzt bzw. kann in zahlreicher Literatur nachgelesen werden (z. B. DATE 1986, CODD 1990, SAUER 1992, CONNOLLY & BEGG 2002; mit Bezug zu GIS: ADAM & GANGOPADYHAY 1997, MAGUIRE, GOODCHILD & RHIND 1999; RIGAUX, SCHOLL & VOISARD 2002).

3.2.2.1 Exkurs: Das Coverage-Format von Arc/Info

Ein Coverage gruppiert raumbezogene Vektordaten eines bestimmten Gebietes und kann mit zu einem Kartenblatt in der manuellen Kartographie verglichen werden. Geographische Objekte der realen Welt werden bei Arc/Info abstrahiert in eine geometrische Komponente, die vom Arc-Modul verwaltet wird und in eine thematische Komponente, die prinzipiell vom Info-Modul verwaltet wird. In der Terminologie von Arc/Info ist ein geographisches Objekt ein Feature. (Im Folgenden werden die bei Arc/Info verwendeten Begriffe überwiegend im englischen Original wiedergegeben, um übersetzungsbedingte Missverständnisse zu vermeiden.) Ein Feature kann durch eines der folgenden geometrischen Primitive abgebildet werden (vgl. Abb. 3.2):

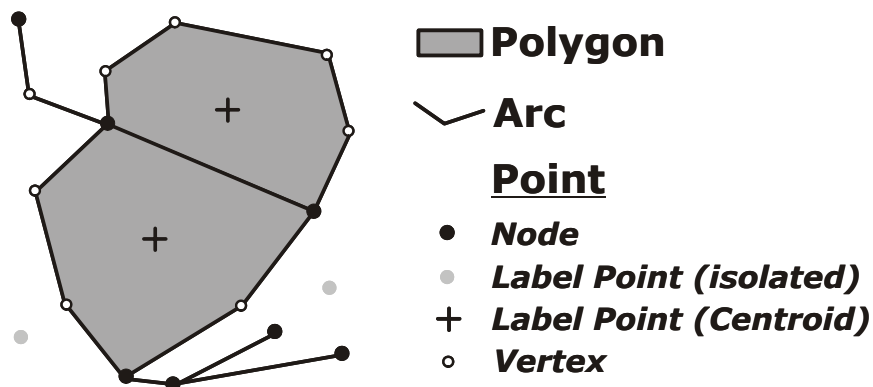


Abb. 3.2: Geometrische Primitive von Arc/Info (nach RIGAUX, SCHOLL & VOISARD 2002, verändert).

- Punkt (Point): beschrieben durch X- und Y-Koordinaten
- Arc: eine Sequenz von Liniensegmenten (Polylinie)

- Polygon: eine Sequenz von Arcs, die die umschließende Grenze des Polygons bilden

Für jede dieser drei Gruppen von Primitiven in einem Coverage verwaltet das Arc-Modul eine geometrische Datei (vgl. Abb. 3.3). In der LAB-Datei werden die Koordinaten der Punkte und ihre Identifier (ID) gespeichert. Die ARC-Datei enthält die Beschreibung der Arcs. Für jeden Arc sind dies: ID, Anzahl der Punkte (Vertizes) und X-, Y-Koordinaten der Punkte. Die PAL-Datei beschreibt die Polygone. Für jedes Polygon sind ein ID, die Anzahl der Arcs und eine Liste der betreffenden Arc-IDs enthalten.

LAB:

point#	x	y
12	3550	5270

ARC:

arc#	number of points	x1	y1	x2	y2	...	x7	y7
231	7	3567	5243	3598	5267	...	3587	5265

PAL:

polygon#	number of arcs	list of arcs
23	14	45, 65, -23, 98, ..., -88

Abb. 3.3: Geometrische Dateien von Arc/Info (nach RIGAUX, SCHOLL & VOISARD 2002, verändert).

Die topologischen Beziehungen (zum Begriff der Topologie siehe auch Abschnitt 4.2.2) der Features werden in weiteren Tabellen gespeichert (vgl. Abb. 3.4). Bevor diese Tabellen erklärt werden, ist es notwendig zu wissen, dass Arc/Info verschiedene Klassen von Features (Feature Classes) unterscheidet:

- Feature Class Arcs: Enthält alle Arcs eines Coverages, also eigenständige Linien-elemente und/oder die Grenzen von Polygonen.
- Feature Class Nodes: Enthält alle Nodes (Knoten). Diese sind vom Geometrie-Typ Punkt und beschreiben die Anfangs- bzw. Endpunkte der Arcs (also nicht der Vertizes).
- Feature Class Polygons: Polygone sind von Arcs umschlossene Flächen, die durch topologische Zeiger auf die entsprechenden Arcs definiert werden.
- Feature Class Label Points: Label Points sind ebenfalls vom Geometrie-Typ Punkt. Es handelt sich entweder um isolierte, topologisch unabhängige Punkte oder um Punkte innerhalb von Polygonen, die topologisch mit diesen verknüpft sind. Letztere Art von Label Points dient dazu, thematische Eigenschaften mit den Polygonen zu verknüpfen.

- Feature Class Tics: Hierbei handelt es sich um die vier Eckpunkte eines Coverages, die dazu dienen, das Coverage zu georeferenzieren.

Die AAT (Arc Attribut Table; vgl. Abb. 3.5) speichert nun die topologischen Beziehungen der Feature Class Arc. Die Tabelle enthält für jeden Arc des Coverages Einträge in mindestens sieben Spalten:

- FNODE#: ID des From-Nodes (Anfangsknoten) des Arcs.
- TNODE#: ID des To-Nodes (Endknoten) des Arcs.
- LPOLY#: ID des links vom Arc befindlichen Polygons.
- RPOLY#: ID des rechts vom Arc befindlichen Polygons.
- LENGTH: Länge des Arcs.
- *cover#*: Geometrischer ID des Arcs und Zeiger zur ARC-Datei. „*cover*“ ist dabei als Platzhalter zu verstehen. Das System trägt hier den Namen des Coverages ein.
- *cover-ID*: Thematischer ID des Arcs (auch als User-ID bezeichnet). Zeiger für die vom Info-Modul gesteuerte(n) Attribut-Tabelle(n). „*cover*“ ist wiederum ein Platzhalter.

Zusätzlich kann der Benutzer weitere Spalten an die Tabelle anhängen, um z. B. thematische Attribute der Arcs zu speichern. Im Info-Modul können weitere Attributtabelle(n) angelegt und mit der AAT verknüpft werden. Zur Attributdatenspeicherung kann aber auch ein anderes kommerzielles DBMS eingesetzt werden, um z. B. die in Kapitel 2, Punkt 3, genannten Vorteile für die Attributdaten zu nutzen. Diese Möglichkeit wird im Datenmodell der GK 25 verwendet, wie in Abschnitt 3.2.2.2 verdeutlicht wird.

Die NAT (Node Attribut Table) enthält die topologischen Beziehungen der Nodes zu den Arcs. Es sind mindestens Einträge in drei Spalten vorhanden (weitere können auch hier vom Benutzer eingefügt werden):

- ARC#: Der ID von dem Arc, der den Node als Anfangs- oder Endknoten hat. Beginnen oder enden mehrere Arcs an diesem Knoten, wird der kleinste ID genommen.
- *cover#*: Geometrischer ID des Nodes (analog zur AAT).
- *cover-ID*: Thematischer ID des Nodes (analog zur AAT).

Die PAT (Polygon Attribut Table; vgl. Abb. 3.4) beschreibt die Label Points. Sie hat mindestens Einträge in vier Spalten:

- AREA: Die Fläche des Polygons. (Bei isolierten Label Points wird der Eintrag auf „0“ gesetzt.)
- PERIMETER: Der Umfang des Polygons. (Bei isolierten Label Points wird der Eintrag auf „0“ gesetzt.)
- *cover#*: Geometrischer ID des Label Points (analog zur AAT).
- *cover-ID*: Thematischer ID des Label Points (analog zur AAT).

Weiterhin ist die CNT-Datei notwendig, um Label Points, die topologische Beziehungen zu Polygonen haben mit diesen zu verknüpfen. Sie enthält also eine Tabelle, in der jedem geometrischen ID eines Label Points der geometrische ID des zugehörigen Polygons zugeordnet ist.

Eine weitere wichtige Eigenschaft eines Coverages ist seine Fuzzy Toleranz. Dieser Wert wird benötigt, wenn die Topologie-bildenden Befehle von Arc/Info (z. B. *Clean* oder *Build*) ausgeführt werden sollen. Der Wert ist ein Maß für die Lagegenauigkeit der Koordinatenspeicherung. Die gespeicherten Koordinaten der Arcs (Knoten und Vertizes) eines Coverages müssen einen minimalen Abstand voneinander haben. Wenn dieser Abstand unterschritten wird, werden die entsprechenden Punkte zu einem Punkt zusammengeführt, dessen Lage nach bestimmtem Algorithmus berechnet wird. Die Fuzzy Toleranz definiert diese minimale Distanz. Das bedeutet, dass die Lage der Koordinaten sich bei der Ausführung der Topologie-bildenden Befehle maximal um die Distanz der Fuzzy Toleranz verschieben kann. Ihre Einheit entspricht der des räumlichen Bezugssystems des Coverages. Bei Gauss-Krüger Koordinaten ist dies z. B. die Einheit Meter. Wenn der Benutzer keine Fuzzy Toleranz vorgibt, berechnet Arc/Info eine Default Fuzzy Toleranz oder bricht den entsprechenden Befehl ab. Typische Werte der Fuzzy Toleranz betragen 1/100.000 bis 1/10.000 des Extents des Coverages (vgl. Abschnitt 3.2.2). Der Extent wird durch die minimalen und maximalen X- und Y-Koordinaten eines Coverages beschrieben.

Soweit zu den Konzepten des Coverage-Formats, die zum Verständnis dieser Arbeit benötigt werden. Weitere Details können z. B. bei MOREHOUSE (1992) oder RIGAUX, SCHOLL & VOISARD (2002) nachgelesen werden.

3.2.2.2 Geometriedaten

Die Geometriedaten eines TK 25-basierten Kartenblattes der GK 25 werden in einem Coverage gespeichert. Wie in Abschnitt 3.2.1 dargelegt, enthält die GK 25 Linien- und Flächenelemente. Die Flächenelemente, welche die geologischen Kartiereinheiten repräsentieren, werden in der Feature Class Polygons des Coverage gespeichert. Um die Verknüpfung mit den geologischen Attributdaten (Schichtenabfolge, Stratigraphie, Petrographie, Genese usw.) in der SQL-Server Datenbank zu gewährleisten, wird an die PAT die Spalte *Inhalt* angehängt. In dieser wird für jedes Polygon ein ID gespeichert, der eindeutig mit der vollständigen geologischen Beschrei-

bung der Schichtenabfolge assoziiert ist. Sobald die Beschreibung der Schichtenabfolge sich nur minimal ändert, z. B. in der Petrographie, muss ein neuer ID in der Spalte *Inhalt* eingetragen werden. Das heißt auch, dass ein ID mehrfach in der Spalte *Inhalt* auftreten kann, nämlich dann, wenn eine exakt identische Beschreibung der Schichtenabfolge mehrfach auf einem Coverage auftritt. Die ID sind weiterhin blattschnittfrei, d. h. der ID einer geologischen Flächenbeschreibung gilt für alle Coverages der GK 25. Die Verbindung mit der SQL-Server Datenbank erfolgt über Verknüpfung der Attribute *Inhalt* (des Coverages) und *KombiNr* (der Datenbank; Näheres in Abschnitt 3.2.2.3).

AREA	PERIMETER	GK252312#	GK252312-ID	INHALT
292161.6	8025.577	1	671841	16105
45678.9	768.9	2	671842	49356
...

Abb. 3.4: PAT von Arc/Info am Beispiel eines Coverages der GK 25.

In der Feature Class Arcs des Coverages werden zum einen – gemäß dem Coverage-Format (vgl. Abschnitt 3.2.1) – die Grenzen der Polygone abgelegt. Zusätzlich werden aber auch alle weiteren Linienobjekte, also Verwerfungen, Terrassenkanten, Erdfälle, Verwaltungsgrenzen usw. (vgl. Abschnitt 3.2.3), in dieser Feature Class gespeichert. Das bedeutet einerseits, dass den Arcs ihre jeweilige Bedeutung zugewiesen werden muss, und andererseits, dass ein Arc zwei Bedeutungen besitzen kann. Letzteres ist dann der Fall, wenn ein Arc sowohl die Grenze einer Kartiereinheit, also eines Polygons, darstellt, als auch gleichzeitig ein weiteres Linienobjekt repräsentiert, wie z. B. eine Verwerfung. Diese Bedeutungen werden den Arcs über das Attribut *Typ* zugewiesen, welches der AAT der Coverages angefügt wurde. Das Attribut *Typ* stellt den Fremdschlüssel des Schlüsselattributes *LinienTypNr* der Relation *LinienTypKat* der SQL-Server Datenbank dar. Über das Attribut *LinienTypKlar* dieser Relation kann die textliche Bedeutung jedes Linientyps herausgefunden werden. *Typ* „1“ bedeutet bspw. „Geologische Grenze“ (vgl. Abschnitt 3.2.3), also die Grenze zwischen zwei Kartiereinheiten. *Typ* „2“ bedeutet „Störung, flächentrennend“, d. h. es handelt sich um eine Störung, die gleichzeitig als Flächengrenze zwei geologische Kartiereinheiten voneinander trennt.

FNODE#	TNODE#	LPOLY#	RPOLY#	LENGTH	GK252312#	GK252312-ID	TYP
3	8	4	7	345.29	1	1	6
34	35	16	23	32.69	2	2	1
...	153

Abb. 3.5: AAT von Arc/Info am Beispiel eines Coverages der GK 25.

Das Koordinatenreferenzsystem der GK 25 sind Gauß-Krüger Koordinaten, die als Zentralmeridian den 3. Streifen (9° geographischer Länge) verwenden.

Des Weiteren sei hier angemerkt, dass die Fuzzy Toleranzen der Coverages (vgl. Abschnitt 3.2.1) stichprobenartig überprüft wurden (ca. 30 zufällig ausgewählte Coverages). Die Toleranzen lagen zwischen 1,0 und 1,3 Metern.

3.2.2.3 Attributdaten

Die Attributdaten der geologischen Kartiereinheiten der GK 25 werden in einer relationalen Datenbank verwaltet, welche in SQL-Server 2000 von Microsoft implementiert ist. Einen Überblick über die wichtigsten Relationen und ihre Beziehungen dieser Datenbank liefert Anhang A2. Die zentrale Relation hat den Namen Schichten und wird auch als Schichtenkatalog von Niedersachsen bezeichnet (BOMBIEN & STOLZ 1993). In dieser Tabelle ist jede in Niedersachsen kartierte Gesteinsschicht mit ihren Eigenschaften (Stratigraphie, Genese, Petrographie u. a.) einmalig abgelegt. Die Relation hat somit einen blattschnittfreien Charakter. Diese Relation steht in einer n:m-Beziehung zu den in den Coverages gespeicherten Flächendaten. Diese Beziehung ist über die attributierte Verknüpfungstabelle SchichtFolgen modelliert. Diese verknüpft die beiden eindeutigen Fremdschlüssel, nämlich den Schlüssel *Adresse* aus der Relation Schichten und den Schlüssel *KombiNr* aus der Relation KombiNrn. Als notwendige Information enthält die Verknüpfungstabelle SchichtFolgen noch das Attribut *SchichtNr*. Hierdurch wird auf ordinalem Skalenniveau die Schichtenabfolge (mit derzeit 1 bis 5 Schichten) jeder ausgewiesenen Kartiereinheit festgehalten. Da die Verknüpfungstabelle SchichtFolgen mit Gewährleistung von referentieller Integrität nur erstellt werden kann, wenn eine Relation vorliegt, in der als Schlüssel *KombiNr* geführt wird, muss die Relation KombiNrn in der Datenbank geführt werden. Das Schlüsselattribut *KombiNr* dieser Relation findet sich als Fremdschlüssel unter dem Namen *Inhalt* in der PAT der Coverages wieder. Die Verbindung zwischen der PAT und der SQL-Server Datenbank erfolgt über einen *Relate* (AML-Befehl) zwischen diesen beiden Attributen. Da in der SQL-Server Datenbank auch Tabellen bzw. Attribute abgefragt werden sollen, die eine 1:n- bzw. eine n:m-Beziehung aufweisen, kann nur über das Cursor-Processing (Befehle *DBMSCursor* u. a.) von AML gearbeitet werden, was einen relativ aufwendigen AML-Code erfordert. Gerade in diesem Punkt stellt ein integriertes System mit ArcGIS 8.x eine wesentliche Erleichterung dar (vgl. Kapitel 2, Punkt 4).

In der Datenbank ist weiterhin die Relation *Flaechen* geführt. Diese listet alle ausgewiesenen Kartiereinheiten der GK 25 blattschnittfrei und eindeutig (über das Schlüsselattribut *FlaechenNr*) auf. In dieser Relation werden über entsprechende Attribute auch Informationen über die Kombi-Nummern (*KombiNr*), die geographische Lage des Polygonmittelpunktes (*Rechts* und *Hoch*) und die Flächengröße (*Area*) der jeweiligen Kartiereinheit gespeichert. Diese Informationen müssen aus den entsprechenden Tabellen der Coverages ausgelesen werden. Das heißt, dass bei jeder Bearbeitung der Flächengeometrien im GIS die Relation *Flaechen* entsprechend aktualisiert werden muss und dass z. T. eine redundante Datenhaltung erfolgt. Über weitere Attribute (*Thema* und *ThemaNr*) der Relation *Flaechen* wird die Beziehung zu den Blattschnitten der TK 25 hergestellt. In den Relationen *Titel*daten, *Themen*Kat und

TK_Eckwerte finden sich dann Angaben über TK-Nummer und -Name, Koordinaten der Blattecken, Autor und Erscheinungsjahr der GK u. a.

Die Kodierung der Liniengeometrien wird in der Relation LinientypenKat aufgelöst. Hier findet sich zu jeder Linientypen-Nummer eine textliche Beschreibung, welche die Natur der Linie erläutert (vgl. Anhang A3).

Neben diesen meist fachlichen Attributen enthält die Datenbank auch Tabellen, die für die Visualisierung bzw. den Kartendruck benötigt werden. Da die Legende auf einer GK 25 nur die geologischen Einheiten zeigen soll, die tatsächlich auf dem jeweiligen Kartenblatt ausgewiesen sind, wird die Relation Legenden benötigt. In dieser ist gespeichert, welche Kombi-Nummern auf welchem GK-Blatt zu finden sind. Analoges gilt für die Linientypen und die Relation LinienTypen. Die Farbgebung und die Signaturen der geologischen Kartiereinheiten richten sich, wie in der Geologie üblich (z. B. BLASCHKE et al. 1989, VOSSMERBÄUMER 1991), nach Stratigraphie, Petrographie und Genese der einzelnen Schichten der Schichtenabfolge. Um dies zu gewährleisten, werden in den Relationen StratPetHGenese_Typen und StratPetHGenese_FarbKat Farbgebungstypen gebildet und ihnen Werte des CMYK-Farbschemas bzw. Signaturen zugewiesen.

3.2.3 Einbindung der Flächendaten in die Aufgaben und Strukturen des Niedersächsischen Landesamtes für Bodenforschung

Wie in Abschnitt 3.1 bereits dargelegt, hat die Verwaltung geologischer Flächendaten in einem GIS zwei übergeordnete Ziele, nämlich effiziente Möglichkeiten zur Visualisierung und zur Analyse der raumbezogenen geologischen Informationen bereitzustellen.

Die Zielgruppen, für die diese Möglichkeiten bereitgestellt werden sollen, lassen sich in zwei Nutzergruppen einteilen. Zum einen sind dies die Mitarbeiter des NLFb selbst, die die geologischen Daten benötigen. Zum anderen sind dies alle Personen außerhalb des NLFb, die einen Bedarf an geologischen Karten haben. Die folgenden Ausführungen beziehen sich nicht nur auf die digitale GK 25 sondern prinzipiell auch auf alle weiteren Kartenserien des NLFb (siehe Abschnitt 3.3), die prinzipiell in dieselben Arbeitsabläufe und Strukturen eingebunden sind.

Innerhäusig können prinzipiell alle Mitarbeiter über das Intranet auf die Coverages und die Attributdatenbanken der digitalen Kartenserien zugreifen. Mit dem auf dem lokalen Rechner installierten GIS können dann die Daten dem eigenen Bedarf entsprechend visualisiert oder analysiert werden. Änderungen der Daten dürfen nur von einem bestimmten Mitarbeiterkreis vorgenommen werden. Alle weiteren Mitarbeiter müssen für persönliche Änderungen lokale Kopien der Daten anlegen.

In einigen Arbeitbereichen, insbesondere der Bodenkunde, ist es häufig notwendig, bestimmte fachliche Methoden auf die Flächendaten anzuwenden, um z. B. Auswertungskarten für ein bestimmtes Thema zu erstellen. Als Beispiele seien die Erstellung einer Karte zur Hochwassergefährdung (vgl. CASPERS, ELBRACHT & SCHNEIDER 2004 folgt), zur Grundwasserneu-

bildung oder zur potentiellen Nitratauswaschungsgefährdung angeführt (vgl. MÜLLER et al. 1992). Die Methoden, die angewendet werden, sind teilweise sehr komplex und erfordern einen Dateninput aus vielen verschiedenen Datenquellen. Da die Arbeitsschritte zur Ausführung einer Methode jedoch immer gleich sind, wurde am NLfB eine Software, Methodenmanagementsystem (MeMaS) genannt, entwickelt, die diesen Prozess weitestgehend automatisiert. (Näheres zu MeMaS kann bei HEINEKE et al. (2002) oder unter <http://www.nlfb.de/boden/downloads/info-mappememas.pdf> (Stand 24.02.04) nachgelesen werden.) Im Zusammenhang mit dieser Arbeit ist wichtig, dass MeMas zur Bereitstellung von GIS-Funktionalitäten ArcView 3.x benutzt. Das bedeutet, dass alle Flächendaten nicht nur als Coverages sondern auch als Shapefiles vorliegen müssen, um auch in ArcView 3.x einsetzbar zu sein.

Insbesondere für Kunden des NLfB, aber auch für die Mitarbeiter, werden die Kartenblätter der Kartenserien auch als Plotausgabe zur Verfügung gestellt (vgl. Anhang A4). Die Plotausgabe erfolgt, unter Verwendung umfangreichen AML-Codes, aus Arc/Info heraus. Der AML-Code steuert dabei im wesentlichen das Kartenlayout, inkl. der Kartenrandobjekte und der dynamisch erzeugten Legende, und den Zugriff auf die Attributdatenbank. Neben der Plotausgabe haben die Kunden des NLfB auch die Möglichkeit, digitale Formate (Shapefile, Arc/Info E00) der gewünschten Kartenblätter zu erwerben (siehe <http://www.nlfb.de/produkte/preisliste/preisliste.htm>, Stand 13.12.03).

3.3 Die Datenmodelle der weiteren Kartenserien des Niedersächsischen Landesamtes für Bodenforschung

Wie in der Zielsetzung (Abschnitt 1.1) dargelegt, soll in dieser Arbeit bei der Konzeption des Datenmodells für die GK 25 auch die Übertragbarkeit auf die weiteren Kartenserien des NLfB beachtet werden. Dass eine solche Übertragbarkeit des Datenmodells nicht gänzlich ohne Modifikationen gegeben sein kann, ist auf Grund der unterschiedlichen Natur der Ausgangsdaten wohl unzweifelhaft. Es gilt also herauszuarbeiten, an welchen Stellen des Datenmodells diese Veränderungen vorgenommen werden müssen. Hierzu ist zunächst eine Analyse der derzeitigen Datenmodelle der weiteren Kartenserien des NLfB nötig. Zu diesen Kartenserien gehören:

- Geologische Karte von Niedersachsen im Maßstab 1:50.000 (GK 50)
- Bodenkarte von Niedersachsen im Maßstab 1:25.000 (BK 25)
- Bodenkarte von Niedersachsen im Maßstab 1:50.000 (BK 50)
- Bodenübersichtskarte im Maßstab 1:50.000 (BÜK 50)

Das Datenmodell der GK 50, inkl. der Attributdatenbank, unterscheidet sich nicht wesentlich von dem der GK 25. In der Praxis werden die Attributdaten der GK 25 und GK 50 sogar in den-

selben Tabellen derselben Datenbank gehalten. Der Bezug zur jeweiligen Kartenserie wird dabei durch das Attribut *Thema* hergestellt (vgl. Anhang A2). Durch die Generalisierung bzw. die Aggregation der Attributdaten verfügt die GK 50 z. T. über Attributwerte, die bei der GK 25 nicht vorkommen, wie z. B. die Linientypen „*Allgemeine Vereisungsgrenze*“ oder „*Sedimentologische Grenze im Watt*“ u. a. Des Weiteren hat bei der GK 50 bereits ein technischer und fachlicher Abgleich der Blattränder stattgefunden. Das heißt, bei dieser Kartenserie ist nicht mit geometrischen oder thematischen Fehlern am Blattrand zu rechnen.

Die Datenmodelle der drei bodenkundlichen Kartenserien sind untereinander identisch und das grundsätzliche Konzept entspricht dem der GK 25. Ein Unterschied ergibt sich bei der Verknüpfung zwischen Geometrie- und Attributdaten. Bei den bodenkundlichen Karten wird das System-Datenfeld *cover-ID* (vgl. Abschnitt 3.2.2.1) der PAT des Coverages zur Verknüpfung verwendet. Anders als bei der GK 25 handelt es sich hier also um ein eindeutiges und blattschnittfreies Schlüsselfeld. Weiterhin enthält die Feature Class *Arcs* (vgl. Abschnitt 3.2.2.1), im Gegensatz zur GK 25 (und GK 50), nur die Grenzen der Polygone und keine weiteren Linienobjekte. Auch hier enthält die AAT weitere benutzerdefinierte Spalten, um Typisierungen der *Arcs* vornehmen zu können. Neben den Polygongrenzen (hier Bodengrenzen genannt) existieren noch die Typen „*Gewässer- und Landesgrenze*“, sowie „*Grenze von anthropogen veränderter Fläche*“ und „*Kartenrahmen*“. Da eine Bodengrenze eine Mehrfachbedeutung haben kann (z. B. ist sie gleichzeitig Gewässergrenze), wurden der AAT vier Spalten hinzugefügt (*Typ_A*, *_B*, *_C*, *_D*), in denen diese Mehrfachbedeutung (also theoretisch bis zu vier Linientypen) kodiert werden kann. Die Spalten werden bei diesem System hierarchisch gewertet, d. h. *Typ_A* ist wichtiger als *Typ_B* und soll bei der Visualisierung vorrangig dargestellt werden (z. B. soll eine Bodengrenze, die zugleich Gewässergrenze ist, im Blau der Gewässergrenzen und nicht im Schwarz der Bodengrenzen erscheinen).

Die Datenstruktur der Attributdatenbank ist, bedingt durch die unterschiedlichen geowissenschaftlichen Thematiken, natürlich vollkommen anders als die der GK 25. Eine genaue Analyse wird an dieser Stelle - entsprechend der Zielsetzung des Projektes - jedoch nicht durchgeführt. Details über diese Kartenserien sind z. B. in BOESS, MÜLLER & SBRESNY (1999) und NIEDERSÄCHSISCHES LANDESAMT FÜR BODENFORSCHUNG (2000) zu finden.

Die Ergebnisse zur Übertragbarkeit der Datenmodelle werden in Abschnitt 5.1 berichtet.

3.4 Fazit: Anforderungen an das Datenmodell und die Datenmigration

Die in diesem Kapitel vollzogene Analyse der Ausgangssituation sollte dazu dienen, die Rahmenbedingungen, unter denen das Datenmodell der GK 25 zu entwickeln ist, kennen zu lernen und zu verstehen. Durch die Analyse des derzeit implementierten physischen Datenmodells kann das dahinterstehende, aber nicht explizit formulierte, konzeptuelle Modell abgeleitet

werden. Ein konzeptuelles Datenmodell wird, in gewissem Sinne rückwirkend, in Abschnitt 4.2 erarbeitet. Ebenso ist die Kenntnis des physischen Modells notwendig, um die Implementierung des neuen Datenmodells und die Datenmigration auch praktisch durchführen zu können (Abschnitte 4.3 und 4.4). Weiterhin wurden die Rahmenbedingungen untersucht, unter denen am NLFb mit der GK 25 gearbeitet wird (Abschnitte 3.2 und 3.3). Hieraus lassen sich Rückschlüsse auf weitere Anforderungen, die die Benutzer an ein Datenmodell stellen, ziehen. Bisher unerwähnt blieben zahlreiche intensive Diskussionen mit den Mitarbeitern des Referates „Fachinformationssystem Geologie, Flächen- und Raumdaten“, die für die GK 25 und die GK 50 verantwortlich sind. Diese Gespräche führten zu weiteren wichtigen Informationen und Anforderungen, die bei der Datenmodellierung, der Datenmigration bzw. dem gesamten neuen Systemdesign beachtet werden sollten. All diese Anforderungen und Aspekte werden im Folgenden zusammengefasst:

1) Das Datenmodell soll einfach sein

Das Datenmodell soll minimale Anforderungen an die Personen stellen, die mit ihm arbeiten. Da Personen mit unterschiedlicher Ausbildungsrichtung und unterschiedlichem Ausbildungsstand mit dem Modell arbeiten, sollen zum Verständnis des Modells geringst mögliches geologisches Fachwissen und geringst mögliche Softwarekenntnisse vorausgesetzt werden müssen. Einfachheit ist ein wesentlicher Aspekt, der zur Akzeptanz, zur Fehlervermeidung und zur langfristigen Funktionsfähigkeit eines Datenmodells beiträgt (ALLEN 1996, JOHNSON et al. 1999, PARENT, SPACCAPIETRA & ZIMÁNYI 1999).

2) Die geologischen Informationen sollen nicht verändert werden

Dieser Punkt ist einerseits generell zu sehen, d. h. bei der Migration dürfen die geologischen Informationen nicht verfälscht oder weggelassen werden. Dies ist jedoch nicht zu verwechseln mit der Genauigkeit der Positionsinformationen, der durchaus ein gewisser Spielraum zugebilligt wird (vgl. Abschnitt 4.3.3). Andererseits ist zu beachten, dass es sich bei den meisten Kartenblättern um bereits veröffentlichte Werke handelt, deren Inhalt nachträglich nicht leichtfertig verändert werden kann. Dieser Aspekt lässt jedoch auch eine gewisse Diskrepanz zum Konzept einer blattschnittfreien Datenbasis erkennen und wird in Abschnitt 5.1 diskutiert.

3) Die Struktur der Attributdatenbank soll möglichst erhalten werden

Die bereits in SQL-Server implementierte Attributdatenbank wurde entsprechend dem relationalen Datenmodell entworfen und im Praxiseinsatz getestet. Es besteht also keine zwingende Notwendigkeit, die Datenbankstruktur grundsätzlich zu ändern; es sollen nur notwendige Anpassungen vorgenommen werden.

4) Das Gauß-Krüger Koordinatensystem soll (zunächst) beibehalten werden

Vor dem Hintergrund, dass die amtliche Landesvermessung bereits auf UTM-Koordinaten umgestellt worden ist (vgl. <http://www.geodatenzentrum.de>, Stand 24.02.04), stellt sich prinzipiell die Frage, in welchem Koordinatenreferenzsystem die Daten gespeichert werden sollen. In dieser Arbeit wird weiterhin das Gauß-Krüger-Koordinatensystem verwendet. Eine Umstellung des Koordinatensystems ist eine Aufgabe, die nach der Umstellung des Systems in Betracht gezogen wird. In dieser Arbeit wird dieses Thema nicht berührt, nicht zuletzt, um den Rahmen der Arbeit nicht zu sprengen.

5) Der Arbeitsaufwand muss überschaubar bleiben

Die Migration zum integrierten System und notwendige Editierarbeiten am Datenbestand müssen in einem vertretbaren Zeit- und Personalrahmen vollzogen werden können. Ansonsten ist das Projekt unter den personellen und finanziellen Rahmenbedingungen des NLfB schlicht nicht durchführbar. Gegebenenfalls ist zu prüfen, ob – und wenn ja, wie – aufwändige Arbeitsschritte automatisierbar sind.

6) Die auf die GK 25 angewiesenen Arbeitsabläufe müssen unterstützt werden

Es ist unter den am NLfB gegebenen Rahmenbedingungen praktisch unmöglich, die Migration zu einem integrierten System, das sofort alle Arbeitsabläufe (vgl. Abschnitt 3.2.2) unterstützt, in einem großen Schritt durchzuführen. Es wird bspw. einige Zeit dauern bis MeMaS, die Plotausgabe von Karten oder der Internetvertrieb von Karten und Daten an die neuen Datenstrukturen angepasst werden können. In diesem Zeitraum werden jedoch Editierarbeiten bereits in dem neuen System durchgeführt werden müssen (vgl. Abschnitt 4.4.2). Das bedeutet, dass bei der Migration beachtet werden muss, dass alle Arbeitsabläufe stets auf die aktuellen Daten zugreifen können.

7) Das Datenmodell soll erweiterbar sein

Wohl niemand kann von sich behaupten, alle zukünftigen Anforderungen an ein Datenmodell vorhersagen zu können. Je komplexer ein System ist, desto schwieriger ist selbstverständlich seine Planung und desto wahrscheinlicher sind auch Ergänzungs- oder Änderungswünsche in der Zukunft. Schon bei der Datenmodellierung muss daher versucht werden, dass Modell erweiterbar zu gestalten, damit mögliche Veränderungen vorgenommen werden können.

8) Das Datenmodell soll auf weitere Kartenserien des NLfB übertragbar sein

Dieser Punkt wurde bereits mehrfach angesprochen (Abschnitte 1.1 und 3.1) und ist hier nur der Vollständigkeit halber noch einmal aufgeführt.

9) Ein technischer Abgleich der Blattränder soll erzielt werden

In den Abschnitten 1 und 1.1 wurde dargelegt, dass nach der Datenmigration mit Positions- und Attributfehlern an den ehemaligen Blattschnittgrenzen der Coverages zu rechnen ist. Als erster Schritt soll ein effektiver Weg gefunden werden, die Positionsfehler zu beheben. Das bedeutet, dass sämtliche Überlappungen und Lücken an den ehemaligen Blattschnittgrenzen eliminiert werden sollen. Für die Attributfehler ist vorerst keine einfache Lösung in Sicht, da hier umfangreiche fachliche Neubewertungen der geologischen Informationen notwendig sind (weitere Diskussion in Abschnitt 5.1).

4. Konzeption und Implementierung des Datenmodells für die Geologische Karte 1:25.000

Wie in Abschnitt 1.3 bereits erläutert, wird in diesem Kapitel die erarbeitete Datenmodellierung für die GK 25 beschrieben. Grundsätzlich soll dabei das in Abschnitt 1.2 erläuterte Schema des Datenmodellierungsprozesses eingehalten werden. Aus praktischen Gründen erscheint es in diesem Fall sinnvoll, das logische und das physische Modell in einem Schritt zusammenzufassen. Begründet ist dies folgendermaßen: Die Attributdaten sind bereits in einer relationalen Datenbank implementiert, d . h. der Modellierungsprozess ist bereits abgeschlossen und die Datenbank muss – wie noch gezeigt wird – nur geringfügig verändert werden. Die Geometriedaten werden vom konzeptuellen Modell direkt in das physische Modell für ArcGIS 8.3 überführt. Dies liegt daran, dass es bei den Geometriedaten für die Ebene des logischen Modells kein Äquivalent gibt. Es gibt kein allgemein verwendetes, dem relationalen Datenmodell entsprechendes Konzept, welches dem proprietären Datenmodell der Geodatabase übergeordnet ist (vgl. Abschnitt 1.2). Dies ist bei dem vorliegenden Datenmodell aber kein Nachteil, da dessen Komplexität durchaus noch überschaubar ist.

Als erstes wird also das konzeptuelle Datenmodell vorgestellt, welches auf Grundlage von Kapitel 3 entworfen wurde (Abschnitt 4.2). Dann folgen schwerpunktmäßig das physische Modell (Abschnitt 4.3) und Aspekte zu dessen Implementierung (Abschnitt 4.4). Da die Datenmodellierung mit Hilfe von UML-Diagrammen visualisiert wird, erfolgt im nächsten Abschnitt zunächst eine kurze Erläuterung zur UML und zu Konzepten der Objektorientierung, soweit dies zum Verständnis dieser Arbeit notwendig ist.

4.1 Exkurs: Objektorientierung und die *Universal Modeling Language*

Bei der objektorientierten Modellierung werden die Bestandteile der realen Welt, welche in das Modell eingehen sollen, in Objekte gegliedert. Bestandteile der realen Welt können hierbei physikalisch existierende Objekte, wie z. B. ein Baum, oder jedwede Art von Konzepten sein, wie z. B. eine geologische Kartiereinheit. Objekte sind dabei gekennzeichnet durch einen Namen, ihren Zustand und ihr Verhalten. Als Zustand werden Attribute (auch Eigenschaften genannt) aufgefasst, die das Objekt beschreiben. Eine bestimmte geologische Kartiereinheit hat z. B. eine bestimmte Ausprägung der Eigenschaften *Petrographie*, *Genese* oder *geographische Lage*. Das Verhalten wird beschrieben durch Methoden (auch Operationen genannt), die ein Objekt ausführen kann, um z. B. Zustandsänderungen zu bewirken. Bei den in dieser Arbeit modellierten Objekten sind Methoden weniger wichtig, da die Objekte zumeist unter einer statischen Sichtweise betrachtet werden. Ein Beispiel für eine Methode wäre *Starten* des Objektes *Flugzeug*. Ein Objekt wird durch seinen Namen, seinen Zustand und sein Verhalten komplett beschrieben. Diese Bündelung wird als Kapslung bezeichnet. Die Kapslung beinhaltet auch, dass auf die Attribute eines Objektes nicht ohne weiteres zugegriffen werden kann – sie sind nur über bestimmte Methoden des Objektes zu manipulieren. Damit zwischen den verschiedenen Objekten Interaktion stattfinden kann, müssen diese miteinander kommunizieren können. Dies geschieht über sogenannte Nachrichten. Kapslung und Nachrichten sind eine wichtige Voraussetzung für den in der objektorientierten Programmierung erwünschten modularen Aufbau von Programmen. Ein weiteres Prinzip der Objektorientierung ist in den vorherigen Sätzen bereits angeklungen: Objekte werden in Kategorien zusammengefasst, welche Klassen genannt werden. Ein bestimmtes Objekt einer Klasse wird als Instanz bezeichnet. Die bestimmte (vereinfachte) Kartiereinheit *Löß über Geschiebelehm* an einem bestimmten Ort ist bspw. eine Instanz der Klasse *Geologische Kartiereinheit*. In der UML (zur Definition der UML siehe Abschnitt 1.2) werden Klassen durch Klassendiagramme visualisiert (vgl. Abb. 4.1). Wie in der Abbildung zu sehen ist, können nach den Attributen noch weitere Informationen, wie z. B. Datentyp oder Domäne (zulässiger Wertebereiche) genannt sein.

Das nächste wichtige Konzept der Objektorientierung ist das der Vererbung. Klassen können von anderen bereits bestehenden Klassen abgeleitet werden und erben dann deren definierte Attribute und Methoden. Die übergeordneten Klassen werden als Superklassen und die abgeleiteten als Subklassen bezeichnet. Hierdurch wird im allgemeinen eine Spezialisierung bzw. - in umgekehrter Richtung - eine Generalisierung ausgedrückt. Durch weitere oder anders definierte Attribute und Methoden wird der Subklasse gegenüber der generellen Klasse eine spezialisierte Funktionsweise zugewiesen. Aus einer Klasse *Gesteinstypen* könnten bspw. die Subklassen *Magmatite*, *Sedimente* und *Metamorphite* abgeleitet werden. Dies verdeutlicht auch gleich den Unterschied zwischen abstrakten Klassen und konkreten Klassen. Abstrakte Klassen, in diesem Fall die Klasse *Gesteinstypen*, dienen dazu, Muster zu definieren. Sie werden aber nicht instanziiert, d. h. es gibt keine Objekte von abstrakten Klassen. Die Objekte gehören immer den untergeordneten, konkreten Subklassen an. Ein Vorteil hiervon ist leicht zu erkennen: Muster, die in mehreren thematisch ähnlichen Klassen vorkommen, brauchen nur einmal in

einer entsprechenden abstrakten Klasse definiert zu werden. Damit die Vererbung funktioniert, muss erlaubt sein, dass Objekte unterschiedlicher Klassen auf dieselbe Nachricht unterschiedlich reagieren. Schickt z. B. ein Objekte *A* die Nachricht *Fläche berechnen* an ein Objekt *B*, so muss dieses, je nachdem ob es sich um ein Objekt der Klasse *Kreis*, *Dreieck* oder *Rechteck* handelt, seine Fläche nach einer anderen Methode berechnen. Dieses Konzept wird Polymorphismus genannt.

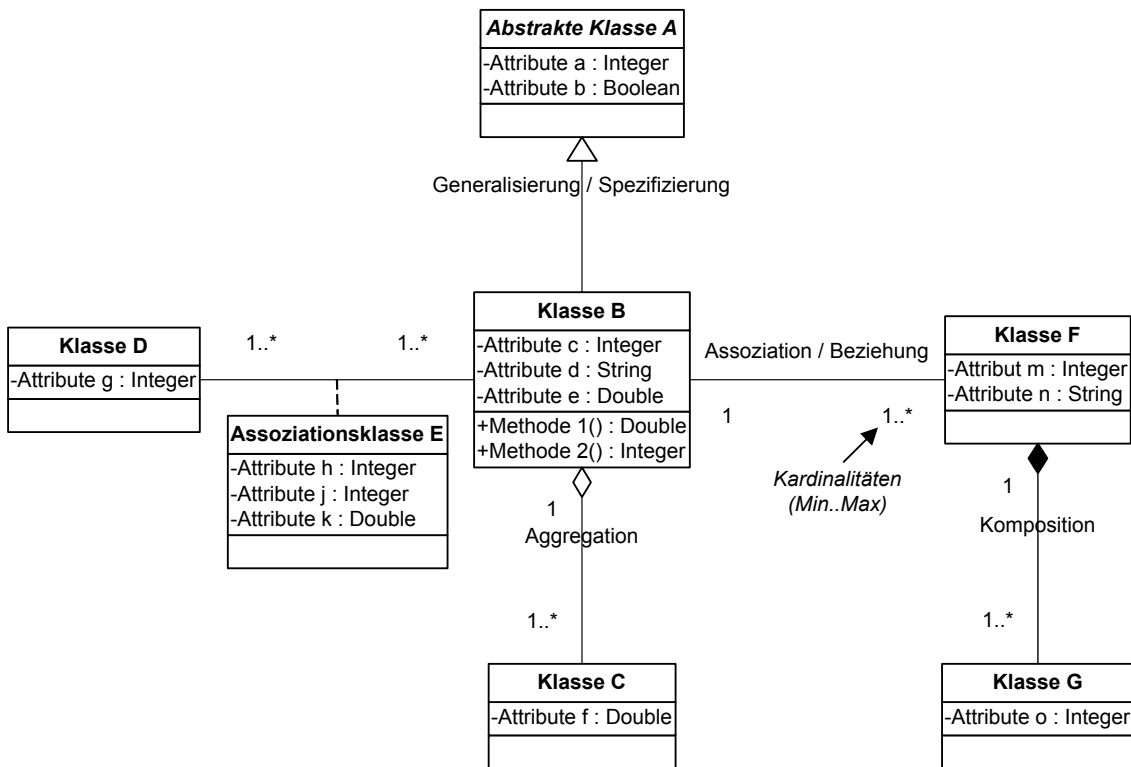


Abb. 4.1: Schema eines UML-Klassendiagramms und UML-Notation.

In der UML-Notation werden die Vererbung, aber auch andere Beziehungen zwischen Objekten, als sogenannte Assoziationen visualisiert (vgl. Abb. 4.1). Assoziationen können, neben ihrem Namen, durch Kardinalitäten gekennzeichnet sein. Kardinalitäten geben an, wie viele Objekte der an der Assoziation beteiligten Klassen miteinander in Beziehung stehen (vgl. relationales Datenmodell). Ist nur eine Zahl als Kardinalität eines Assoziationsendes angegeben, so ist dies die genaue Zahl der beteiligten Objekte. Bei zwei angegebenen Zahlen (z. B. 0..5) handelt es sich um eine Spanne, wobei links der Minimal- und rechts der Maximalwert aufgeführt werden. Das Zeichen * steht für eine unbestimmte Anzahl von Objekten. Eine Assoziation kann weiterhin als Aggregation oder Komposition gekennzeichnet sein (vgl. Abb. 4.1). Aggregation bedeutet, dass Objekte einer Klasse aus Objekten einer anderen Klasse zusammengesetzt werden. Ein Objekt der Klasse *Geologische Kartiereinheit* kann z. B. aus Objekten der Klasse *Geologische Schicht* zusammengesetzt werden. Die Raute befindet sich dabei auf der Seite der zusammengesetzten Klasse. Eine Komposition ist eine strenge Form der Aggregation, die zugleich bedeutet, dass das Teilobjekt nur einem Aggregat angehören und dass es nicht unab-

hängig von diesem existieren kann. Bei der Datenmodellierung kann es sinnvoll sein, auch Assoziationen mit Attributen zu versehen. Ein Beispiel hierfür ist die Modellierung von n:m-Beziehungen. Im relationalen Modell wird eine n:m-Beziehung in eine Verknüpfungstabelle und zwei 1:n-Beziehungen aufgelöst. In UML-Diagrammen kann dies mit einer Assoziation und einer ihr (über eine gestrichelte Linie, s. Abb. 4.1) zugeordneten Assoziationsklasse dargestellt werden. Neben den zur Zuordnung notwendigen Schlüsselattributen kann die Assoziationsklasse auch weitere Attribute beinhalten. Ein weiterer Bestandteil der UML-Notation ist das Paket. Mit Hilfe von Paketen können komplexere Modelle besser gegliedert werden, in dem z. B. sinnvoll abgegrenzte Submodelle gebildet werden.

Nicht alle erläuterten Konzepte der Objektorientierung kommen bei der Datenmodellierung in dieser Arbeit zur Anwendung. Ein etwas umfangreicherer Einblick in die Objektorientierung hilft jedoch dabei, die UML-Klassendiagramme richtig zu verstehen. Weitere Details zur Objektorientierung und der UML können bei Bedarf der in Abschnitt 1.2 genannten Literatur entnommen werden.

4.2 Konzeptuelles Datenmodell

In diesem Abschnitt wird das konzeptuelle Datenmodell vorgestellt, welches auf Grundlage der in Kapitel 3 erarbeiteten Ergebnisse entwickelt wurde. Wie in Abschnitt 1.2 erläutert, stehen hierbei die geologischen, raumbezogenen Fachdaten im Vordergrund, d. h. dass das Modell nicht vollständig alle benötigten Aspekte eines Fachinformationssystems abdeckt. Beispielsweise werden darstellungsbezogene Themen, wie z. B. die Einbindung von topographischen Flächendaten, nicht berücksichtigt. Andere darstellungsbezogene Aspekte, wie einige Kartenrandobjekte, werden jedoch in das Modell aufgenommen und kurz diskutiert, um Anregungen für die Weiterentwicklung des Datenmodells zu geben. Das konzeptuelle Datenmodell ist als UML-Klassendiagramm im Anhang B1 visualisiert.

4.2.1 Geologische Entitäten

Die geologischen Entitäten, welche bei der GK 25 von Bedeutung sind, können aus den Coverages der GK 25 und dem Linientypen-Katalog (Anhang A3) abgeleitet werden und sind in Tab. 4 aufgelistet.

Nicht alle Linientypen wurden in das konzeptuelle Modell übernommen, da einige Typen aus verschiedenen Gründen wegfallen können (siehe Tab. 4.2). Dies sind einige nicht-geologische Linientypen, wie z. B. *Verwaltungsgrenzen*, *Grenzen der Bearbeitung*, *Kartenrahmen* u. a. Zukünftig sollen diese Art von Grenzen als separate Themen bzw. z. T. gar nicht mehr gespeichert werden. Andere Linientypen, wie z. B. *Rutschmassen*, *Umgrenzung einer Wurte* oder *Gewässergrenze* werden nicht übernommen, da sie keine eigenständigen Entitäten sondern Visualisierungen von Attributen der Entität Geologische Kartiereinheit sind und somit aus den Flächenbeschreibungen abgeleitet werden können.

Kurzname	Bedeutung	Geometrie- typ	Arc/Info Linientyp Nr.
Geologische Kartiereinheit	Homogene Schichtenfolge	Polygon	-
Störung	Alle Typen von Störungen	Polylinie	2-7, 13-17, 59-93
Erdfall	Erdfälle und Subrosionssenken	Polylinie	24-28, 164-165
Terrasse	Alle Typen von Terrassenkanten	Polylinie	20, 39, 45
Endmoräne	Streichrichtung von Endmoränen	Polylinie	151
Stauchmoräne	Streichrichtung Streichmoränen	Polylinie	30
Flussrinne	Ehemalige Flussrinnen	Polylinie	156
Transgression	Übergreifende Lagerung	Polylinie	12
Schichtbank	Schichtbänke	Polylinie	49-57
Gipseinlagerung	Gips- u. Anhydriteinlagerungen	Polylinie	48
Braunkohleabbau	Grenzen Braunkohleabbau	Polylinie	22
Tagebau	Tagebaugrenzen	Polylinie	21
Kohleflöz	Kohleflöze	Polylinie	170
Mesoz. Grenze	Schichtgrenze im Mesozoikum unter mehr als 2 m Quartär	Polylinie	23
Kalkige Schicht	Verbreitung kalkhaltiger Schichten im Quartär	Polylinie	35
Granitkontakthof	Grenze des Granitkontakthofes	Polylinie	58

Tab. 4.1: Geologische Entitäten im konzeptuellen Modell der GK 25 (vgl. Anhang B1).

Typ-Nr.	Bedeutung	Grund des Wegfallens
1	Grenze der Kartiereinheit	wird durch Polygongrenzen ersetzt
8	Kartenrahmen	blattschnittfrei
9	Gewässergrenze	wird über Attribut <i>KombiNr</i> der Kartiereinheiten visualisiert
11, 110	Bundeslandgrenzen	eigene Feature Class
31	Grenze glaziärer Schollen oder Schuppen	wird über Attribut <i>Genese</i> der Kartiereinheiten visualisiert
33-34, 36	Grenzen von Rutschmassen	wird über Attribut <i>Genese</i> der Kartiereinheiten visualisiert
40	Umgrenzung einer Wurte	wird über Attribut <i>Genese</i> der Kartiereinheiten visualisiert
115	Grenze der geologischen Bearbeitung	fällt weg

Tab. 4.2: Im konzeptuellen Modell wegfallende Linientypen des Arc/Info-Linientypenkataloges.

Im folgenden werden die Klassen des UML-Klassendiagramms (Anhang B1), welche die geologischen Entitäten repräsentieren oder in direktem Bezug zu ihnen stehen, aufgelistet und erläutert.

Objekt

Alle Klassen des Klassendiagramms leiten sich aus dieser abstrakten Klasse ab und verfügen daher über einen eindeutigen *Objekt-Identifer* (OID).

Geographisches Objekt

Aus dieser abstrakten Klasse werden alle Klassen abgeleitet, die einen Raumbezug aufweisen (symbolisiert durch das Attribut Geometrie).

Flächenobjekt

Eine abstrakte Klasse aus der alle Klassen abgeleitet werden, die eine Fläche (Polygon mit bestimmter *Flächengröße* und bestimmtem *Umfang*) beschreiben.

Linienobjekt

Eine abstrakte Klasse aus der alle Klassen abgeleitet werden, die eine Linie (Polylinie mit bestimmter *Länge*) beschreiben. Im vorliegenden konzeptuellen Modell wird nur die Klasse Geologisches Linienobjekt von dieser Klasse abgeleitet. Es ist jedoch denkbar, das Modell durch weitere Linien-Klassen, wie z. B. Verwaltungsgrenzen, zu erweitern.

Geologische Kartiereinheit, Geologische Schicht, Schichtenabfolge, Sonstige Kartiereinheit

Eine der elementaren Klassen der GK 25 ist die Klasse Geologische Kartiereinheit. Diese Klasse repräsentiert die vom Geologen kartierte Flächeneinheit, für die die Schichtenabfolge als homogen betrachtet wird. Dieselbe Schichtenabfolge kann in mehreren Instanzen der Klasse Geologische Kartiereinheit realisiert sein. Alle möglichen Schichtabfolgen sind in der Klasse Schichtenabfolge einmal enthalten und durch den von der Klasse Objekt geerbten *OID* (hier in Anlehnung an das Arc/Info-Modell *KombiNr* genannt) eindeutig gekennzeichnet. Die Klasse weist keinen Raumbezug auf. Die Klasse Geologische Kartiereinheit setzt sich demnach aus genau einem Objekt der Klasse Schichtenabfolge zusammen. Diese wiederum setzt sich zusammen aus Null oder mehreren (in der Praxis nicht mehr als fünf) Objekten der Klasse Geologische Schicht. In dieser Klasse sind alle beschriebenen geologischen Schichten einmal enthalten und auch sie weist keinen Raumbezug auf. Zwischen den Klassen Schichtenabfolge und Geologische Schicht herrscht eine n:m-Beziehung. Diese Beziehung muss ein Attribut besitzen, welches die Rangfolge der Schichten von der Erdoberfläche in die Tiefe auf ordinalem Skalenniveau wiedergibt. Diese Funktion nimmt im UML-Diagramm das Attribut *AbfolgeNr*

der Assoziationsklasse Schichtenrangfolge wahr. Die weiteren Attribute dieser Klasse (*KombiNr* und *GS OID*) repräsentieren die Fremdschlüssel der Klassen Schichtenabfolge bzw. Geologische Schicht.

Nicht jede kartierte Fläche ist durch eine geologische Schichtenabfolge gekennzeichnet. Vorkommen können auch nicht-geologische Flächenbeschreibungen, wie z. B. Gewässer. Solcherart Flächenbeschreibungen werden durch die Klasse Sonstige Kartiereinheiten modelliert.

Geologisches Linienobjekt

Aus dieser abstrakten Klasse werden alle Klassen abgeleitet, die ein geologisches Linienobjekt repräsentieren sollen. Das Attribut *Linientyp* dient zur Klassifizierung der verschiedenen Linien. Eine aggregierende Klassifizierung der Linienobjekte ist aus Darstellungsgründen nötig, denn die verschiedenen Linienobjekte werden auf einer Karte z. T. durch dieselbe Symbolik visualisiert (vgl. Abschnitt 4.3).

Störung, Störung mit Flächenbezug, Störung ohne Flächenbezug

Aus der abstrakten Klasse Störung werden die beiden Klassen Störung mit Flächenbezug und Störung ohne Flächenbezug abgeleitet. Dies ist notwendig, um die topologischen Beziehungen zu verdeutlichen. Nur Objekte der Klasse Störung mit Flächenbezug weisen eine topologische Beziehungen zur Klasse Geologische Kartiereinheit auf (vgl. Abschnitt 4.2). Die Attribute sind, mit Ausnahme von *Hierarchie*, vom Datentyp Boolean (Ja/Nein) und wurden im alten System als Linientypen kodiert. *Einfallsrichtung: Ja* bedeutet, dass in analogen Manuskriptkarten die Einfallsrichtung und evtl. der Einfallswinkel der Störung eingetragen sind. Die Einfallswinkel wurden bisher nicht digitalisiert. Sollte dies in Zukunft geschehen, kann die Klasse um entsprechende numerische Attribute erweitert werden. *Hierarchie* kann drei Zustände, nämlich *Haupt-* bzw. *Nebentörung* oder *keine Angabe*, aufweisen. Einige Störungen/Gänge im Harz sind durch Namen, wie z. B. *Rodentaler Gang* gekennzeichnet, welcher im Attribut *Name* festgehalten werden kann.

Erdfall, Erdfall mit Flächenbezug, Erdfall ohne Flächenbezug

Diese Klasse ist analog zur Klasse Störung aufzufassen. Das Attribut *Unmaßstäblich* kennzeichnet Erdfälle/Subrosionssenken, deren Größe in der 1:25.000er Karte aus kartographischen Gründen übertrieben dargestellt worden ist. Prinzipiell ist zu überdenken, ob es sinnvoll ist, Objekte, die für die Darstellung im Zielmaßstab zu klein sind, weiterhin mitzuführen. Es ist denkbar, dass ihre kleinen Geometrien zukünftig zu Problemen bei der Datenhaltung, z. B. bei einer Generalisierung der Geometrien, führen können (vgl. Abschnitt 4.3.3). Zunächst wird jedoch die Anforderung beachtet, dass die geologischen Informationen der Originale nicht verändert werden sollen (vgl. Punkt 2 in Abschnitt 3.4).

Terrassenkante, Terrassenkante mit Flächenbezug, Terrassenkante ohne Flächenbezug

Diese Klasse ist ebenfalls analog zur Klasse Störung aufzufassen. Im Attribut *Typ* kann die Art der Terrasse, z. B. *Mittelterrasse*, festgehalten werden.

Schichtbank

Diese Klasse weist lediglich ein Attribut zur Speicherung der *Stratigraphie* auf. Analog dazu verhält sich die Klasse Gipseinlagerung.

Kohleflöz

Die Objekte dieser Klasse weisen keine weiteren Eigenschaften auf. Analog hierzu sind die Klassen Endmoräne (Streichrichtung), Stauchmoräne (Streichrichtung), (ehemalige) Flussrinne, Granitkontakthof (Harz), Transgression, Mesozoische Schichtgrenze (unter 2 m Quartär) und Kalkige Schicht (im Quartär) zu betrachten.

Tagebau, Braunkohleabbau

Grundsätzlich wird eine anthropogene Beeinflussung von geologischen Schichten in deren Beschreibung gekennzeichnet (z. B. durch „y ,Tagebau““ im Attribut *Genese*). Da eine Tagebaugrube sich jedoch über mehrere Kartiereinheiten erstrecken kann, ist es nicht (einfach) möglich die Außengrenze der Grube über die Grenzen der Kartiereinheiten zu visualisieren, wie es z. B. bei Rutschmassen geschieht. Daher wird eine eigenständige Klasse Tagebau benötigt. Weiterhin ist zu bemerken, dass der Braunkohleabbau in der Vergangenheit z. T unter Tage erfolgte, so dass die obersten Schichten nicht direkt beeinflusst worden sind und folglich auch nicht anders beschrieben wurden als angrenzende Schichten ohne Braunkohleabbau. Um dies zu visualisieren, wird die Klasse Braunkohletagebau benötigt.

Die weiteren in Anhang B1 dargestellten Klassen und Assoziationen dienen entweder zur Modellierung der Topologie oder zur Modellierung der Visualisierung der geologischen Informationen auf einem Kartenblatt und werden in den folgenden Abschnitten erläutert.

4.2.2 Topologie

Im Zuge der Definition eines neuen Datenmodells kann die Chance genutzt werden, über einige, vielleicht für selbstverständlich gehaltene, Konzepte neu nachzudenken. Die Daten der GK 25 wurden bisher im Coverage-Format von Arc/Info gespeichert, welches die topologischen Beziehungen zwischen den Daten explizit speichert (vgl. Abschnitt 3.2.2.1). Weitverbreitete, erfolgreiche Produkte, wie z. B. das Desktop-GIS ArcView 3.x, benutzen jedoch nicht-topologische Datenstrukturen. (In Anlehnung an THEOBALD (2001) wird die nicht-topologische

Datenstruktur, wie sie im Shapefile von ArcView 3.x benutzt wird, im Folgenden mit KDS, für kartographische Datenstruktur, abgekürzt. Einige Autoren (z. B. RIGAUX, SCHOLL & VOISARD 2002) zählen den Shapefile zu den sogenannten Spagetti-Datenstrukturen, was jedoch nach den differenzierteren Definitionen von THEOBALD (2001) oder MAGUIRE & DANGERMOND (1991) ungenau ist.)

Die Fragen, die sich bei der Reflexion stellen, lauten: Ist für die betrachteten Daten eine Speicherung in einer topologischen Datenstruktur (TDS) überhaupt notwendig? Und: Welche Vorteile und Nachteile ergeben sich daraus? Um diese Fragen zu beantworten, wird in diesem Abschnitt etwas ausführlicher auf die Topologie eingegangen. Zunächst eine Definition der Begriffe *topologische Beziehungen* zwischen raumbezogenen Daten, wie sie im Zusammenhang mit GIS verwendet werden (nach EGENHOFER 1994; HARDY 2001; THEOBALD 2001; WATSON 2002; HOEL, MENON & MOREHOUSE 2003): Es sind räumliche Beziehungen, die gegenüber planaren Verzerrungen von metrischen Eigenschaften, wie Richtung und Distanz, invariant bleiben. Die wichtigsten topologischen Beziehungen sind Kontiguität, Konnektivität und Eingeschlossenheit. Geographische Objekte (*Features*, vgl. Abschnitt 3.2.2.1), die in einem topologischen Zusammenhang stehen, haben z. T. die gleichen geometrischen Eigenschaften (Vertizes, Kanten).

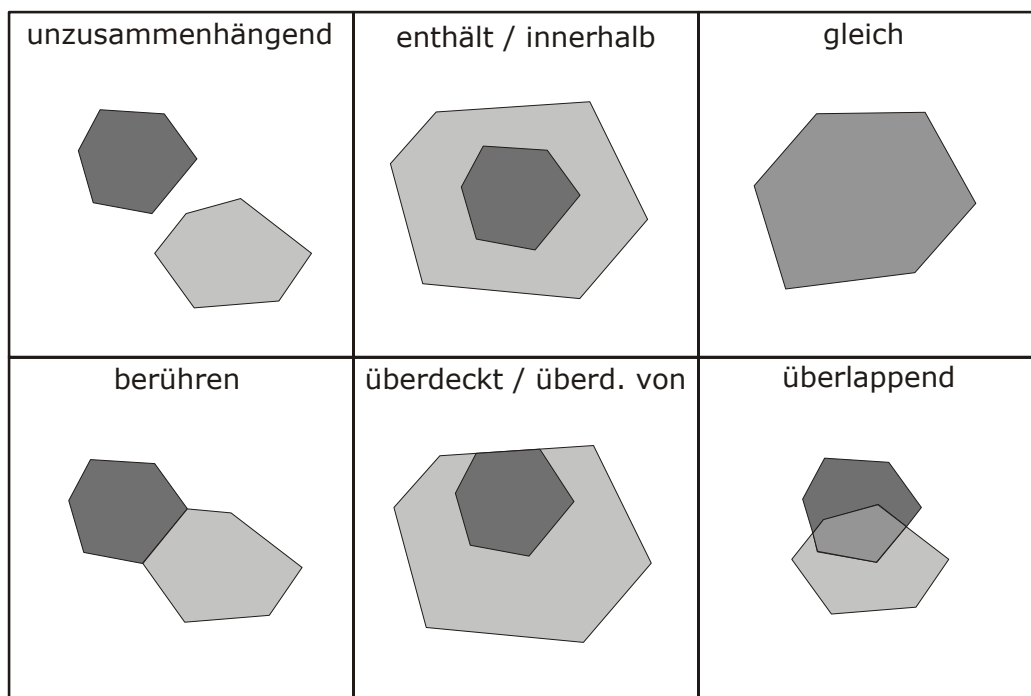


Abb. 4.2: Topologische Beziehungen zwischen zwei Polygonen im zweidimensionalen Raum (nach EGENHOFER 1994, verändert).

Die Abb. 4.2 zeigt beispielhaft die möglichen topologischen Beziehungen zwischen zwei Polygonen im zweidimensionalen Raum.

Die Anzahl topologischer Beziehungen, die zwischen den Gruppen der geometrischen Primitiven (Punkt, Linie, Polygon) im zweidimensionalen Raum möglich sind, gibt Tab. 4.3 wieder. Zwischen Objekten derselben Dimensionalität (also Punkt/Punkt, Linie/Linie oder Polygon/Polygon) können Umkehr-Beziehungen auftreten, wie z. B. *Polygon A überdeckt Polygon B* bedeutet gleichzeitig *Polygon B ist überdeckt von Polygon A* (vgl. Abb. 4.3). Diese Beziehungen werden auf Grund der Austauschbarkeit der Polygone in Tab. 4.1 nur als eine Beziehung gewertet.

Gruppe von Primitiven	Anzahl der Beziehungen
Punkt / Punkt	2
Punkt / Linie	3
Punkt / Polygon	3
Linie / Linie	23
Linie / Polygon	19
Polygon / Polygon	6

Tab. 4.3: Anzahl topologischer Beziehungen per Gruppierung von geometrischen Primitiven (nach UBEDA & EGENHOFER 1997; Erläuterung im Text).

THEOBALD (2001) weist darauf hin, dass alle geographischen Daten implizit topologische Beziehungen enthalten. Als wesentliche Vorteile einer expliziten Speicherung von Topologie (wie im Coverage-Format) werden genannt (THEOBALD 2001; WATSON 2002):

- Sicherung der Datenintegrität. (Fehler, wie Lücken, Überlappungen usw., können vermieden bzw. automatisch eliminiert werden.)
- Bessere Performance bei Abfragen auf Kontiguität, Konnektivität oder Eingeschlossenheit. Hier ist hervorzuheben, dass eine TDS nicht zwingend erforderlich ist, um diese Art von Abfragen auszuführen, denn auch mit KDS ist dies „on the fly“ möglich. Dazu sind jedoch relativ aufwändige und Performance-intensive Algorithmen notwendig, die die Geometrien der betroffenen Features vergleichen (THEOBALD 2001; WATSON 2002). Der Einsatz dieser Algorithmen mit akzeptabler Performance in kommerziellen GIS war historisch angewiesen auf die technologische Entwicklung von ausreichender Prozessorleistung.
- Geringerer Speicherplatzbedarf, wenn die geteilten Geometrien nur einmal gespeichert werden. Die Praxis zeigt allerdings, dass dies nicht immer der Fall sein muss. Traditionelle TDS brauchen mehrere Dateien um die Geometrien zu speichern, während KDS nur eine Datei dazu brauchen. Damit können sie den erhöhten Speicherplatzbedarf durch die doppelt gespeicherten Geometrien kompensieren. Weiterhin beansprucht die

Speicherung der Attributdaten in vielen Fällen den größten Anteil an Speicherplatz. Der Vorteil des geringeren Speicherplatzbedarfs muss also - insbesondere bei der steigender Verfügbarkeit von günstigem Speicherplatz –stark relativiert werden (THEOBALD 2001).

Es zeigt sich also, dass die beiden letztgenannten Vorteile von TDS mit der zunehmenden Leistungsfähigkeit der Rechner immer mehr an Bedeutung verlieren. Doch allein der erste Punkt ist für viele Nutzer so wichtig, dass Topologie nach wie vor ein wichtiger Bestandteil eines GIS ist (vgl. UBEDA & EGENHOFER 1997).

Die Vorteile einer TDS müssen jedoch durch einige Nachteile erkauft werden:

- Topologie muss in vielen Fällen erst einmal gebildet werden. Hierzu sind z. T. aufwändige Editierarbeiten notwendig.
- Der Bildschirmaufbau ist bei einer traditionellen TDS oft langsamer als bei einer KDS. Dies liegt daran, dass bei einer TDS die Geometrien aus mehreren Dateien bzw. Tabellen gelesen werden müssen, was kostspieliger ist als das Auslesen der meist sequentiellen KDS-Dateien (THEOBALD 2001).
- Traditionelle TDS bedingen strenge Restriktionen bei der Datenmodellierung. Das Modellieren einiger Gegebenheiten der realen Welt ist nur über Umwege, Kompromisse oder gar nicht möglich. Zum Beispiel kann die dritte Dimension bei Features wie Überwegen oder Tunneln nicht ohne unerwünschte Knotenbildung am Kreuzungspunkt abgebildet werden.
- Traditionelle TDS (und z. T. KDS) weisen keine „offene“ Struktur auf und sind somit prinzipiell nicht für weitere Softwareprodukte zugänglich. Zur Definition des Begriffs „Offenheit“ von Datenmodellen sei auf OGC (2003a & 2003b) verwiesen.

Die ArcGIS Geodatabase verwendet ein neues Konzept zur Implementierung von Topologie (HOEL, MENON & MOREHOUSE 2003; ESRI 2003a, ESRI 2003c). Mit diesem Konzept können einige der Nachteile von traditionellen TDS vermieden werden. Die Geometrien werden Feature-bezogen in der Geodatabase gespeichert, d. h. geteilte Geometrien von topologisch in Zusammenhang stehenden Features werden durchaus doppelt gespeichert. Hierdurch kann der langsamere Bildschirmaufbau umgangen werden. Die Topologie-Maschine benutzt zum Validieren der Topologie sowohl Tabellen, die in der Geodatabase angelegt worden sind, als auch außerhalb der Geodatabase befindliche Programmlogik (HOEL, MENON & MOREHOUSE 2003). Das Modellieren der topologischen Beziehungen ist für den Benutzer wesentlicher flexibler gestaltet worden. Dieser bestimmt bspw. den Zeitpunkt und das geographische Gebiet für welches die Topologie validiert werden soll. Dies bedeutet, dass es möglich ist, Features mit

nicht validierter Topologie in der Datenbank zu verwalten. Weiterhin kann der Benutzer eine Auswahl an Topologie-Regeln treffen, die für seine Daten zur Anwendung kommen sollen. Dies sind nur einige Aspekte zum Topologie-Konzept von ArcGIS 8.3. Näheres kann in der oben genannten Literatur nachgelesen werden und auch in Abschnitt 4.3.3 wird noch einmal auf dieses Thema eingegangen.

Wie bereits erwähnt, sind in wohl allen geographischen Datensätzen implizit topologische Informationen vorhanden, so auch in den Daten der GK 25. Topologische Beziehungen haben stets eine semantische Komponente (UBEDA & EGNHOFER 1997), d. h. eine Linie vom Typ *Strasse* darf z. B. ein Polygon vom Typ *Grünland* kreuzen, ein Polygon vom Typ *Haus* jedoch nicht. Bei der Datenmodellierung ist die Grenze zwischen Objekt und Eigenschaft fließend und zwar insofern, dass sie von der Sichtweise des Nutzers abhängig ist. Im obigen Beispiel könnte eine Klasse *Landnutzung* implementiert werden, in der die Typen *Haus* und *Grünland* als Attribute abgelegt werden, oder es könnte als wichtig erachtet werden, die Typen gleich zu trennen und zwei Klassen *Haus* und *Grünland* zu implementieren. Diese Entscheidung kann von vielen Dingen abhängen, u. a. von den Möglichkeiten der Software, die topologischen Beziehungen zu überprüfen. Im Fall der Klasse *Landnutzung* müsste eine topologische Validierung auf das Attribut *Typ* und die Geometrien zurückgreifen, im Fall der zwei Klassen *Haus* und *Grünland* wären nur die Geometrien erforderlich. Wie eine topologische Validierung durchgeführt werden kann, ist jedoch abhängig von der GIS-Software. Das bedeutet, dass es bei der konzeptuellen Modellierung von geographischen Daten derzeit unpraktikabel erscheint, nicht die zur Implementierung vorgesehene Software zu beachten. Dies wiederum bedeutet einen Bruch mit dem Paradigma der konzeptuellen Modellierung, welches besagt, dass dieser Schritt der Datenmodellierung Software-unabhängig erfolgen sollte. Natürlich ist es denkbar, konzeptuelle Modelle für Geodaten zu entwickeln, die vollständig unabhängig von der Software sind. Es ergibt sich dann jedoch die Gefahr, dass das Modell soweit abstrahiert ist, dass die Umsetzung in das logische und physische Modell sehr aufwändig wird. Dann wäre das konzeptuelle Modell u. U. mehr von akademischem Wert als von praktischem Nutzen. Ein weiteres Problem hierbei ist, dass die gängigen Werkzeuge zur Datenmodellierung, wie z. B. UML-Klassendiagramme, keine spezielle Notation für topologische Beziehungen bereithalten. Die Visualisierung kann nur über unbefriedigende Umwege erfolgen. In dieser Arbeit werden topologische Beziehungen als binäre Assoziationen modelliert, deren graphische Repräsentation in Abb. 4.3 zu sehen ist. Reflexive Assoziationen sind notwendig, um eine topologische Beziehung zwischen Instanzen derselben Entität zu modellieren (z. B.: Polygone der Entität *Geologische Kartiereinheit* sollen direkt aneinander grenzen, vgl. Klasse C in Abb. 4.3).

Bei den Daten der GK 25 sind zahlreiche topologische Beziehungen denkbar. Beispiele hierfür sind:

- Schichtbänke mit einer bestimmten Stratigraphie sollten im Allgemeinen in Kartiereinheiten liegen, die ebenfalls diese Stratigraphie in der Schichtenabfolge aufweisen.

- Streichrichtungen von End- oder Stauchmoränen sollten in Kartiereinheiten liegen, die bestimmte quartäre Schichten aufweisen.
- Tagebaugrenzen sollten Kartiereinheiten umschließen, die die Genese „y ,Tagebau“ aufweisen.

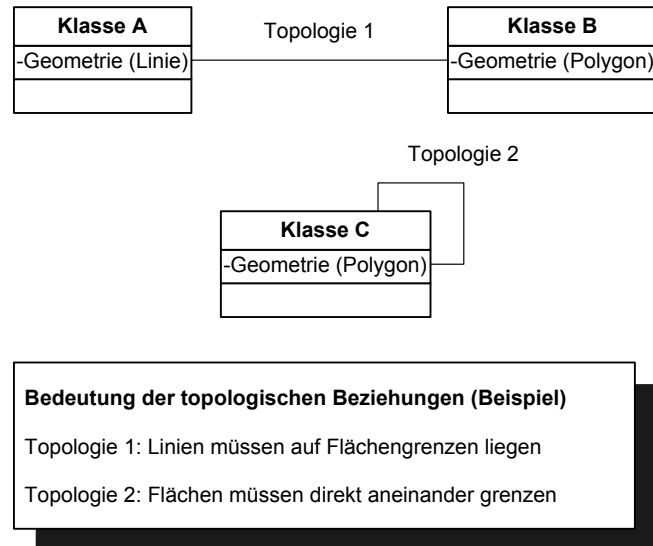


Abb. 4.3: Graphische Repräsentation von topologischen Beziehungen im UML-Klassendiagramm (Erläuterung im Text).

Da diese Beziehungen auf Attribute der Klasse Geologische Schicht zugreifen, können sie im UML-Diagramm nicht als binäre Beziehung dargestellt werden. Eine Visualisierung kann nur erreicht werden, wenn die Kartiereinheiten in entsprechende Subklassen untergliedert werden. Dies würde das Diagramm erheblich komplizierter machen und ist nicht erstrebenswert, wie im Folgenden noch erläutert wird. Die Implementierung dieser Beziehungen in ArcGIS 8.3 würde voraussetzen, dass auf Grundlage der für die jeweilige Beziehung semantisch notwendigen Attribute zumindest entsprechende Subtypen (oder auch eigene Feature Classes, vgl. 4.3.3) der Klasse Geologische Kartiereinheiten gebildet werden. Auch dies würde zu einem extrem komplexen Datenmodell führen, welches in keinem Verhältnis zu Aufwand und Nutzen steht und dem Prinzip der Einfachheit (vgl. Abschnitt 3.4, Punkt 1) widerspricht. Nichtsdestotrotz gibt es einige topologische Beziehungen in der GK 25, die sehr wichtig sind und die sowohl im UML-Diagramm als auch bei der Implementierung Beachtung finden sollen. Diese topologischen Beziehungen sind in Tab. 4.4 aufgeführt. Wichtig sind diese Beziehungen deshalb, weil sie als grundlegend für die Datenintegrität erachtet werden und/oder viele Objekte betreffen. Für die Datenintegrität ist es z. B. sehr entscheidend, dass es keine Lücken oder Überlappungen zwischen den Polygonen der geologischen Kartiereinheiten gibt, denn diese führen nicht nur zu erheblichen Problemen bei der Visualisierung sondern verursachen auch Folgefehler bei allen Verschneidungsoperationen. Auch die Anzahl der betroffenen Objekte (vgl. Anhang A3) wurde bei der Auswahl der topologischen Beziehungen beachtet, denn es ist wichtiger die topo-

Klasse 1	Klasse 2 (bzw. reflexiv)	Bedeutung	UML-Typ
Geol. Kartiereinheit	reflexiv	Polygone müssen aneinandergrenzen (keine Überlappungen oder Lücken)	Topologie 1
Störung mit Flächenbezug	Geol. Kartiereinheit	Linien der Klasse 1 müssen auf Poly- gongrenzen der Klasse 2 liegen	Topologie 2
Erdfall mit Flächenbezug	Geol. Kartiereinheit	Linien der Klasse 1 müssen auf Poly- gongrenzen der Klasse 2 liegen	Topologie 2
Terrasse mit Flächenbezug	Geol. Kartiereinheit	Linien der Klasse 1 müssen auf Poly- gongrenzen der Klasse 2 liegen	Topologie 2
Erdfall	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Terrasse	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Endmoräne	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Stauchmoräne	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Transgression	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Schichtbank	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Gipseinlagerung	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Braunkohleabbau	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Tagebau	Geol. Kartiereinheit	Linien der Klasse 1 müssen auf Poly- gongrenzen der Klasse 2 liegen	Topologie 2
Tagebau	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Mesoz. Grenze	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Kalkige Schicht	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3
Granitkontakthof	reflexiv	Linien der Klasse 1 dürfen sich nicht selbst schneiden	Topologie 3

Tab. 4.4: Wichtige topologische Beziehungen im konzeptuellen Datenmodell der GK 25.

logischen Beziehungen von ca. 38.000 Störungen zu implementieren als die von 26 Streichrichtungen von Endmoränenwällen.

Linienobjekte der Klassen Störung, Erdfall oder Terrassenkante können die Eigenschaft haben, flächentrennend zu sein, d. h. sie trennen zwei geologische Kartiereinheiten. Um diese topologische Eigenschaft im UML-Diagramm zu visualisieren, wurden jeweils die Subklassen mit Flächenbezug und ohne Flächenbezug gebildet.

4.2.3 Kartenrandobjekte

Dem Benutzer einer blattschnittfreien Datenbasis sollte es möglich sein, ein beliebiges Gebiet am Bildschirm auszuwählen (z. B. mit Hilfe eines Auswahl-Rechteckes) und sich diesen Ausschnitt als Kartenblatt mit Legende und weiteren Informationen (z. B. Metadaten, Maßstabsleiste usw.) visualisieren oder ausdrucken zu lassen. Das bedeutet, dass das GIS dynamisch auf die Auswahl der Kartiereinheiten und Linienobjekte reagieren und eine entsprechende Legende zusammenstellen muss. Im UML-Diagramm ist dies folgendermaßen modelliert: Die Auswahl des Gebietes erzeugt ein Objekt der Klasse Kartenausschnitt, welches 1 oder viele Objekte von Geologische Kartiereinheit und 0 oder viele Objekte von Geologisches Linienobjekt enthält. Weiterhin wird ein Objekt der Klasse Legende kreiert, welches sich aus Objekten der Klassen Legeneinheit Fläche (LE Fläche im Diagramm) und, wenn vorhanden, Legeneinheit Linie zusammensetzt. Weiterhin verfügt die Klasse Legende über die Methoden *Gruppieren Flächen*, *Gruppieren Linien* und *Flächensymbolik bilden* sowie *Liniensymbolik bilden*. Die Methoden müssen durch Code implementiert werden und dienen zum einen dazu, die ausgewählten Objekte von Geologische Kartiereinheit, Sonstige Kartiereinheit und Geologisches Linienobjekt nach *KombiNr* bzw. Klassenzugehörigkeit zu gruppieren (Methoden *Gruppieren Flächen* bzw. *Gruppieren Linien*), damit für jede Gruppe nur eine Legeneinheit gebildet wird. Zum anderen sind die Methoden vorgesehen, um die Symbolik der Flächen- bzw. Linieneinheiten nach einer bestimmten Vorgehensweise zu bilden. Die Vorgehensweise zur Bildung der Flächensymbolik richtet sich nach der Schichtenabfolge der geologischen Kartiereinheit und den Attributen der einzelnen Schichten (eine Beschreibung hierzu findet sich im Erläuterungstext der Karte in Anhang A4), d. h. für jede *KombiNr* (bzw. nach dem Typ von Sonstige Kartiereinheit) gibt es eine Darstellungsweise. Bei den Linienobjekten richtet sich die Darstellung nach der Klassenzugehörigkeit und z. T. nach den Attributen.

Neben der Legende sind als Kartenrandobjekte im UML-Diagramm noch die Klassen Metadatenkontainer und Maßstabsleiste aufgeführt. Diese sollen nur als Beispiele verstanden werden, denn es sind selbstverständlich noch andere Objekte denkbar, die am Kartenrand erscheinen können. Auf den Metadatenkontainer soll hier etwas näher eingegangen werden, denn es ist bei den derzeitigen Rahmenbedingungen der GK 25 nicht einfach, ein Konzept für die Verarbeitung von Metadaten zu entwickeln. Als Metadaten (Daten über Daten; vgl. ISO 2003) sollen in diesem Zusammenhang alle Informationen aufgefasst werden, die die GK 25 dokumentieren, so dass Nutzer die geologischen Informationen der GK 25 fachlich korrekt und ohne Unklarheiten für ihre Aufgaben verwenden können. Zu derartigen Metadaten gehören z. B. Informationen über Verantwortlichkeit, Kontaktadresse, Koordinatensystem, Maßstab, Datenaktualität, Datenqualität, Konventionen zur Visualisierung usw. (Näheres über Metadaten für

Geodaten z. B. in ISO 2003; FGDC 1998). Unproblematisch zu handhaben sind Metadaten, die sich auf das gesamte System der GK 25 beziehen, wie z. B. Verantwortlichkeit, Kontaktadresse, Koordinatensystem, Maßstab u. a. Diese Daten können einfach aus der Datenbank ausgelesen (oder auch als vorgefertigtes Textobjekt gespeichert) und am Kartenrand ausgedruckt bzw., bei digitaler Weitergabe der GK 25, in Dateiform weitergegeben werden.

Einige der notwendigen Metadaten sind jedoch schwieriger handhabbar, da sie dynamisch von dem jeweiligen ausgewählten Kartenausschnitt abhängig sind. Diese Daten wiederum besitzen nach wie vor einen Bezug zum Blattschnitt der TK 25, der ja bei der Erstellung der Daten zu Grunde gelegt wurde (vgl. Abschnitt 3.2). So ist es bspw. denkbar, dass ein Nutzer einen Kartenausschnitt auswählt, in dem nebeneinander Daten aus Kartenblättern dargestellt sind, die 1910, 1950 und 1990 kartiert worden sind. Da die Informationen über die Erstellungsjahre der Karten (und andere Blattschnitt-bezogene Informationen, wie z. B. Autor oder Revision der Karte) für den Nutzer von großer Wichtigkeit sind, müssen sie ihm auch zur Verfügung gestellt werden. Im UML-Diagramm ist die Beziehungen zwischen Metadaten und Kartenausschnitt durch die 1:1-Beziehung von Metadatenkontainer zu Kartenausschnitt dargestellt. Die Klasse Metadatenkontainer ist zusammengesetzt aus mehreren Objekten der Klasse Metadatum und verfügt über die Methode *Get_Metadaten*, um die notwendigen Metadaten zu generieren. Die abstrakte Klasse Metadatum vererbt die Methode *Return_Metadatum* an alle abgeleiteten Klassen. Diese Methode muss überprüfen, ob das jeweilige Metadatum einen Bezug zum Blattschnitt der TK 25 hat oder nicht (im UML-Diagramm symbolisiert durch die Assoziation zwischen Metadatum und TK25 Blatt). Wenn ein Bezug vorhanden ist, wird die Information aus einer blattschnittbezogenen Tabelle ausgelesen; wenn nicht, aus einer Tabelle für allgemeine Informationen über die GK 25. Dies ist natürlich nicht das einzig denkbare Konzept zur Lösung der Metadaten-Problematik, doch es sollte dazu dienen, die Problematik zu veranschaulichen.

Wie bereits erwähnt (vgl. Abschnitt 1.1 und 4.2), wurden die in diesem Abschnitt vorgestellten Konzepte bis zur Fertigstellung der Arbeit noch nicht implementiert und sollen als Anregung verstanden werden (siehe auch die Diskussion in Abschnitt 5.1).

4.3 Logisches und physisches Modell

In diesem Abschnitt wird der Schritt vom konzeptuellen zum physischen Datenmodell vollzogen. Der Schwerpunkt liegt dabei auf der Modellierung der Geometriedaten und ihren topologischen Beziehungen (Abschnitt 4.3.2 und 4.3.3). Wie in Abschnitt 4 erläutert, ist es strenggenommen nur für die Sachdaten sinnvoll, ein logisches Datenmodell zu entwickeln. Auf dieses wird in Abschnitt 4.3.4 eingegangen. Im Abschnitt 4.4 werden einige wichtige Aspekte zur Implementierung des Datenmodells und zur Datenmigration erläutert. Da in diesem Kapitel die Terminologie der Geodatabase von ESRI Verwendung findet, werden in Abschnitt 4.3.1 zunächst die in dieser Arbeit verwendeten Konzepte der Geodatabase vorgestellt (wie in Abschnitt

1.3 erläutert, wird dabei in den meisten Fällen die amerikanische Original-Terminologie verwendet).

Bei der physischen Modellierung besteht die Aufgabe nicht nur darin, das konzeptuelle Modell in ein zur ausgewählten Software kompatibles Datenmodell umzusetzen, sondern es müssen auch weitere von der Software (und u. U. Hardware) vorgegebene Rahmenbedingungen beachtet werden. Bei relationalen Datenbanken können z. B. bewusst Normalisierungsregeln verletzt werden, um die Performance des Systems zu verbessern. Auch für die Geodatabase wurde in Abschnitt 4.3.1 bereits ein Aspekt genannt, der starken Einfluss auf die Performance hat, nämlich die Anzahl von Feature Classes in einem Feature Dataset. Oft müssen bei der Modellierung Kompromisse zwischen einer „sauberen“ Datenmodellierung und weiteren systembedingten Zwängen geschlossen werden. Auch hier ist der Umgang mit geographischen Daten zumeist komplizierter als der mit „herkömmlichen“ Daten. Geographische Daten leben davon, als Karten visualisiert zu werden. Die Möglichkeiten, die der Nutzer zur Visualisierung seiner Daten hat, werden wiederum vom GIS vorgegeben, d. h. auch diesbezügliche Aspekte sollten bei der physischen Modellierung berücksichtigt werden. Ein Beispiel aus den Daten der GK 25 soll dies verdeutlichen: Im konzeptuellen Modell wurden alle Linien, die geologische Störungen repräsentieren, als Entität *Störung* modelliert (vgl. Anhang B1). Auf Grund der Eigenschaften, die eine bestimmte Störung besitzen kann, wurde die Entität mit verschiedenen Attributen, wie *vermutet*, *Haupt-/Nebenstörung*, *mineralisiert* usw. versehen. Im Hinblick auf die Zielsetzung bei der konzeptuellen Modellierung – ein logisch nachvollziehbares Modell eines Ausschnitts der realen Welt zu schaffen – ist gegen diesen Entwurf nichts einzuwenden. Probleme ergeben sich jedoch, wenn dieses Modell im GIS implementiert wird und die Störungen gemäß ihrer Attribute dargestellt werden sollen, denn die Darstellung soll anhand mehrerer Attribute vorgenommen werden. Zumindest ArcGIS 8.3 bietet jedoch als standardmäßige Funktionalität nur die Visualisierung anhand eines Attributes an. (Es gibt auch eine Darstellungsform anhand von zwei Attributen, wobei jedoch eines davon quantitativer Natur sein muss). Es präsentieren sich zwei Möglichkeiten, dieses Problem zu lösen, nämlich einerseits die Visualisierung über speziellen Programmcode zu implementieren, was relativ aufwändig ist, und andererseits, die Klassifizierung der Störungen anhand der darzustellenden Attribute.

Letzterer Weg würde bedeuten, dass die Klassifizierung der Störungen über Subtypen der Feature Class *Störungen* implementiert werden müsste. So entstünde z. B. ein Subtyp, der alle Störungen beinhaltet, die als vermutete, mineralisierte Hauptstörung kartiert worden sind. Ein Nachteil dieser Vorgehensweise ist, dass sich nun verschiedene Attribute kodiert hinter einer Kennzahl verbergen, was bedeutet, dass sie nicht mehr ohne weiteres abzufragen sind. Die Abfrage „*Selektiere alle vermuteten Störungen*“ kann nun nicht über das Attribut *Vermutet* geschehen, sondern muss über die Kennzahl der Klassifizierung ausgeführt werden. In diesem Fall muss zunächst ermittelt werden, welche Kennzahlen eine vermutete Störung enthalten. Um dieses Problem zu umgehen, wäre es denkbar, neben der Klassifizierungs-Kennzahl auch die Attribute in der Feature Class zu führen. Dies wiederum würde eine redundante Datenhaltung zur Folge haben, die zu einer inkonsistenten Datenbasis führen kann. Dieses Problem könnte dadurch gelöst werden, dass die Klassifizierung automatisch aktualisiert wird, wenn die ihr zu

Grunde liegenden Attribute geändert werden. Wie diese Fragestellungen in dieser Arbeit tatsächlich behandelt worden sind, zeigen die nächsten Abschnitte. An dieser Stelle sollte lediglich deutlich gemacht werden, dass die Datenmodellierung von Geodaten in vielen Fällen stark an dem internen Datenmodell der GIS-Software ausgerichtet werden muss.

4.3.1 Exkurs: Die Geodatabase von ArcGis 8.3

Die GEODATABASE (GDB) von ESRI ist ein mit ArcGIS 8.x eingeführtes generisches Datenmodell zur Speicherung von Geodaten. Es ermöglicht die Speicherung von Geodaten unter Verwendung von Technologien, wie sie bei kommerziellen (objekt-) relationalen Datenbanken verwendet werden. Ab der Version 9.0 von ArcGIS soll das Datenmodell der GDB vollständig offengelegt werden (vgl. <http://esri.com/software/arcgis/arcgis90.html>, Stand 24.02.04).

Es lassen sich zwei Typen von GDB unterscheiden: die Personal Geodatabase (PGDB) und die ArcSDE Geodatabase (SDE GDB). Die PGDB verwendet zur Speicherung das Microsoft Access MDB-Format und ist nicht für den Multi-User Zugriff geeignet. Das bedeutet, dass sie im Allgemeinen auf dem lokalen Rechner gespeichert wird. Die Datenmenge, die mit einer GDB verwaltet werden kann, ist – wie auch bei Access – auf ca. 2 Gigabyte limitiert. Die SDE GDB wird serverseitig installiert, ist für den Multi-User Zugriff geeignet und besitzt theoretisch keine Limitierung der speicherbaren Datenmenge. Die SDE GDB wird in einem kommerziellen (objekt-) relationalen Datenbankmanagementsystem (DBMS), wie Oracle, Microsoft SQL Server, IBM DB2 oder Informix, implementiert, wobei die Software ArcSDE als Middleware zwischen DBMS und ArcGIS fungiert.

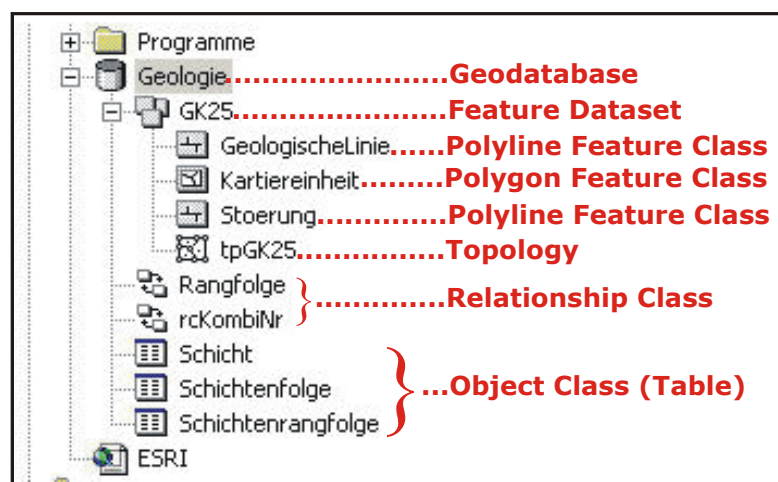


Abb. 4.4: Blick auf eine Geodatabase in ArcCatalog.

Innerhalb einer GDB werden geographische Objekte (Features nach ESRI-Terminologie) in FEATURE CLASSES gespeichert (vgl. Abb. 4.4). In einer Feature Class werden Objekte desselben Geometrietyps (Punkt, Polylinie, Polygon) zusammengefasst. Auch Attribute der Features, die in einer 1:1-Beziehung zum Feature stehen, können in der Feature Class gespeichert werden.

chert werden. Es empfiehlt sich insbesondere Attribute, die zur Visualisierung der Features verwendet werden sollen, direkt in der Feature Class (und nicht in verknüpften Tabellen) abzuliegen, denn dadurch kann eine bessere Performance bei der Visualisierung erzielt werden. Mit Hilfe eines geeigneten Attributes in der Feature Class können SUBTYPES angelegt werden. Subtypes sind eine Möglichkeit, ähnliche Features (mit denselben Attributen) feiner zu untergliedern, um ihre Eigenschaften und ihr Verhalten besser zu steuern. Subtypes können z. B. über unterschiedliche Anfangswerte, über Attribute Domains bei den Attributen oder über unterschiedliche topologische Beziehungen verfügen (zu Attribute Domains und Topologie siehe weiter unten).

Es ist möglich, Feature Classes isoliert in der GDB zu speichern, dann Standalone Feature Classes genannt, oder sie in einem FEATURE DATASET zu gruppieren. Ein Feature Dataset ist vergleichbar mit einem Ordner, in dem zusammengehörige Feature Classes und andere Objekte der GDB (siehe weiter unten) aggregiert werden können. Für die Modellierung von topologischen Beziehungen ist es sogar unabdingbar, die Feature Classes, welche an der Topologie teilnehmen sollen, in einem Feature Dataset zu gruppieren. Ein Feature Dataset verfügt über Eigenschaften, die vom Nutzer definiert werden müssen, und die für alle Feature Classes in diesem Feature Dataset gelten. Hierzu gehören das Koordinatensystem und die Ausdehnung der Koordinaten in X- und Y-Richtung (X/Y Domain) sowie die Präzision (Precision) mit der die Koordinaten gespeichert werden sollen. Das Koordinatensystem kann aus einer vordefinierten Liste ausgewählt oder benutzerdefiniert angegeben werden. Die X/Y Domain ist ein durch maximale und minimale X- und Y-Koordinaten gekennzeichnetes Rechteck. Die Präzision ist die Genauigkeit der Koordinatenspeicherung ausgedrückt in 1/ Präzision Maßeinheiten des Koordinatensystems. Eine Präzision von 100 bei Gauss-Krüger Koordinaten bedeutet also eine Speichergenauigkeit der Koordinaten von 1 cm. X/Y Domain und Präzision müssen angegeben werden, da ArcGIS die Koordinaten intern in positiven 32-Bit-Integer Zahlen speichert. Dies bedeutet, dass zur Speicherung der Koordinaten ein festgelegter Wertebereich zur Verfügung steht, nämlich 31 Bit (entspricht 2^{31} Zahlenwerten). Somit können bei der Anlage eines Feature Datasets entweder die minimalen und maximalen Koordinaten angegeben werden und die Präzision wird von ArcGIS berechnet oder es werden die minimalen (bzw. maximalen) Koordinaten und die Präzision angegeben, wobei ArcGIS dann die maximalen (bzw. minimalen) Koordinaten berechnet. Für Niedersachsen mit einer maximalen Ausdehnung in X-Richtung von ca. 350 km ergibt sich somit eine maximale Speichergenauigkeit der Koordinaten von 0,16 mm (Präzision ca. 6136). Weiterhin ist zu beachten, dass in einem Feature Dataset nicht zu viele Feature Classes gespeichert werden sollten, denn dies wirkt sich nachteilig auf die Performance beim ersten Öffnen einer zu einem Feature Dataset gehörenden Feature Class aus. Der Grund hierfür liegt darin, dass ArcGIS beim Öffnen alle Feature Classes eines Feature Dataset öffnet.

Weitere Objekte, die in einer GDB gespeichert werden können, sind Attribute Domains, Object Classes und Relationship Classes. ATTRIBUTE DOMAINS sind festgelegte Wertebereiche, die für einen bestimmten Datentyp gültig sind und mit einer (oder mehreren) entsprechenden Spalte(n) in einer Feature Class oder Object Class assoziiert werden können. Sie helfen dabei, die Datenintegrität zu verbessern. Es werden *Range Domains* und *Coded Value Domains* unter-

schieden. Range Domains definieren einen fließenden Wertebereich (z. B. 0 bis 100) für numerische Attribute und Coded Value Domains liefern eine Auswahl von vordefinierten Werten und sind für alle Datentypen von ArcGIS einsetzbar.

OBJECT CLASSES speichern Objekte ohne räumlichen Bezug und sind vergleichbar mit den Relationen in einem relationalen DBMS. Sie dienen z. B. dazu, Attribute von Features zu speichern, die nicht direkt in der Feature Class gespeichert werden (können). RELATIONSHIP CLASSES dienen dazu, Verknüpfungen zwischen zwei verschiedenen Feature Classes, Object Classes oder zwischen einer Feature Class und einer Object Class herzustellen. Wie im relationalen Datenmodell können 1:1- und 1:n-Beziehungen abgebildet werden. N:m-Beziehungen müssen in zwei 1:n-Beziehungen und eine Verknüpfungstabelle aufgelöst werden. Die Verknüpfungstabelle kann weitere Attribute enthalten, die die Beziehung charakterisieren und wird dann *Attributed Relationship Class* genannt. Bei Bedarf kann jede Beziehung als *Attributed Relationship Class* implementiert werden.

Topologische Beziehungen zwischen Feature Classes werden in der GDB in einem separaten Objekt, der TOPOLOGY (im folgenden Topologie) gespeichert. Alle Feature Classes, deren topologische Beziehungen zueinander in einer Topologie modelliert werden sollen, müssen in demselben Feature Dataset vorliegen (und weisen daher dasselbe räumliche Bezugssystem auf). In einem Feature Dataset können mehrere Topologien existieren; jede Feature Class darf aber nur an einer Topologie beteiligt sein. Nachdem eine Topologie angelegt worden ist (siehe unten), muss sie validiert werden. Der Nutzer bestimmt den Zeitpunkt der Validierung und das räumliche Gebiet, welches validiert werden soll (d. h. es muss nicht die gesamte Ausdehnung des Feature Datasets auf einmal validiert werden). Während der Validierung wird überprüft, ob die Geometrien der Features mit den benutzerdefinierten Eigenschaften der Topologie korrespondieren, wobei es zu Veränderungen der Geometrien kommen kann. Das Ausmaß der irreversiblen Veränderungen ist ebenfalls von den Eigenschaften der Topologie abhängig. Zu diesen Eigenschaften gehören, neben den beteiligten Feature Classes, die *Cluster Toleranz* (Cluster Tolerance), *Ranks* für die Feature Classes (eine Rangfolge der Feature Classes) und die zu verwendenden *Topologie-Regeln* (Topology Rules).

Die Cluster Toleranz ist z. T. vergleichbar mit der Fuzzy Toleranz von Arc/Info (vgl. Abschnitt 3.2.2.1). Laut ESRI (2003a; vgl. S. 112) definiert die Cluster Toleranz den minimal möglichen Abstand zwischen Vertizes in Karteneinheiten. Vertizes, deren Abstand kleiner ist als die Cluster Toleranz werden als übereinstimmend betrachtet und während des Validierungs-Prozesses zusammengezogen (*snapped* in ArcGIS-Terminologie), wobei die neue Position des resultierenden Vertizes nach bestimmten Regeln festgelegt wird. Der Validierungs-Prozess bedeutet also nicht nur eine Überprüfung und ggf. Veränderung der gemeinsam benutzten Geometrien von Features, sondern auch eine Generalisierung aller beteiligten Geometrien gemäß der eingestellten Cluster Toleranz. Während der praktischen Arbeiten hat sich jedoch gezeigt, dass die oben zitierte Definition der Cluster Toleranz irreführend ist, denn bei den Daten der GK 25 wurden bei der Validierung auch Vertizes zusammengezogen, die weiter voneinander entfernt waren als die Cluster Toleranz. Dies ist einerseits begründet durch einen im nächsten Absatz beschriebenen Prozess, der erst in ESRI (2004a) beschrieben wurde. (Anm: ESRI (2004a)

ist erst nach Beendigung der praktischen Arbeiten erschienen.) Andererseits liegt dies daran, dass der minimal mögliche Abstand zwischen Vertices nicht die Cluster Toleranz sondern die Cluster Toleranz mal $\sqrt{2}$ ist (siehe auch Abschnitt 4.3.3). Dies konnte nur durch eine Kontaktaufnahme zu ESRI in Erfahrung gebracht werden.

Die an der Topologie beteiligten Feature Classes können mit einer Rangfolge versehen werden. Die Rangplätze (Ranks) geben an, wie zuverlässig die Koordinatenangaben der jeweiligen Feature Class im Vergleich zu den anderen Feature Classes ist. Sind die Abstände zwischen Vertices bei der Validierung kleiner als die Cluster Toleranz, werden Vertices mit einem niedrigerem Rangplatz auf die Positionen von Vertices mit einem höheren Rangplatz gesnappt, d. h. die Vertices mit dem höheren Rangplatz verändern ihre Lage nicht. Sind die Rangplätze der Vertices gleich, wird die Lage des jeweils resultierenden Vertex durch geometrische Mittelung bestimmt. Nach ESRI 2004a ist in diesem Fall weiterhin zu beachten, dass beide ursprünglichen Vertices auf den Mittelpunkt gezogen werden, wenn sich dieser innerhalb der Cluster Toleranz befindet. Das heißt, dass in diesem Fall Vertices zusammengezogen werden, die ursprünglich eine Entfernung voneinander aufwiesen, die der Cluster Toleranz mal 2 (mal $\sqrt{2}$) entsprach. Eine Tatsache, die nach der oben zitierten Definition der Cluster Toleranz nicht unbedingt vorherzusehen ist.

Die Topologie-Regeln geben an, welche Art von topologischen Beziehungen zwischen den Features einer Feature Class oder zwischen den Features verschiedener Feature Classes modelliert werden sollen. ArcGIS 8.3 bietet insgesamt 25 verschiedene Topologie-Regeln an (zu den einzelnen Regeln siehe die unten angegebene Literatur), die vom Nutzer nach Bedarf zu einem für die Topologie gültigen Regel-Katalog zusammengestellt werden können.

Wie bereits beschrieben, wird während der Validierung der Topologie überprüft, ob die Geometrien der betroffenen Features den benutzerdefinierten Regeln der Topologie entsprechen. Wenn dies nicht der Fall ist, werden die Geometrien automatisch korrigiert, soweit es die Eigenschaften der Topologie zulassen. Wenn die Regeln verletzt werden und die Geometrien der betroffenen Features nicht automatisch korrigiert werden können, werden Topologie-Fehler (Topology Errors) generiert. Der Nutzer kann sich diese Fehler nach der Validierung anzeigen lassen und hat die Möglichkeit, die Geometrien entsprechend zu korrigieren oder die Geometrien unverändert zu lassen und die Fehler als Ausnahmen (exceptions) zu markieren. Es ist also möglich, bewusst Features in der ArcGIS-Topologie zu verwalten, die gegen die Topologie-Regeln verstoßen. Insgesamt zeigt sich, dass das Topologie-Konzept von ArcGIS 8.3 für den Nutzer eine deutlich flexiblere Datenmodellierung zulässt als das von Arc/Info. Näheres zur Topologie von ArcGIS 8.3 kann unter ESRI (2003a), ESRI (2003c), HOEL, MENON & MOREHOUSE (2003) sowie der ArcGIS Online-Hilfe nachgelesen werden.

4.3.2 Geometriedaten

Das konzeptuelle Datenmodell der Geometriedaten der GK 25 ist in Anhang B1 als UML-Klassendiagramm dargestellt. ESRI unterstützt die Modellierung einer Geodatabase mit UML-

Werkzeugen, d. h. die Diagramme können direkt zur Implementierung der Geodatabase verwendet werden (siehe Abschnitt 4.4.1). Hierzu müssen einige spezielle Konventionen beachtet werden, die in Kapitel 13 von ESRI (2003a) dargelegt sind. Weiterhin können zum Zeitpunkt von ArcGIS 8.3 nicht alle Eigenschaften der Geodatabase und auch keine topologischen Beziehungen mit Hilfe von UML modelliert werden (weiteres hierzu in Abschnitt 4.3.3 und 4.4.1).

Objekte aller Klassen von Geometriedaten können in topologischem Bezug zueinander stehen. Aus diesem Grund müssen alle Feature Classes der GK 25 in einem Feature Dataset (mit dem Namen GK25) gebündelt werden.

Die flächenbeschreibenden Klassen Geologische Kartiereinheit und Sonstige Kartiereinheit des konzeptuellen Modells (s. Anhang B1) werden im physischen Modell in der Polygon Feature Class Kartiereinheit zusammengefasst. Diese Klasse verfügt über das Attribut *KombiNr*, mit dem die Verknüpfung zu den Tabellen der Attributdatenbank erfolgt, in der die Schichtenabfolge und die Schichtbeschreibungen abgelegt sind. Die einzige zur Zeit existierende nicht-geologische Kartiereinheit „Gewässer“ wird über eine spezielle *KombiNr* und die Kürzeln „qh“ (für Holozän) sowie „W“ (für Wasser) unter den Attributen *Stratigraphie* bzw. *Petrographie (Hauptbestandteile)* der Relation Schicht (s. Abschnitt 4.3.4) kodiert. Es ist, u. a. aus topologischen Gründen, nicht praktikabel für nicht-geologische Kartiereinheiten eine eigene Feature Class zu implementieren. Die Darstellung der Kartiereinheiten gemäß der üblichen Konventionen, wie sie bei der Plotausgabe verwendet werden (vgl. Abschnitt 3.2.2.3), ist nur über Programmcode möglich. Theoretisch wäre es zwar denkbar, für jede *KombiNr* ein vorgefertigtes Legendensymbol als Style-Element abzuspeichern, bei derzeit über 13.000 verschiedenen Werten für *KombiNr* ist dies jedoch praktisch kaum durchführbar. Eine schnelle Visualisierung der Kartiereinheiten am Bildschirm (ohne Plotausgabe) kann daher nur in vereinfachter Form ausgeführt werden.

Die geologischen Linienobjekte werden aus Gründen der Übersichtlichkeit und der Datenintegrität in zwei Feature Classes aufgeteilt, nämlich *Stoerung* und *GeologischeLinie*.

Wie aus dem konzeptuellen Modell (siehe Anhang B1) ersichtlich, können sowohl Störungen als auch andere geologische Linienobjekte Attribute aufweisen. Gemäß dem relationalen Modell können diese Attribute als Spalten der jeweiligen Feature Class modelliert werden, wie dies im physischen Datenmodell der Geometriedaten, Variante 1 (Anhang B2) dargestellt ist. Durch die Trennung in zwei Feature Classes kann die Datenintegrität besser gewährleistet werden, denn für geologische Linienobjekte, wie z. B. Kohleflöze, können nicht versehentlich Attribute eingetragen werden, die nur für Störungen Gültigkeit besitzen. Eine weitere Möglichkeit, die Datenintegrität zu verbessern, bietet ArcGIS durch die Attribute Domains (vgl. Abschnitt 4.3.1). Bei der Feature Class *Stoerung* wird die Coded Value Domain *dmJaNein* für viele Attribute verwendet (vgl. Anhang B1). Diese Domain ersetzt den aus verschiedenen Datenbanken bekannten Datentyp Boolean (Ja/Nein-Datenfeld), der von ESRI bei der UML-Modellierung nicht direkt angeboten wird. Für die Modellierung von Haupt- bzw. Nebenstörungen (oder Gängen) findet weiterhin die Coded Value Domain *dmStoerungHierarchie* Anwendung. Diese ermöglicht die ausschließliche Eingabe der Werte *Haupt-* bzw. *Nebenstörung* oder *keine Angabe*. Bei

der Feature Class Geologische Linie stellt sich das Problem, dass nicht alle Subtypen über alle in der Feature Class definierten Attribute verfügen dürfen. Um dieses Problem zu vermeiden, könnten prinzipiell die Linienobjekte in mehreren Feature Classes abgelegt werden, welche jeweils nur die gültigen Attribute aufweisen. Dies hätte jedoch den Nachteil, dass stets mit einer größeren Zahl von Feature Classes gearbeitet werden müsste, von denen einige nur sehr wenige Datensätze enthalten würden. Im vorliegenden Datenmodell ist dieses Problem über die Coded Value Domain dmFixed gelöst, welche nur die Angabe des Wertes *nein* zulässt. Bei Subtypes, die nicht über die Attribute *Vermutet* oder *Unmasstaeblich* verfügen sollen, wird die Domain den Attributen zugeordnet und verhindert somit einen positiven Wert in den Ja/Nein-Attributen.

Die Feature Class Stoerung erhält nur zwei Subtypes (OhneFlaechenbezug und MitFlaechenbezug), die dazu dienen, die topologischen Beziehungen zur Feature Class Kartiereinheit abbilden zu können.

Alle weiteren geologischen Linienobjekte, die nicht zu der Feature Class Stoerung gehören, werden in der Polyline Feature Class GeologischeLinie zusammengefasst, wobei jede in Anhang B1 aufgeführte Subklasse als Subtype modelliert wird. Da auch hier einige Objekte vom Typ Erdfall/Subrosionssenke und Terrassenkante über topologische Beziehungen zu der Polygon Feature Class Kartiereinheit verfügen, müssen entsprechende Subtypes (ohne und mit Flächenbezug) gebildet werden. Die Feature Class GeologischeLinie verfügt über das Attribut Bemerkung (Datentyp String), in dem ergänzende Informationen abgelegt werden können, wie z. B. die stratigraphische Zugehörigkeit der Schichtbänke. Die Modellierung der Linienobjekte als Subtypes ermöglicht es, topologische Regeln für die einzelnen Linienobjekte festzulegen und bei Bedarf in der Zukunft gezielt zu ergänzen bzw. zu ändern (wobei u. U. die in Abschnitt 4.3.1 erläuterten Konsequenzen des Clustering-Prozesses zu beachten sind).

Beide Polyline Feature Classes (Stoerung und GeologischeLinie) erhalten weiterhin das Attribut *TypAlt*, in welchem der Integer-Wert des Linientyps des alten Arc/Info-Datenmodells mitgeführt wird. Dies ist zumindest in der Anfangszeit des neuen Datenmodells sinnvoll, denn so können die Schritte vom neuen zum alten Datenmodell besser zurückverfolgt werden, um mögliche Unklarheiten aufzudecken. Es ist jedoch zu bedenken, dass das Attribut nach erfolgten Editierarbeiten nicht mehr den aktuellen Stand widerspiegeln kann. Dieses Attribut wurde mit der Attribut-Domäne dmLinienTypAlt versehen, welche den Wertebereich der Integer-Werte des Arc/Info-Modells abdeckt. Die Zuordnung der Linientypen des Arc/Info-Modells zu den Feature Classes und Subtypes der Geodatabase ist in Anhang B5 aufgeführt.

Das in Anhang B2 dargestellte Datenmodell hat allerdings einen Nachteil, der dann zum Tragen kommt, wenn Linienobjekte anhand mehrerer unterschiedlicher Attribute visualisiert werden sollen. Beispielsweise sollen bei der Plotausgabe Störungen auf Grund ihrer Attribute *Hierarchie*, *Vermutet* (Ja/Nein) und *Quartaerbedeckung* (Ja/Nein) in unterschiedlicher Strichstärke und mit unterschiedlichen Liniensignaturen dargestellt werden. In ArcGIS 8.3 ist eine Visualisierung anhand *eines* Attributes z. B. über die interaktive Legendenbildung oder über das Anfertigen von benutzerdefinierten Styles, die einem Attribut zugewiesen werden, problemlos möglich. Die Visualisierung anhand mehrerer Attribute ist jedoch nur für ein nominal skaliertes

und ein quantitatives Attribut vorgesehen, was für die vorliegenden Daten nicht zutreffend ist. Das bedeutet, dass das Datenmodell für eine Visualisierung, die auf mehrere Attribute zugreifen soll, noch durch externe Programmlogik ergänzt werden muss. Für die Plotausgabe der GK 25 wäre es z. B. denkbar, die Visualisierung der Linien aufzubauen, indem die Attribute über If/Then-Schleifen abgefragt werden.

Eine Möglichkeit, die Programmierarbeit zu umgehen, wäre ein anderes physisches Datenmodell, wie es z. B. in Anhang B3 dargestellt ist. In dieser zweiten Variante werden die Störungen anhand der in Anhang B1 aufgeführten Attribute kategorisiert und diese Kategorien werden als Subtypes modelliert. Dies entspricht im Prinzip dem Linientypenkatalog des alten Arc/Info-Modells. Die Linientypen können dann individuell über das Subtype-Attribut visualisiert werden. Um die topologischen Beziehungen zwischen den Linienobjekten und den Grenzen der Polygone von Kartiereinheit darstellen zu können, müssen weiterhin für die entsprechenden Objekte zwei Subtypes vorliegen, nämlich jeweils einer für Objekte ohne und mit Flächenbezug. Diese Variante hat jedoch den Nachteil, dass sich mehrere Attribute hinter einem Subtype „verbergen“, was direkte Abfragen auf die Attribute (z. B. *Selektiere alle Störungen unter Quartärbedeckung*) unmöglich macht. Um solche Abfragen durchführen zu können, müsste der Bearbeiter die textlichen Beschreibungen der Linientypen durchgehen und alle entsprechenden Kodenummern auswählen.

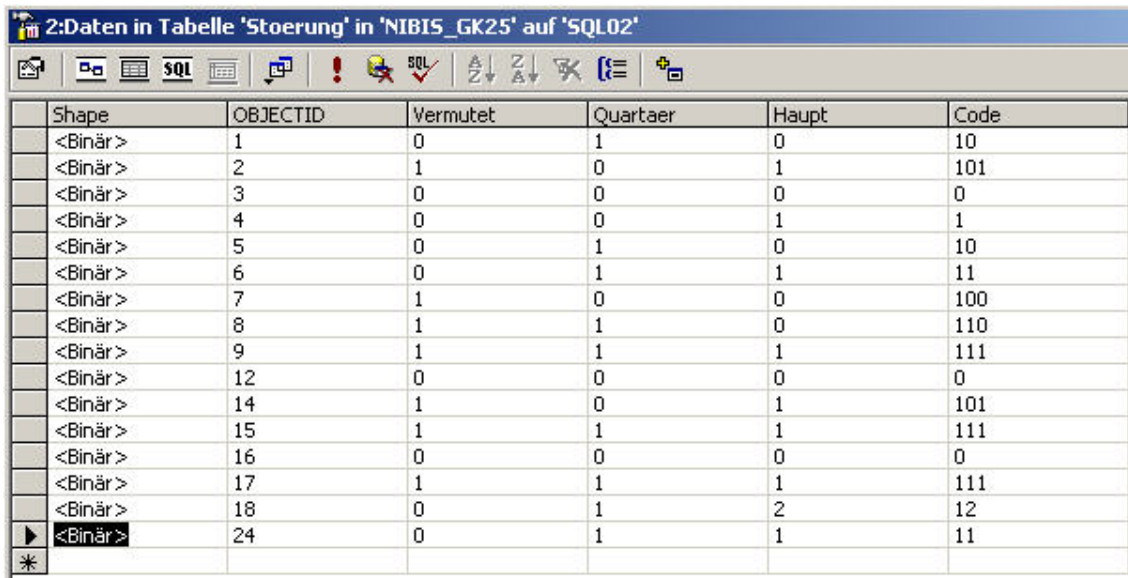
Es wäre weiterhin denkbar, die Vorteile der Varianten 1 und 2 miteinander zu verbinden, indem neben den in der Variante 1 aufgeführten Attributen auch ein Attribut mit einer zur Visualisierung bestimmten Kodierung in den Feature Classes geführt wird. Dies hätte aber eine redundante Datenhaltung mit all ihren Nachteilen zur Folge. Eine Änderung des Attributes *Vermutet* bspw. müsste auch eine Änderung der Visualisierungs-Kodierung nach sich ziehen, was ohne eine dahinterliegende Automatisierung zwangsläufig zu Fehlern führen würde. In einem Standard-DBMS kann für eine derartige Automatisierung eines Updates die Trigger-Technologie eingesetzt werden, für die ArcSDE aber keine alternativen Funktionen bietet. Das heißt, ein entsprechender Trigger müsste auf der Ebene von SQL-Server implementiert werden. Hierbei ist aber darauf zu achten, dass dieser dann nicht auf der Ebene der SDE-Geodatabase „bekannt“ ist. Ein derartiger Trigger könnte folgendermaßen konzipiert werden: Die Feature Class *Stoerung* erhält neben ihren Attributen, wie *Vermutet*, *Quartaerbedeckung*, *Hierarchie* usw., ein weiteres Attribut, in dem die zuerst genannten Attribute kodiert werden. Dieses Attribut wird im folgenden als *Code* (Datentyp Integer) bezeichnet. Der Trigger wird als Update- und Insert-Trigger auf den SDE-Business-Table (vgl. ESRI 1998 und ESRI 2003a) der Feature Class *Stoerungen* gelegt. Der Trigger sorgt für eine Aktualisierung des Attributes *Code*, wenn die Attribute der Feature Class geändert worden sind oder wenn ein neuer Datensatz eingefügt worden ist (vgl. Abb. 4.5). Die SQL-Syntax (T-SQL für SQL-Server 2000) für das vereinfachte Beispiel in Abb. 4.5 SQL-Server kann lauten:

```
CREATE TRIGGER trgUpdateCode ON Stoerung  
FOR INSERT, UPDATE
```

AS

```
Update Stoerung SET Stoerung.Code = CONVERT(varchar, Vermutet) + CONVERT(varchar,
Quartaer) + CONVERT(varchar, Haupt)
```

GO



Shape	OBJECTID	Vermutet	Quartaer	Haupt	Code
<Binär>	1	0	1	0	10
<Binär>	2	1	0	1	101
<Binär>	3	0	0	0	0
<Binär>	4	0	0	1	1
<Binär>	5	0	1	0	10
<Binär>	6	0	1	1	11
<Binär>	7	1	0	0	100
<Binär>	8	1	1	0	110
<Binär>	9	1	1	1	111
<Binär>	12	0	0	0	0
<Binär>	14	1	0	1	101
<Binär>	15	1	1	1	111
<Binär>	16	0	0	0	0
<Binär>	17	1	1	1	111
<Binär>	18	0	1	2	12
<Binär>	24	0	1	1	11

Abb 4.5: Kodierung der Attribute der Feature Class Stoerung im Attribut Code. Das Attribut Code kann zur Visualisierung der Features verwendet werden (vereinfachtes Beispiel; Erläuterung im Text).

Auf diese Weise kann das Attribut *Code* zur Visualisierung benutzt werden, z. B. indem den verschiedenen Werten ein vorgefertigter Linien-Style zugeordnet wird.

Hierbei ist zu beachten, dass der Trigger erst feuert, wenn die Änderungen einer versionierten Feature Class durch entsprechende Befehle in den Business-Table zurückgeschrieben werden. Solange die Feature Class versioniert ist, werden die Features nicht im Business-Table sondern in sogenannten delta-Tables gespeichert, die von ArcSDE verwaltet werden (und auf die praktisch keine Trigger gesetzt werden können; vgl. ESRI 2004b). Es zeigt sich hier, dass das Zusammenspiel zwischen SDE-Geodatabase und dem darunter liegendem DBMS kompliziert ist und darauf aufbauende Anwendungen eingehend getestet werden sollten. Der vorgestellte Trigger konnte in dieser Arbeit wegen technischer Probleme mit der SDE-Datenbank (die außerhalb des Einflussbereichs des Autors liegen) nicht eingehend getestet werden. Sollten entsprechende Trigger für die GK25-Datenbank in Betracht gezogen werden, empfiehlt es sich dieses Thema noch genauer zu untersuchen.

Abschließend soll betont werden, dass prinzipiell der ersten Variante des Datenmodells der Vorzug gewährt werden sollte, da sie nicht, wie die zweite Variante, nur auf den Anwendungsfall der Visualisierung zugeschnitten ist, sondern an grundlegenden Richtlinien des relationalen Schemas orientiert ist.

4.3.3 Topologie

Wie in Abschnitt 4.3.1 dargelegt, muss für eine Topologie in einer Geodatabase eine Cluster Toleranz, eine Rangfolge der beteiligten Feature Classes und eine Auswahl von Topologie-Regeln festgelegt werden. Die Cluster Toleranz sollte gut an die Genauigkeit der Geometriedaten und an den für die Datenbasis gültigen Maßstabbereich angepasst sein. Eine sehr kleine Cluster Toleranz hat den Vorteil, dass bei der Validierung der Topologie die Geometrien der Features kaum in ihrer Lage verändert werden. Ein Nachteil kann jedoch sein, dass bei der Validierung viele Features gegen die Topologie-Regeln verstoßen und ihre Geometrien nicht automatisch angepasst werden können. Dies wiederum bedeutet eine große Anzahl von Topologie-Fehlern, die vom Nutzer gesichtet und mit einer angemessenen Methode korrigiert werden müssen, was einen erheblichen Arbeitsaufwand verursachen kann. Auch bei den Editierarbeiten selber kann eine sehr kleine Cluster Toleranz ineffektiv sein, denn sie hat Auswirkungen auf das Snapping-Verhalten. Bei kleinen Cluster Toleranzen muss dementsprechend genau gearbeitet werden, was u. a. relativ zeitaufwendiges Zoomen mit sich bringt. Zu große Cluster Toleranzen können jedoch noch schwerwiegendere Folgen haben, nämlich dann, wenn das Zusammenziehen von Vertices bei der Validierung zu unerwünschten Ergebnissen führt. Beispielsweise können Geometrien von dicht benachbarten, aber eigentlich getrennten Features verbunden werden, was letztendlich zu einer Zerstörung der geometrischen Datenbasis führen kann. Weiterhin können die Geometrien von Features zu klein sein, um mit der gewählten Cluster Toleranz abgebildet werden zu können. Das bedeutet, dass die Geometrie des Features durch das Zusammenziehen von Vertices bei der Validierung in sich kollabiert. In diesem Fall generiert ArcGIS für das betroffene Feature den Topologie-Fehler *must be larger than cluster tolerance*, der z. B. durch Löschen oder Vergrößern des Features behoben werden kann. ESRI (2003a) empfiehlt, die Cluster Toleranz so zu wählen, dass sie eine Größenordnung kleiner ist, als die Lagegenauigkeit der genauesten Feature Class der Topologie (z. B. eine Cluster Toleranz von 0,1 m bei Daten mit einer Lagegenauigkeit von 1 m).

Für die geologischen Daten der GK 25 lässt sich eine eindeutige Lagegenauigkeit nicht mit Bestimmtheit angeben. Es liegt in der Natur der Daten (vgl. Abschnitt 3.1), dass die Lagegenauigkeit sehr heterogen ist. Anhand von Bohrungen ausgewiesene Grenzverläufe von Kartiereinheiten, die in der Realität u. U. kontinuierlich ineinander übergehen, haben eine Lagegenauigkeit von einigen zehn Metern. Andere Objekte, wie beobachtete Störungen, Gewässergrenzen, Tagebaugrenzen usw. können durchaus eine weitaus höhere räumliche Genauigkeit besitzen, die bis in den Meterbereich reichen kann.

Eine stichprobenartige Überprüfung der Raumdaten der GK 25 hat gezeigt, dass der kleinste Abstand zwischen den Geometrien verschiedener Objekte ca. 2 m beträgt. Das bedeutet, dass die Cluster Toleranz nicht größer als 2 m gewählt werden sollte, da es sonst zu unerwünschtem Snapping zwischen benachbarten Geometrien kommt. Weiterhin haben Testläufe mit verschiedenen Cluster Toleranzen gezeigt (vgl. Abschnitt 4.4.3), dass bei der Polygon Feature Class Kartiereinheit bei der Cluster Toleranz von 1 m der erste *must be larger than cluster tolerance* Fehler auftritt und die Anzahl dieser Fehler bei größeren Cluster Toleranzen kontinuierlich zu-

nimmt (vgl. Tab. 4.7). Aus diesen Betrachtungen heraus lässt sich ableiten, dass es bei den Daten der GK 25 sinnvoll erscheint, mit einer Cluster Toleranz von ca. 0,1 bis 1 m zu arbeiten. Bei den Testläufen mit verschiedenen Cluster Toleranzen wurde weiterhin festgestellt, dass bei einer vorgegebenen Cluster Toleranz von 1 m trotzdem Vertizes zusammengezogen werden, die ursprünglich mehr als 1 m auseinander gelegen haben, was der Definition der Cluster Toleranz (ESRI 2003a, S. 112) widerspricht. Analoge Resultate zeigen sich auch bei der Vorgabe von anderen Werten für die Cluster Toleranz.

Klasse 1	Subtype Klasse 1	Bedeutung	Klasse 2
Geol. Kartiereinheit		Must not overlap	-
		Must not have gaps	-
Stoerung		Must be single part	-
		Must not overlap	-
		Must not self-intersect	-
	MitFlaechenbezug	Must be covered by boundary of	Geol. Kartiereinheit
GeologischeLinie		Must not overlap	-
		Must be single part	-
		Must not self-intersect	-
	Erdfall	Must not intersect	-
	Erdfall_f	Must be covered by boundary of	Geol. Kartiereinheit
		Must not intersect	-
	Terrassenkante	Must not intersect	-
	Terrassenkante_f	Must be covered by boundary of	Geol. Kartiereinheit
		Must not intersect	-
	Mittelterasse	Must not intersect	-
	Moraenen	Must not intersect	-
	Transgression	Must not intersect	-
	Schichtbank	Must not intersect	-
	Gipseinlagerung	Must not intersect	-
	Braunkohleabbau	Must not intersect	-
	Tagebau_f	Must be covered by boundary of	Geol. Kartiereinheit
		Must not intersect	-
	Mesozoische Schicht	Must not intersect	-
	Kalkige Schicht	Must not intersect	-
	Granitkontakthof	Must not intersect	-

Tab. 4.5: *Verwendete Topologie-Regeln im physischen Datenmodell der GK 25 (Variante 1).*

Nachdem der Herstellerfirma dieses Problem berichtet wurde, äußerte sich diese (kurz vor Abschluss der Arbeit) mit dem Hinweis, dass die Cluster Toleranz mal $\sqrt{2}$ der minimal erlaubte Abstand zwischen Vertices sei (vgl. Abschnitt 4.3.1). Da dies zu Zeiten der praktischen Arbeiten noch nicht bekannt war, wurde vorsichtshalber mit einer relativ kleinen Cluster Toleranz von 0,1 m gearbeitet.

Die Topologie-Regeln, welche auf die Feature Classes und Subtypes angewendet wurden, sind Tab. 4.5 zu entnehmen. Für alle beteiligten Feature Classes wurde der Rangplatz 1 vergeben, da es keine Unterschiede zwischen deren Lagegenauigkeiten gibt.

4.3.4 Attributdaten ohne Raumbezug

Bei den Attributdaten ohne Raumbezug handelt es sich um die geologischen Fachdaten, welche die Schichtenfolgen der Kartiereinheiten beschreiben. Dies sind im konzeptuellen Modell die Relationen Schichtenfolge, Schichtenrangfolge und Geologische Schicht (vgl. Abschnitt 4.2.1 und Anhang B1). Im physischen Datenmodell (vgl. Anhang B4) werden die Relationen Schichtenfolge und Geologische Schicht durch die Object Classes (Tabellen ohne Raumbezug) Schichtenfolge bzw. Schicht repräsentiert. Diese, wie auch die folgenden Objekte, werden nicht im Feature Dataset sondern direkt auf der Datenbank-Ebene geführt. Die n:1-Beziehung zwischen der Feature Class Kartiereinheit und der Object Class Schichtenfolge (Schlüsselattribut *KombiNr*) wird in der Geodatabase über eine Relationship Class modelliert. Die n:m-Beziehung zwischen den Object Classes Schichtenfolge und Schicht wird durch eine Attributed Relationship Class beschrieben. Bei der Implementierung sorgt ArcGIS dabei automatisch für die Erstellung der Verknüpfungstabelle Schichtenrangfolge, in der die Abfolge der zu einer Kartiereinheit gehörenden Schichten abgelegt ist. Alternativ könnte diese Beziehung auch durch eine separate Object Class Schichtenrangfolge und zwei n:1-Relationship Classess zu den Object Classes Schichtenfolge bzw. Schicht modelliert werden.

Dieses Datenmodell setzt voraus, dass die Tabellen mit nicht-raumbezogenen Attributdaten in derselben SDE-Datenbank gehalten werden, wie die Geometriedaten. Dies muss nicht zwangsläufig der Fall sein, da es nach wie vor auch andere Wege gibt, die Geometrie- und Attributdaten miteinander zu verknüpfen. Es ist jedoch für einen Bearbeiter im GIS der komfortabelste Weg, diese Daten integriert zu analysieren, zu visualisieren oder zu editieren, denn die Navigation von insbesondere 1:n-verknüpften Datensätzen ist auf diese Weise am einfachsten.

Wie in Abschnitt 4.4.2 gezeigt werden wird, können für die Relationen Schichtenfolge, Schichtenrangfolge und Schicht die bestehenden SQL-Server-Tabellen des Arc/Info-Modells verwendet werden. Es ist nicht notwendig, wie das UML-Diagramm vielleicht vermuten lassen mag, diese Relationen als Object Classes neu zu erstellen und zu befüllen.

4.4 Implementierung des Datenmodells und Datenmigration

4.4.1 Implementierung des Datenmodells mit Hilfe von CASE-Tools

ArcGIS 8.3 bietet die Möglichkeit an, ein Datenmodell für eine Geodatabase mit Hilfe von UML-Diagrammen zu entwerfen, und diese Diagramme dazu zu nutzen, um das Datenbank-Schema in ArcCatalog zu erstellen. Die Funktionalitäten, welche dies ermöglichen, werden unter dem Begriff *Computer Aided Software Engineering Tools* (CASE-Tools) zusammengefasst. Generell sind für diesen Prozess folgende Arbeitsschritte notwendig:

- 1) Entwurf des Datenmodells mit der UML. Hierzu muss Software verwendet werden, mit der es möglich ist, UML-Diagramme zu erstellen. In der vorliegenden Arbeit wurde die Software Visio von Microsoft verwendet. Mit Hilfe von ArcGIS-Templates für Visio und einigen ESRI-spezifischen Konventionen zur Verwendung der UML (siehe Kapitel 13 in ESRI 2003a), kann auf diese Weise ein Geodatabase-Schema entwickelt werden. Nicht alle Aspekte einer Geodatabase, wie z. B. die Topologie, können derzeit mit der UML modelliert werden. Diese Aspekte müssen entweder nachträglich implementiert oder in den Schema-Wizard (siehe unten) eingegeben werden. Es ist auch möglich, das Datenmodell einer bestehenden Geodatabase mit der UML zu ändern und das neue Schema erneut auf die bestehende Geodatabase anzuwenden. Weiterhin bietet ArcGIS einen Semantics-Checker an, mit dem es möglich ist, die meisten Fehler in einem UML-Diagramm aufzudecken.
- 2) Die UML-Diagramme müssen in ein vom Schema-Wizard lesbares Format exportiert werden. In dieser Arbeit wurde dafür der *XML Metadata Interchange* (XMI) Standard benutzt, der es ermöglicht, die UML-Diagramme in einer *Extensible Markup Language* Datei (XML-Datei) zu speichern.
- 3) Der Schema-Wizard in ArcCatalog wird dazu benutzt, die XML-Datei zu lesen und das Datenbank-Schema zu implementieren. In den einzelnen Fenstern des Schema-Wizards können und müssen dabei noch einige Eigenschaften der Geodatabase, wie z. B. das räumliche Bezugssystem von Feature Datasets, eingegeben werden. In dieser Arbeit wird, wie in Abschnitt 3.4 erläutert, das Gauß-Krüger Koordinatensystem (bei ArcGIS unter *Germany Zone 3* oder *DHDN Degree 3 Gauss Zone 3* zu finden) mit einer X/Y Domain mit den Eigenschaften $\text{MinX} = 3000000$, $\text{MinY} = 5000000$ und $\text{Precision} = 100$ (vgl. Abschnitt 4.3.1) verwendet.

Nach Ablauf dieses Prozesses liegt eine leere Struktur für eine Geodatabase vor, die mit Daten befüllt werden kann.

4.4.2 Datenmigration

4.4.2.1 Geometriedaten

Die Coverages der GK 25 wurden zunächst unter Verwendung des Befehls *Merge* des Geoprocessing-Wizards von ArcMap zu einem Coverage zusammengefügt. Danach wurde der *Simple Data Loader* von ArcCatalog dazu verwendet, die Features der Feature Classes der Coverages in die entsprechenden Feature Classes der Geodatabase zu laden. Hierbei ist einerseits darauf zu achten, dass die System-Felder von Arc/Info (z. B. Cover-ID, Area, FNODE etc.) nicht in die Geodatabase überführt werden, da sie bei bestimmten Operationen in ArcGIS zu Fehlern führen können. Andererseits müssen bzw. sollen natürlich die *KombiNr* der Polygone und der Linientyp der Arcs mit überführt werden. Bei den Arcs ist darauf zu achten, dass durch entsprechend formulierte Abfragen im Simple Data Loader die verschiedenen Linientypen in die dafür vorgesehenen Subtypes geladen werden (vgl. Anhang B5).

Nachdem alle Geometrien geladen sind, kann die Topologie mit den in Abschnitt 4.3.3 genannten Einstellungen erstellt und validiert werden. Hierbei zeigt sich, dass in der Datenbasis eine Reihe von Topologie-Fehlern enthalten sind (siehe Tab. 4.6).

Klasse 1: Subtype	Bedeutung	Klasse 2	Anzahl Fehler
Geol. Kartiereinheit	Must not overlap	-	13.453
Geol. Kartiereinheit	Must not have gaps	-	1635
Störung: Flächenbezug	Must be covered by boundary of	Geol. Kartiereinheit	132
Störung	Must not overlap	-	2

Tab. 4.6: Topologie-Fehler nach der Validierung der Topologie.

Versuche mit derselben Datenbasis und denselben Einstellungen bei der Topologie haben gezeigt, dass die Anzahl der Fehler bei verschiedenen Durchläufen nicht konstant bleibt, sondern sich in einem gewissen Rahmen ändern kann. Gründe hierfür sind wohl in den Algorithmen der Topologie-Engine zu suchen, die jedoch nicht offengelegt sind. Die Polygon Feature Class weist eine Vielzahl von *Must not have gaps* und *Must not overlap* Fehlern auf, welche räumlich ausschließlich an den ehemaligen Kartenblatträndern der Coverages angesiedelt sind. Dieser wichtigen Problematik ist der eigene Abschnitt 4.4.3 gewidmet. Bei den Topologie-Fehlern der Polyline Feature Classes handelt es sich nicht um systematische oder auf irgendeine Weise prozessbedingte Fehler, sondern schlicht um fehlerhafte Kodierungen des Linientyps (*Must be covered by boundary of* Fehler) oder um doppelt abgelegte Features (*Must not overlap* Fehler). Diese Fehler müssen individuell gesichtet und korrigiert werden. Die Korrektur der *Must be covered by boundary of* Fehler muss durch eine Neuzuweisung zum Subtype OhneFlächenbezug vollzogen werden und der *Must not overlap* Fehler durch die Löschung eines der doppelten Features. Da die Anzahl der Polyline-Fehler relativ gering ist, bleibt der Arbeitsaufwand überschaubar.

4.4.2.2 Attributdaten ohne Raumbezug

Wie in Abschnitt 4.3.4 erläutert, handelt es sich bei den Attributdaten ohne Raumbezug um die Relationen Schichtenfolge, Schichtenrangfolge und Schicht des physischen Datenmodells in Anhang B4. Diese Relationen sind im Arc/Info-Modell bereits als die SQL-Server Tabellen KombiNrn, SchichtFolgen und Schichten (vgl. Anhang A2) vorhanden und können direkt in die SDE-Datenbank überführt werden. Hierzu müssen die Tabellen in die SDE-Datenbank kopiert und mit dem Befehl *Register with Geodatabase* von ArcCatalog in der SDE-Datenbank „angemeldet“ werden. Damit auch die Beziehungen zwischen den Tabellen in der SDE-Datenbank bekannt sind, müssen weiterhin die entsprechenden 1:n-Relationship Classes gebildet werden. Die weiteren in Anhang A2 aufgeführten Tabellen sind für Verwaltungs- und Visualisierungsaspekte des alten Arc/Info-Systems notwendig (vgl. Abschnitt 3.2.2). Da diese Aufgaben mit der Geodatabase anders wahrgenommen werden müssen, brauchen diese Tabellen nicht in die SDE-Datenbank überführt werden. Eine Ausnahme bilden die Relationen StratPetH-Genese_FarbKat und StratPetHGenese_Typen, welche wahrscheinlich auch im noch zu entwickelnden Visualisierungs-System unter ArcGIS Verwendung finden werden.

4.4.3 Blattrandproblematik

In Abschnitt 4.4.2.1 wurde bereits berichtet, dass nach der Validierung der Topologie sehr viele *Must not have gaps* und *Must not overlap* Fehler, welche sich auf die Polygon Feature Class Kartiereinheit beziehen, vorhanden sind (vgl. Tab. 4.6). Es ist äußerst wünschenswert, diese Fehler aus der Datenbasis zu entfernen, da sie nicht nur kartographisch unbefriedigend sind, sondern auch bei Verschneidungsaktionen stets Folgefehler nach sich ziehen können. ArcGIS 8.3 bietet standardmäßig keine Funktionen an, um diese Fehler in einer Art Bulk-Processing zu beheben. Inwiefern solche Funktionen überhaupt sinnvoll wären, wird weiter unten diskutiert. Die Anzahl der Fehler ist jedenfalls viel zu hoch, um sie einzeln anzusehen und zu korrigieren. Aus diesem Grund wurde nach einer Möglichkeit gesucht, mit der diese Fehler automatisch oder zumindest weitestgehend automatisch behoben oder gänzlich vermieden werden können. Hierbei ist es zunächst zweckdienlich, die Fehler genauer zu charakterisieren und nach ihrer Ursache zu suchen. Wie die Abb. 4.6 und 4.7 zeigen, sind die Fehler ausschließlich an den ehemaligen Blatträndern der Coverages zu finden und über die gesamte räumliche Ausdehnung der Datenbasis zufällig verteilt.

Ein Vergleich von Abb. 4.6 und 4.7, sowie auch die Vergrößerung in Abb. 4.8, machen weiterhin deutlich, dass an einem ehemaligen Blattrand zwischen zwei Coverages sowohl *Must not have gaps* als auch *Must not overlap* Fehler existieren können. Eine stichprobenartige Sichtung dieser Fehler ergab, dass die Positionsungenauigkeiten in den meisten Fällen einige Dezimeter, in einigen Fällen jedoch bis zu 8 m betragen können.

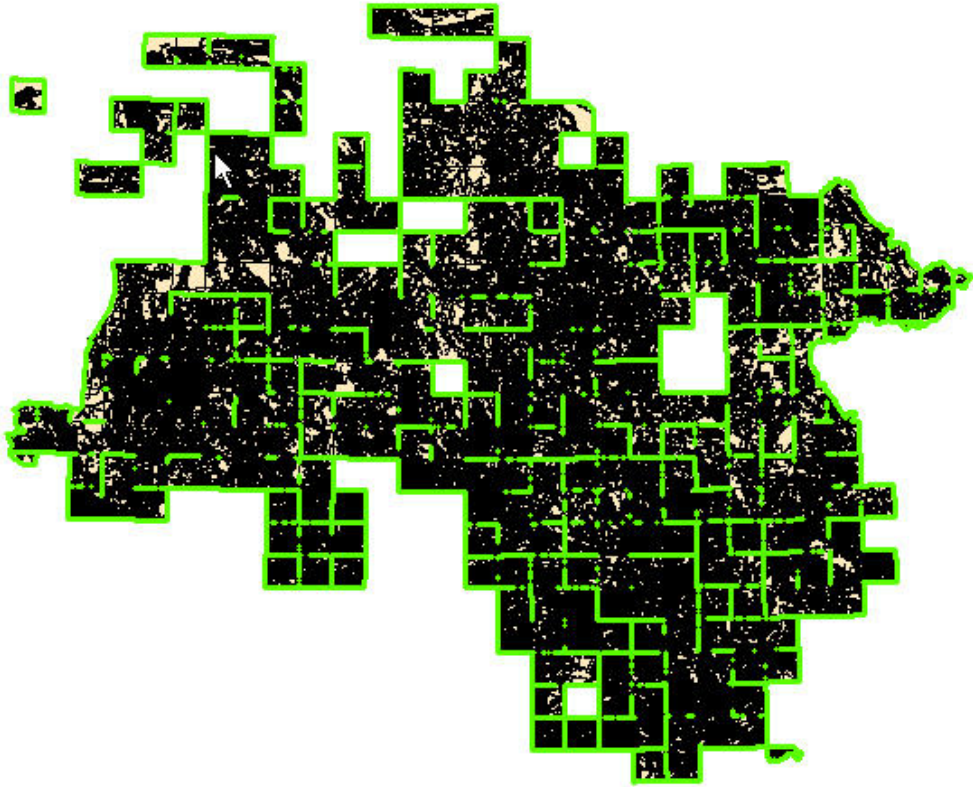


Abb. 4.6: Must-not-have-gaps Fehler nach der Validierung der Topologie.

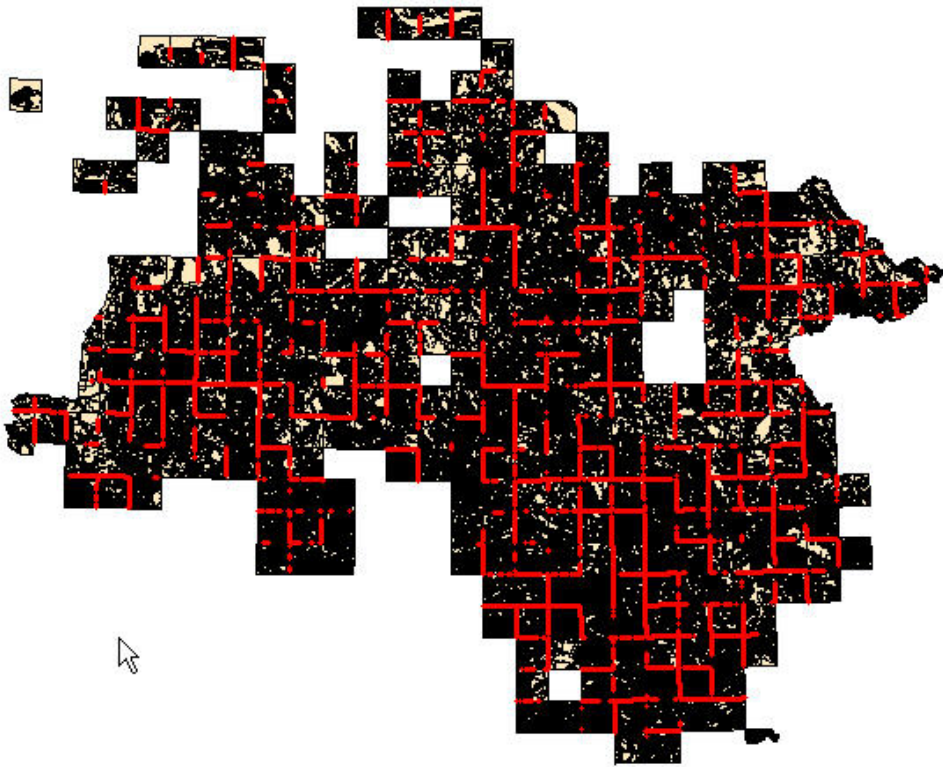


Abb. 4.7: Must-not-overlap Fehler nach der Validierung der Topologie.

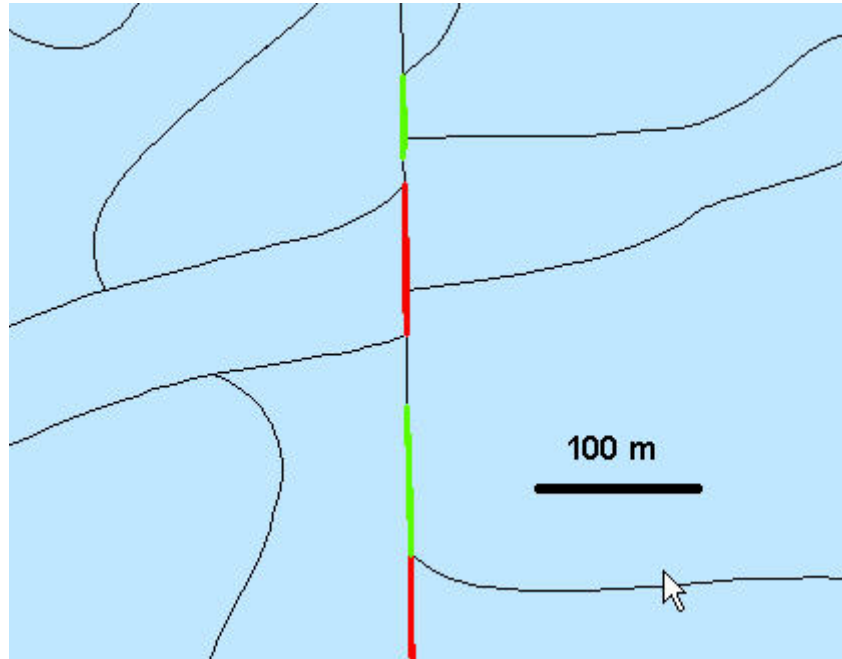


Abb. 4.8: *Must-not-have-gaps (grün) und Must-not-overlap (rot) Fehler am ehemaligen Blatt-
rand zwischen zwei Coverages. (Anm: Die Darstellung der Topologie-Fehler kor-
respondiert bei dieser Zoom-Stufe nicht mit den tatsächlichen Abmessungen der
Fehler).*

Eine Betrachtung der ursprünglichen Coverages beweist, dass diese Fehler bereits in den Coverages vorgelegen haben, d. h. sie sind nicht durch die Migration zur Geodatabase entstanden. Die Fehler waren früher nicht so augenscheinlich, da nur in seltenen Fällen mit mehreren Coverages parallel gearbeitet wurde. Die Fehler sind wahrscheinlich im Wesentlichen auf zwei Aspekte zurückzuführen: Zum einen ist dies eine ungenaue Georeferenzierung. Der Kartenrahmen wurde in den Coverages der GK 25 durch vier TIC-Punkte (vgl. Abschnitt 3.2.2.1) definiert, die durch vier Arcs, gekennzeichnet als Linientyp Kartenrahmen, miteinander verbunden waren. Die Koordinatenangaben für diese TIC-Punkte decken sich nicht immer genau mit den entsprechenden Koordinaten der TIC-Punkte der Nachbarblätter. Zumindest teilweise ist dies wohl auf das unterschiedliche Alter der Coverages zurückzuführen, welches dazu führte, dass bei benachbarten Kartenblättern unterschiedliche Koordinaten zur Georeferenzierung benutzt wurden. Überprüfungen haben weiterhin gezeigt, dass in einigen Fällen auch die Arcs der Kartenrahmen nicht exakt durch die TIC-Punkte verlaufen, was auf ungenaues Digitalisieren des Bearbeiters zurückzuführen ist. Der zweite Aspekt wird durch die Befehle zur Topologie-Bildung unter Arc/Info verursacht (z. B. *Build* oder *Clean* Befehl). Bei deren Ausführung wird überall dort, wo ein Arc auf den Kartenrahmen trifft, ein Knoten gesetzt. Dieser Knoten liegt nicht exakt auf dem Kartenrahmen, denn in Abhängigkeit von der Genauigkeit der Koordinatenspeicherung und der Fuzzy Tolerance (vgl. Abschnitt 3.2.2.1) kommt es zur Verschiebung des Knotens und damit des Kartenrahmens. Der Kartenrahmen ist also nur in Ausnahmefällen eine direkte Verbindung zwischen den TIC-Punkten des Coverages. Im Normalfall entspricht er einer in einem gewissen Rahmen verlaufenden „Zickzack“-Linie.

Grundsätzlich sind mehrere Ansätze denkbar, mit denen die *Must not have gaps* und *Must not overlap* Fehler korrigiert oder von vornherein weitestgehend vermieden werden können. Es könnte versucht werden, schon vor der Migration die ursprünglichen Daten der Coverages mit Funktionen von Arc/Info in ihrer Qualität zu verbessern. Dies würde jedoch bedeuten, dass entweder jedes Coverage einzeln, oder dass der zu einem Coverage zusammengefügte Datenbestand (vgl. Abschnitt 4.4.2.1) insgesamt bearbeitet werden muss. Im ersten Fall lässt sich alleine aus der Anzahl der zu bearbeitenden Coverages ein erheblicher Arbeitsaufwand ableiten. Im zweiten Fall lässt sich kein großer Vorteil zur Bearbeitung in der Geodatabase erkennen. Aus diesen Gründen wurde nach weiteren suchen Methoden Ausschau gehalten.

Eine weitere Möglichkeit, den Topologie-Fehlern zu begegnen, bietet die Cluster Toleranz der Topologie in ArcGIS 8.3 (vgl. Abschnitt 4.3.1). Die Cluster Toleranz, die zu der hohen Zahl von *Must not have gaps* und *Must not overlap* Fehler führte, war mit 0,1 m sehr klein gewählt. Eine größere Cluster Toleranz würde bedeuten, dass Fehler während des Validierungs-Prozesses automatisch korrigiert werden würden. Hierbei ist jedoch zu beachten, dass sich die Cluster Toleranz auf die gesamte Datenbasis auswirkt. Wie in Abschnitt 4.3.1 erläutert, bedeutet dies sowohl eine Generalisierung aller Features als auch eine mögliche Verschiebung der Vertizes um maximal den Wert der Cluster Toleranz mal $\sqrt{2}$. Weiterhin ist abzusehen, dass auf diese Weise nicht alle Fehler behoben werden können, denn die Fehler gehen zurück auf Positionsungenauigkeiten von bis zu 8 m. Wie in Abschnitt 4.3.3 dargelegt, ist jedoch schon ab einer Cluster Toleranz von ca. 2 m damit zu rechnen, dass die Vertizes benachbarter Features zusammengezogen werden würden, was nicht wünschenswert ist. Trotz dieser wenig erfolgversprechenden Voraussetzungen wurden einige Probeläufe mit verschiedenen Cluster Toleranzen gerechnet, um die Auswirkungen auf die Datenbasis besser einschätzen zu können. Die Ergebnisse dieser Probeläufe sind in Tab. 4.7 aufgeführt.

Cluster Toleranz [m]	Art des Topologie-Fehlers			Summe
	must be larger than Cluster T.	must have no gaps	must have no overlaps	
0,1	0	1.592	13.406	14.998
1,0	1	302	1.629	1.932
2,0	5	205	562	772
2,5	12	191	354	557
3,0	25	251	190	466
4,0	133	275	109	517
10,0	6.439	1.460	419	8.318

Tab. 4.7: Vergleich der Topologie-Fehler der Feature Class Kartiereinheit bei Verwendung verschiedener Cluster Toleranzen.

Die Tab. 4.7 zeigt, dass es mit der benutzten Datenbasis möglich ist, die Gesamtzahl der Topologie-Fehler deutlich zu senken. Das Minimum von 466 Fehlern ist bei einer Cluster Toleranz von 3,0 m erreicht ist. Allerdings ist auch ersichtlich, dass ab einer Cluster Toleranz von 1,0 m die Anzahl der Features, deren Geometrie zu klein ist, um dargestellt zu werden (*Must be larger than Cluster Tolerance* Fehler), langsam ansteigt. Dies zeigt deutlich, dass die Datenbasis der GK 25 relativ viele Features enthält, die sehr klein sind. Die Größe der Features, die bei einer Cluster Toleranz von 4 m als *Must be larger than Cluster Tolerance* Fehler ausgewiesen werden, beträgt jeweils bis zu ca. 1700 m² (entspricht einem Kreis von ca. 32 m im Durchmesser). Nach BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE (2002) sollten die kleinsten darzustellenden Flächen einer geologischen Karte im Maßstab 1:25.000 nicht kleiner sein als ca. 1,5 mm (entsprechend 37,5 m in der Natur) im Durchmesser, damit der Leser sie noch gut identifizieren kann. Die Diskrepanz zwischen den kleinsten Flächengrößen in der Datenbasis und den Anforderungen an die Lesbarkeit legt nahe, dass es eine zukünftige Aufgabe sein sollte, die Datenbasis diesbezüglich zu generalisieren (vgl. Abschnitt 5.1). Da es eine Anforderung an diese Arbeit ist, die Datenbasis fachlich nicht zu verändern (vgl. Punkt 2 in Abschnitt 3.4), sollte zum jetzigen Zeitpunkt der Arbeit vermieden werden, Fehler der Art *Must be larger than Cluster Tolerance* zu produzieren. Weiterhin zeigt die Tabelle, dass die *Must not have gaps* Fehler ab einer Cluster Toleranz von 3,0 m wieder ansteigen. Der Grund hierfür ist, dass Vertizes von ursprünglich benachbarten Features zusammengezogen worden sind, wodurch eine vorher nicht vorhandene Lücke (Gap) entstanden ist (vgl. Abb. 4.9). Das bedeutet, dass spätestens bei einer Cluster Toleranz von 3,0 m die Geometrien der Cluster Toleranz zerstört werden. Um derartige Fehler ausschließen zu können, sollte die Cluster Toleranz deutlich kleiner als 3,0 m gewählt werden (vgl. Empfehlung zur Cluster Toleranz in Abschnitt 4.3.3). Dies zeigt, dass allein durch die Wahl der Cluster Toleranz nicht vermieden werden kann, dass eine relativ große Anzahl von *Must not have gaps* und *Must not overlap* Fehlern bei der Validierung ausgewiesen wird. Demzufolge müssen diese Fehler nach der Validierung durch einen möglichst automatisierbaren Prozess behoben werden.

Um eine mögliche Automatisierbarkeit der Behebung von *Must not have gaps* und *Must not overlap* Fehlern zu überprüfen, muss zwischen beiden Arten von Fehlern unterschieden werden. Der *Must not overlap* Fehler bezieht sich stets auf genau zwei Features, welche im *Fix Topology Error Tool* von ArcGIS 8.3 explizit ausgewiesen werden. Das Tool bietet weiterhin die Möglichkeit an, diesen Fehler automatisch zu beheben. Dazu muss eines der beiden Features interaktiv ausgewählt werden, zu welchem dann die Overlap-Fläche hinzugefügt wird (in ArcGIS als *Merge* bezeichnet). Demnach sollte es möglich sein, Programm-Code zu entwickeln, der den *Must not overlap* Fehler im Bulk-Processing korrigiert, in dem die Overlap-Fläche z. B. stets dem größten der betroffenen Polygone zugewiesen wird. Das Resultat hätte allerdings den Nachteil, dass entlang der ehemaligen Blattränder der Coverages „Zickzack“-Linien auftreten würden, da das größere Polygon mal links und mal rechts von dem Blattrand platziert sein kann. In der Plotausgabe wäre dies in den meisten Fällen nicht wahrzunehmen, doch es bleibt kartographisch gesehen eine eher unschöne Maßnahme.

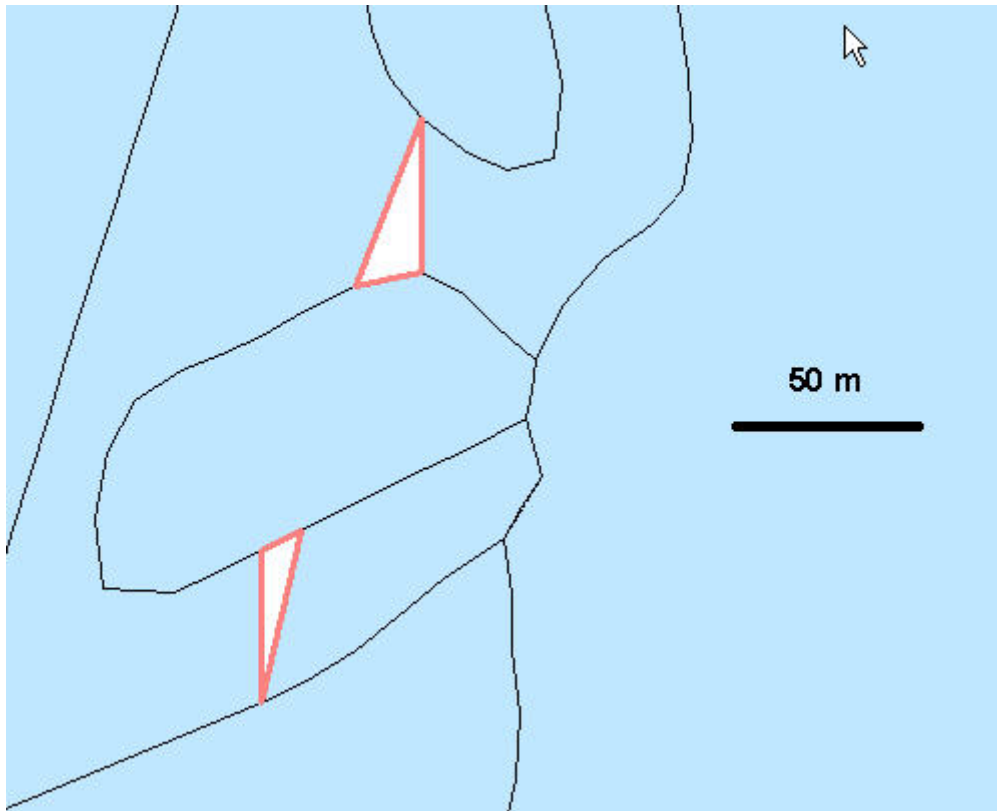


Abb. 4.9: Durch Validierung mit zu großer Cluster Toleranz entstandene Lücken in der Feature Class Kartiereinheit (Gebiet liegt nicht an einem ehemaligen Blattrand).

Für den *Must not have gaps* Fehler bietet sich keine vergleichbare Möglichkeit zum Bulk-Processing an. Bei der Validierung überprüft die Topologie-Engine, ob es in der Ausdehnung des zu überprüfenden Gesamtgebietes Bereiche gibt, die nicht durch ein Polygon abgedeckt sind (der genaue Algorithmus ist nicht offengelegt). Ist dies der Fall, wird ein *Must not have gaps* Fehler ausgewiesen. Das bedeutet, dass ein *Must not have gaps* Fehler sich nicht explizit auf eine bestimmte Anzahl von Features bezieht. Ein Programmierer kann also einen *Must not have gaps* Fehler nicht auf ein oder mehrere verursachende Features zurückverfolgen und so entzieht sich ihm der Ansatzpunkt zur Fehlerbehebung. Weiterhin wäre für jede Lücke (Gap) individuell zu entscheiden, welche Vertizes der betroffenen Features wohin bewegt werden sollen, um die Lücke zu schließen. Diese Entscheidungen können am besten unter visueller Kontrolle eines menschlichen Bearbeiters am Bildschirm getroffen werden. Es scheint mit vertretbarem Aufwand keine Möglichkeit zu geben, diesen Entscheidungsprozess sinnvoll zu automatisieren.

Eine Möglichkeit, die *Must not have gaps* und *Must not overlap* Fehler weitestgehend automatisiert zu beheben, bietet ArcGIS 8.3 mit dem *Zipper Task* genannten Werkzeug (unter den *Developer Samples* zu finden). Mit diesem Werkzeug können alle Vertizes, die sich in einer vom Nutzer zu bestimmenden Entfernung (Zipper Toleranz) zu der Geometrie eines ausgewählten Features befinden, an die Geometrie des ausgewählten Features gezogen werden. Die Idee, die hinter der Verwendung dieses Tools zu einer sinnvollen und möglichst effektiven Bearbeitung der vorliegenden Problematik steht, ist, dass die Polygon-Vertizes, welche sich am

Nächsten zum ehemaligen Blattrand befinden, erneut auf einen neu und exakt definierten Blattrand gezogen werden (vgl. Abb. 4.10).

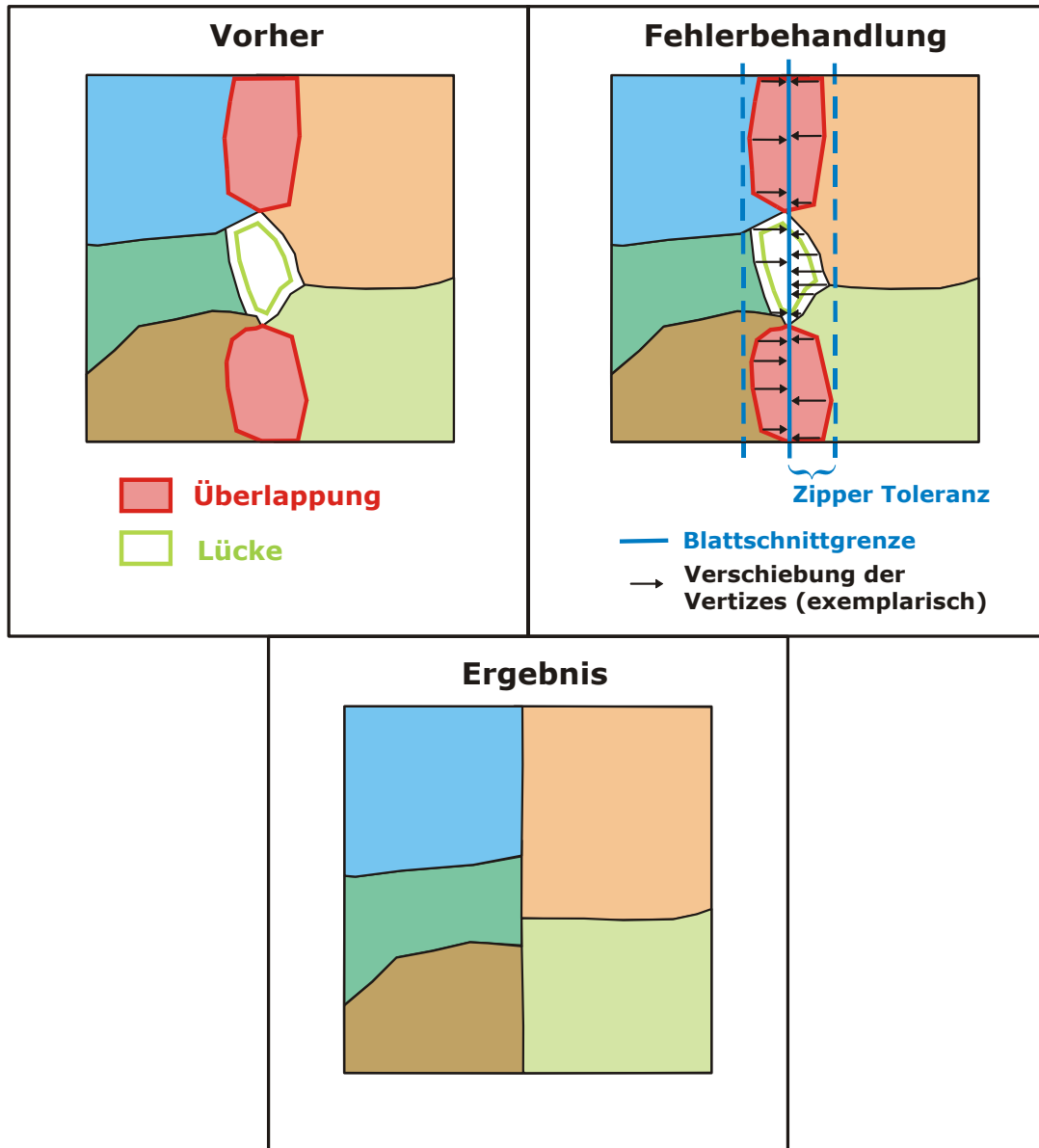


Abb. 4.10: Schema zur Verwendung des ArcGIS-Werkzeugs Zipper Task zur Korrektur der Topologie-Fehler der Feature Class Kartiereinheit.

Hierzu muss zunächst eine Polyline Feature Class geschaffen werden, deren Polylines die exakten Blattecken-Koordinaten (definiert durch den Landesbetrieb Geobasisinformation Niedersachsen in cm-Genauigkeit) der TK 25 Blätter einer bestimmten Breite bzw. Länge von Niedersachsen miteinander verbinden. Diese Polyline Feature Class wird in ArcMap über die Datenbasis der GK 25 Geometriedaten gelegt. Nun werden mit dem Zipper Task Werkzeug nacheinander die Blattrand-Polylinien ausgewählt und mit der definierten Zipper Toleranz werden dann die betroffenen Vertices an die ausgewählte Polylinie herangezogen. Mit einer auf die gesamte

Datenbasis angewendeten Zipper Toleranz von 1,0 m konnte so die Anzahl der *Must not have gaps* und *Must not overlap* Fehler (der Topologie mit der Cluster Toleranz von 0,1 m) um ca. 60 % reduziert werden (vgl. Abb. 4.11). Der Arbeitsaufwand für das Zippen der gesamten Datenbasis mit der 1,0 m Toleranz betrug ca. zwei Stunden (davon reine Rechenzeit ca. eine Stunde).

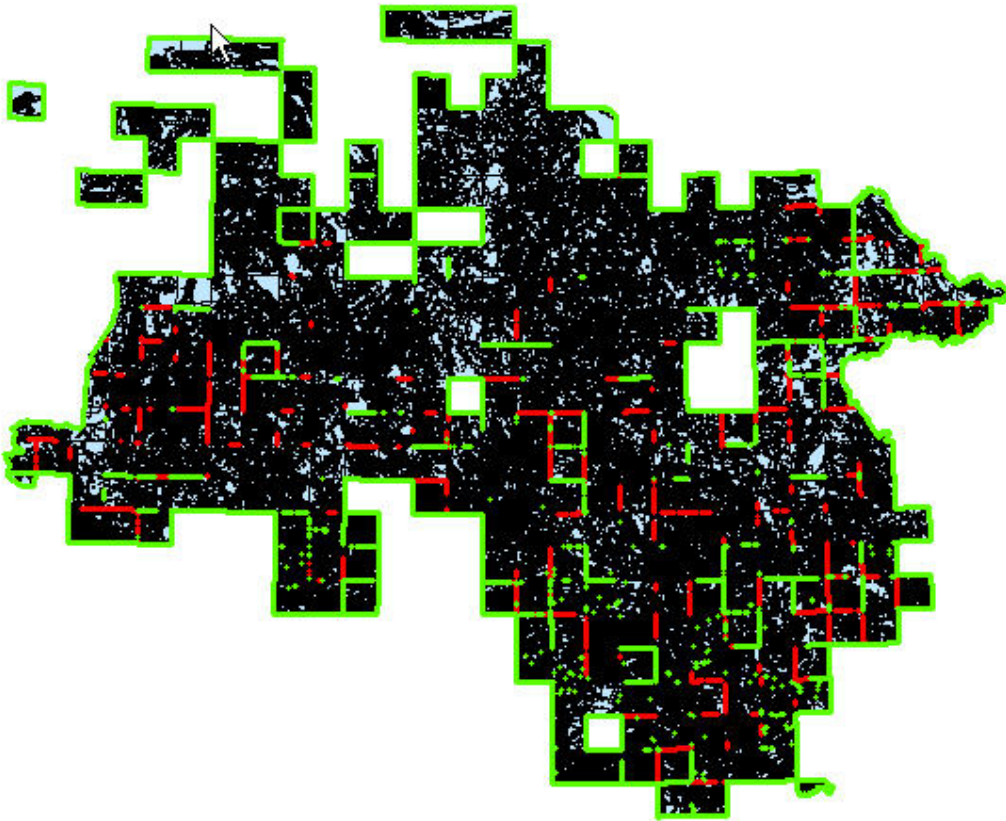


Abb. 4.11: Topologie-Fehler der Feature Class Kartiereinheit nach der Behandlung mit dem Zipper-Task-Werkzeug und einer Zipper-Toleranz von 1,0 m (Reduktion der Fehleranzahl um ca. 60 %; vgl Abb. 4.6 und 4.7).

Bei der Zipper Toleranz von 1,0 m kam es nicht zu unerwünschten Zerstörungen der Geometrien, was jedoch bei größeren Toleranzen vorkommen kann. Daher empfiehlt es sich, nach der Bearbeitung mit der 1,0 m Toleranz räumlich gezielt vorzugehen. Dabei sollten bestimmte geographische Bereiche der Datenbasis mit einer den lokalen Gegebenheiten der Topologie-Fehler angepassten Zipper Toleranz bearbeitet werden. Nach dem Zippen sollte dieser Bereich validiert werden, wobei sich zerstörte Geometrien durch nicht am Kartenrand liegende *Must not have gaps* und *Must not overlap* Fehler aufdecken lassen. Sollten zerstörte Geometrien auftreten, kann die Zipper-Aktion rückgängig gemacht und ein neuer Versuch mit einer geringeren Zipper Toleranz durchgeführt werden. Auf diese Weise kann der überwiegende Teil der *Must not have gaps* und *Must not overlap* Fehler in einer Art schrittweisem Bulk-Processing effektiv korrigiert werden und nur für komplizierte Einzelfälle ist eine individuelle Behandlung der Fehler notwendig. Der gesamte Zeitaufwand für die Korrekturarbeiten wird vom Autor auf

wenige Arbeitstage geschätzt. Nach Vollendung der Korrekturarbeiten kann die Polyline Feature Class mit dem Kartenrahmen gelöscht werden und es liegt eine topologisch fehlerfreie und kartographisch „saubere“ Datenbasis vor. Die ehemaligen Blattränder sind dann natürlich noch als solche zu erkennen, doch ist dies nunmehr ein fachliches und kein technisches Problem (vgl. Abschnitt 5.1). Diese Ausführungen machen deutlich, dass das Zipper Task Werkzeug zur Korrektur der Topologie-Fehler effektiv eingesetzt werden kann. Auch in der Literatur finden sich Hinweise darauf, dass die dem Konzept des Zipper Task Werkzeuges entsprechenden Methoden zur Korrektur von Blattrand-Fehlern viele Vorteile gegenüber anderen Methoden aufweisen (vgl. CHRISMAN 1990).

5 Diskussion und Schlussfolgerungen

Dieses Kapitel soll dazu dienen, die Arbeitsergebnisse abschließend zu bewerten und zu diskutieren (Abschnitt 5.1.1) sowie einen Ausblick auf die weiteren Arbeitsschritte, die zur Erreichung einer vollständigen Umstellung vom dualen zum integrierten System notwendig sind, zu eröffnen (Abschnitt 5.1.2). Neben diesen konkret auf die GK 25 bezogenen Schlussbetrachtungen lassen sich aus der Arbeit auch einige generelle Schlussfolgerungen ableiten, die in Abschnitt 5.2 dargelegt werden.

5.1 Schlussfolgerungen für die Geologische Karte 1:25.000

5.1.1 Diskussion der Arbeitsergebnisse

Gemäß der Zielsetzung dieser Arbeit, wurde ein Verfahren entwickelt, mit dem es möglich ist, den Datenbestand der GK 25 vom dualen Arc/Info-Modell zu einer integrierten Arc-GIS/ArcSDE/SQL-Server System-Architektur zu migrieren. Der erste wesentliche Schritt hierbei ist die Entwicklung eines neuen Datenmodells. Inwieweit das in dieser Arbeit vorgestellte Datenmodell noch Schwächen aufweist und ob es den in Abschnitt 3.4 gestellten Anforderungen entspricht, soll im Folgenden diskutiert werden.

Geologische Linienobjekte sind aus mehreren Features zusammengesetzt

Geologische Linienobjekte sind aus ein oder mehreren Features zusammengesetzt, ohne dass der Bezug zu den Features gespeichert ist. In der Abb. 4.12 ist diese Problematik schematisch am Beispiel einer Störung dargestellt. Das als *eine* Störung aufzufassende Linienobjekt besteht digital aus zwei Features der Feature Class Stoerung, wobei eines vom Subtyp OhneFlaechenbezug und eines vom Subtyp MitFlaechenbezug ist.

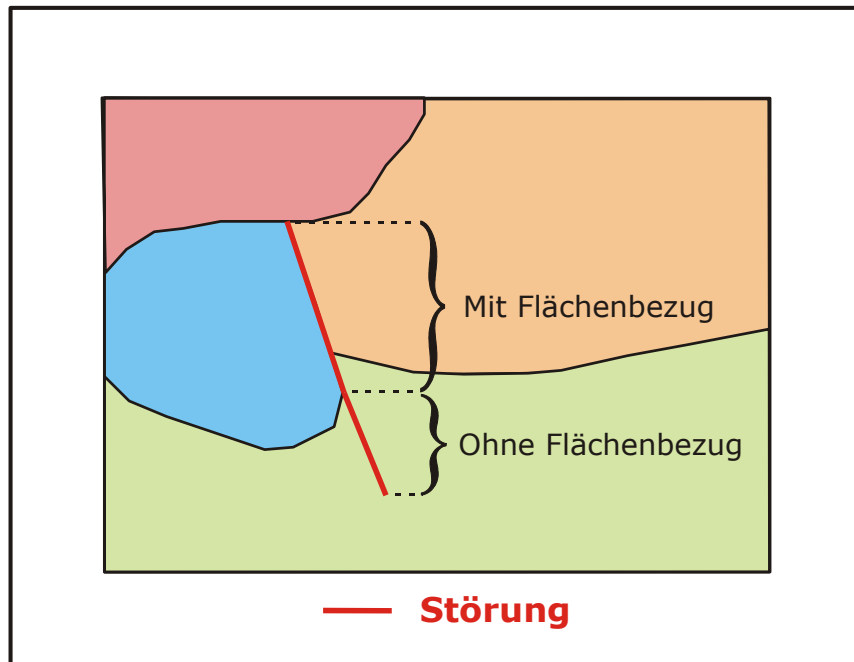


Abb. 4.12: Schematische Darstellung eines geologischen Linienobjekts der Feature Class Störung (Erläuterung im Text).

Das vorliegende Datenmodell ermöglicht also nicht den direkten Zugriff auf die Störung aus „fachlich geologischer“ Sicht, sondern nur aus „technischer“ Sicht. Für die derzeitigen Anforderungen an das Datenmodell der GK 25 stellt dies kein Problem dar. Doch für zukünftige Anwendungen könnte dieser Zugriff möglicherweise von Bedeutung sein. Technisch wäre es kein Problem, den Bezug zwischen den geologischen Linienobjekten und den Features, aus denen sie zusammengesetzt sind, herzustellen. In der Praxis ist diese Zuweisung jedoch sehr schwierig zu realisieren, denn selten ist der Verlauf der Störungen so simpel, wie in der Abb. 4.12 dargestellt (vgl. z. B. Anhang A4). Die Entscheidung darüber, welche Features zu welchen geologischen Objekten gehören, erfordert mindestens geologischen Sachverstand; in vielen Fällen wohl auch den örtlichen Sachverstand des Kartierers selbst. Eine derartige Zuweisung müsste für die GK 25 nachträglich durchgeführt werden, was einen zum jetzigen Zeitpunkt nicht gerechtfertigten Arbeitsaufwand bedeuten würde. Hieraus lässt sich jedoch auch ableiten, dass es zukünftig sinnvoll wäre, schon bei der Kartierung stärker die Erfordernisse des Datenmodells für die digitale Speicherung zu berücksichtigen. Dieser Aspekt wird weiter unten noch einmal aufgegriffen.

Blattrandproblematik

In Abschnitt 4.4.3 wurde erläutert, wie die Positionsfehler an den ehemaligen Blatträndern der Coverages korrigiert werden können. Nach wie vor existieren jedoch die Attribut-bezogenen Fehler in der Datenbasis. Das bedeutet, dass an den ehemaligen Blatträndern z. B. ein aus geologischer Sicht inkonsistenter Übergang zwischen Kartiereinheiten stattfinden kann, der nur durch eine Änderung der beschreibenden geologischen Attribute oder einer Anpassung der

Grenzverläufe behoben werden kann (vgl. Abb. 4.13). Weiterhin wäre es nicht nur aus geologischer sondern auch aus kartographisch-ästhetischer Sicht wünschenswert, die durch den ehemaligen Blattrand begründeten geraden Grenzverläufe aufzulösen bzw. realistischer zu gestalten. Die Attribut-bezogenen Fehler scheinen um so deutlicher zu werden, je weiter die Entstehungsjahre der benachbarten Kartenblätter auseinander liegen, was sich wahrscheinlich auf im Laufe der Zeit gewandelte geologische Konzepte und Kartiermethodiken zurückführen lässt.

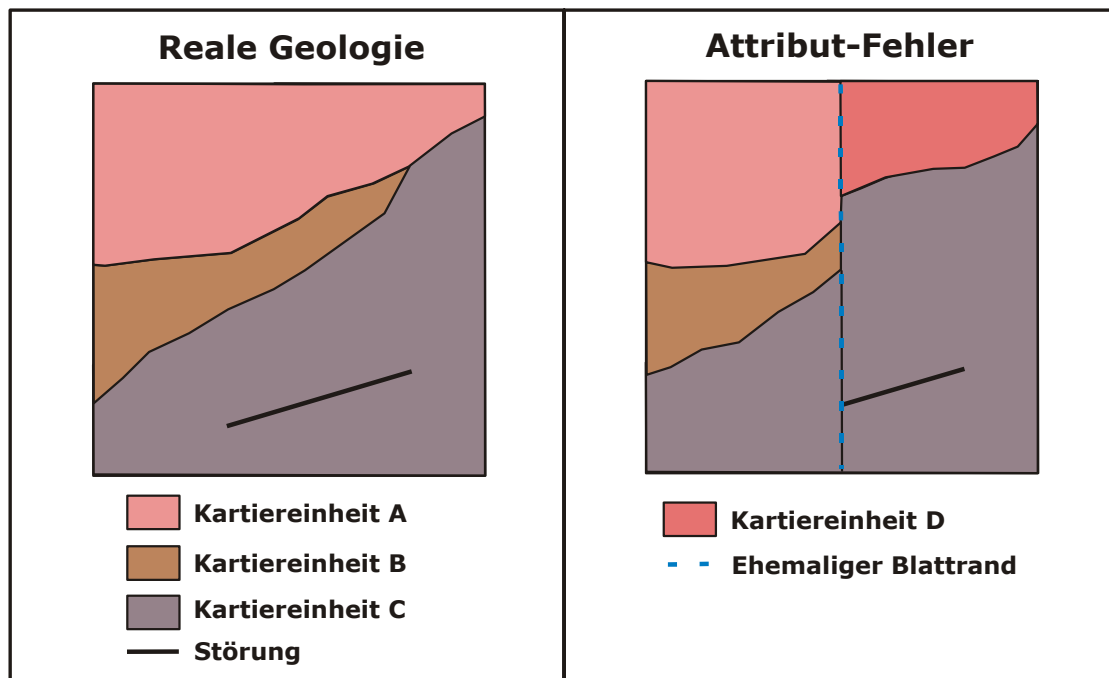


Abb. 4.13: Schematische Darstellung von Attribut-bezogenen Fehlern am ehemaligen Blattrand der Coverages (Erläuterung im Text).

Auch in diesem Fall würden Korrekturarbeiten geologischen Sachverstand und einen hohen Zeitaufwand erfordern. In vielen Fällen wäre es wahrscheinlich notwendig, die analogen Arbeitsgrundlagen (Manuskriptkarten und Feldbücher) des Kartierers zu Rate zu ziehen oder u. U. eine Neukartierung des fraglichen Gebietes durchzuführen. Diese Arbeiten sind auch deshalb so schwierig, weil die geologische Interpretation einen sehr hohen Anteil der Kartierarbeit ausmacht und sehr früh in den Entstehungsprozess einer Karte einfließt. Hieraus lässt sich zum einen ableiten, dass es zukünftig sinnvoll wäre, bei der geologischen Kartierung und selbstverständlich auch im digitalen Datenmodell stärker auf die Trennung von faktischen und interpretierten Arbeitsergebnissen zu achten. Weiterhin wäre es zukünftig sinnvoll und technisch möglich, die im Gelände ermittelten Arbeitsergebnisse (z. B. Position und Ansprache von Sondierbohrungen und Aufschlüssen) nicht analog sondern digital zu speichern. Durch diese beiden Aspekte wäre es wesentlich einfacher, die Nachvollziehbarkeit der Kartierungsergebnisse zu gewährleisten und effektive Revisionen der Kartierungsergebnisse zu ermöglichen.

Abschließend lässt sich festhalten, dass der geologische Randabgleich der GK 25 ein aufwändiges Projekt darstellen würde, das bei den derzeitigen personellen und finanziellen Mitteln des NLFb nur schwer realisierbar ist. Weiterhin ist der Bedarf an diesem Projekt nicht zwingend gegeben, da mit der GK 1:50.000 eine Kartenserie vorliegt, in der dieser Randabgleich stattgefunden hat.

Unterstützung der auf der GK 25 beruhenden Arbeitsabläufe

Dieser Punkt bezieht sich auf die in Abschnitt 3.4 (Punkt 6) formulierte Anforderung. Diese besagt im Wesentlichen, dass zumindest für einen Übergangszeitraum MeMaS und die auf AMLs beruhende Plotausgabe der Kartenblätter unterstützt werden müssen. Dies ist grundsätzlich möglich, da aus der Geodatabase Shapefiles für MeMaS und Coverages für die Plotausgabe exportiert werden können. Für die Coverages ist allerdings festzuhalten, dass die interne Struktur der aus der Geodatabase exportierten Coverages nicht der ursprünglichen Arc/Info Coverages entspricht, denn in letzteren waren sowohl die Polygongrenzen als auch alle weiteren geologischen Linienobjekte als Arcs abgelegt (vgl. Abschnitt 3.2.2). Wenn Coverages aus der Geodatabase exportiert werden, ist es notwendig, die Polygon und Polyline Feature Classes jeweils in separate Coverages zu exportieren. Für die Plotausgabe würde dies bedeuten, dass die AML-Prozedur z. B. dahingehend angepasst werden muss, dass sie die entsprechenden Coverages mit den Polygonen und Polylinien vor dem Ausdruck in geeigneter Weise übereinander legt. Weiterhin greift die AML-Prozedur auf das Linientypen-Attribut des Arc/Info-Modells zu. Dieses Attribut wird in der Geodatabase zwar mitgeführt, jedoch nicht gepflegt (vgl. Abschnitt 4.3.2). Demzufolge müssen die geologischen Linien nun mehr über das Subtypen-Attribut (*Subtype*) der Feature Class GeologischeLinie bzw. über das Kodierungs-Attribut (*Code*) der Feature Class Störungen visualisiert werden, was ebenfalls eine Anpassung der AML-Prozedur zur Folge haben muss. Die Visualisierung der Kartiereinheiten über das Attribut *KombiNr* folgt weiterhin demselben Konzept. Auf Grund der Migration der Daten kann es jedoch zu Änderungen von Namen und Pfadangaben gekommen sein, die ggf. in der AML-Prozedur zu korrigieren sind.

Übertragbarkeit der Arbeitsergebnisse auf die weiteren Flächendaten des NLFb

Gemäß der in Abschnitt 3.4 gestellten Anforderung (Punkt 8), ist das vorgestellte Datenmodell grundsätzlich auf die weiteren Kartenserien (vgl. hierzu Abschnitt 3.3) des NLFb übertragbar. Da die Linientypen der GK 25 und der GK 50 sich nicht vollständig identisch sind, müssen entsprechende Anpassungen des Datenmodells vorgenommen werden. Das bedeutet im Wesentlichen, dass für die Feature Class GeologischLinie einige Subtypen hinzukommen und einige gelöscht werden müssen. Für die Feature Class Störungen sollte überprüft werden, ob alle Attribute auch in der GK 50 benötigt werden.

Bei den bodenkundlichen Kartenserien wäre zu prüfen, ob die Mehrfachkodierung der Arcs in der Geodatabase wegfallen kann. Die Kodierung als Kartenrahmen erübrigt sich durch die blattschnittfreie Datenbasis und die zur Visualisierung dienende Kodierung als *Gewässergrenze* bzw. *Grenze anthropogen veränderter Flächen* kann wahrscheinlich wegfallen, da die Flächenbeschreibungen diese Aufgabe übernehmen können. Voraussetzung dafür ist, dass diese Grenzen nicht Gebiete umfassen, die mehrere Polygone umschreiben. In diesem Fall müsste eine Polyline Feature Class mit entsprechenden Subtypes implementiert werden. Die Kodierung als *Landesgrenze* sollte zukünftig durch eine separate Feature Class geregelt werden. Es bietet sich an, ein Feature Dataset mit verschiedenen Feature Classes für die verschiedenen Hierarchieebenen der offiziellen Verwaltungsgrenzen in der Datenbank bereitzustellen.

5.1.2 Weiterführende Arbeitsschritte

Neben dem Entwurf eines Datenmodells und der Migration der Datenbasis selbst müssen bei der Umstellung auf eine integrierte ArcGIS/SDE-Architektur noch viele weitere Aspekte beachtet werden. Einige Arbeitsschritte, die direkt im Anschluss an diese Arbeit angegangen werden sollten, werden im Folgenden kurz erörtert.

Entwicklung eines Konzepts für die Visualisierung und Plotausgabe der Karteninhalte

Auf die Visualisierung der GK 25 wurde in den vorherigen Abschnitten bereits mehrfach eingegangen. Mit dem vorgestellten Datenmodell sollte es möglich sein, hierfür ein praktikables Konzept zu entwickeln und umzusetzen. Im Wesentlichen besteht die Aufgabe darin, die bisher zur Plotausgabe verwendete AML-Prozedur in Programmcode zu überführen, der von ArcGIS 8.3 ausführbar ist. Das bedeutet, dass der Programmcode in einer Programmiersprache geschrieben werden muss, die kompatibel zum Microsoft Component Object Modell (COM) ist, wie z. B. Visual Basic (VB), Visual Basic for Applications (VBA), Delphi oder C++. Vor den Programmierarbeiten sind noch einige konzeptionelle Überlegungen notwendig, die z. B. untersuchen inwieweit ArcGIS neue Möglichkeiten zur Visualisierung bietet. Zu diesen Überlegungen gehört auch, ein Konzept für die Behandlung der Metadaten (und anderer Kartenrandobjekte) zu entwickeln, denn diese müssen mit der gedruckten Karte in geeigneter Weise an den Kunden weitergegeben werden. Diesem wichtigen Punkt ist der nächste Abschnitt gewidmet.

Entwicklung eines Konzepts zur Behandlung der Metadaten

Auf die Problematik bei der Bereitstellung von Metadaten für die GK 25 wurde ansatzweise bereits in Abschnitt 4.1 eingegangen. Sofern weiterhin an den Kunden nur Daten oder Karten weitergegeben werden, die auf den Blattschnitt einer TK 25 bezogen sind, ist es nicht schwer, auch die Metadaten bereitzustellen. In diesem Fall ergeben sich konzeptionell keine Änderungen gegenüber dem Arc/Info-Modell. Problematisch wird die Bereitstellung dann, wenn der

Kunde dazu in die Lage versetzt wird, ein Gebiet aus der Datenbasis frei auszuwählen – was ein wesentlicher Vorteil einer blattschnittfreien Datenbasis ist. Dieser Ausschnitt kann geologische Daten aus mehreren ehemaligen Coverages enthalten und demzufolge haben auch die Metadaten z. T. weiterhin Bezug zum ehemaligen Blattschnitt (z. B. Erstellungsjahr der Karte). Für diese Problematik ist eine Lösung zu finden, die auch der Art und Weise der Weitergabe bzw. Darstellung der Metadaten beinhaltet.

Schulung der Mitarbeiter und Entwicklung eines Konzepts für den Arbeitsablauf von Editierarbeiten

Es ist selbstverständlich, dass die Mitarbeiter bei einer Umstellung der Technologie entsprechend geschult werden, damit sie ihre Aufgaben möglichst reibungslos in der Umgebung der neuen Technologie bewältigen können. Bei der Umstellung auf ArcGIS 8.x/SDE ist es insbesondere für die Mitarbeiter, welche Editierarbeiten wahrnehmen sollen, wichtig, die neuen Editierwerkzeuge kennen zu lernen und das Konzept des Versionsmanagement (*Versioning* in der ESRI-Terminologie) zu verstehen. Letzteres ist Voraussetzung dafür, die Arbeitsabläufe in der Multi-User Umgebung von ArcSDE erfolgreich zu koordinieren und durchzuführen. Dabei ist es weiterhin notwendig, ein gutes Verständnis für die Administration der SDE-Datenbank(en) zu entwickeln. Die Aufgaben des SDE Administrators stehen z. T. in engem Zusammenhang zu den Editierarbeiten (z. B. Kompression der Datenbanken, vgl. Kapitel 11 in ESRI 2003a) und können sich signifikant auf die Performance der Datenbank(en) auswirken (z. B. durch die Index-Verwaltung).

Steigern der Datenqualität

In den Geo- oder Umweltwissenschaften ist es wohl selten der Fall, dass eine Datenbasis perfekt genannt werden kann. Auch bei der GK 25 bieten sich einige Ansatzpunkte zur Verbesserung der Datenbasis an. Es ist klar, dass einige der folgenden Punkte einen erheblichen finanziellen und personellen Aufwand erfordern würden, der unter den derzeitigen Rahmenbedingungen des NLFb als unrealistisch anzusehen ist. Nichtsdestotrotz sollen diese Aspekte aus Gründen der Vollständigkeit angesprochen werden. Zunächst wäre es wünschenswert, für die geologisch kartierte Fläche von Niedersachsen einen Deckungsgrad von 100 % zu erreichen. Der nächste Punkt wurde bereits in Abschnitt 5.1 erläutert und betrifft den geologischen Randabgleich zwischen den ehemaligen Blatträndern der Coverages. Erstrebenswert wäre weiterhin eine Aktualisierung bestimmter Gebiete der Datenbasis, insbesondere dort, wo die Kartierungen sehr alt sind. Die weiteren Punkte sind mit deutlich weniger Aufwand und ohne Kartierarbeiten zu erreichen. In Abschnitt 4.4.3 wurde berichtet, dass die Datenbasis der GK 25 Geometrien enthält, die nicht optimal an den Ziel-Maßstab von 1:25.000 angepasst sind, da die betreffenden Features zu klein sind oder sehr eng beieinander liegen. Da diese Features u. U. im Zielmaßstab nur schlecht lesbar sind, wäre es angebracht, eine an den Maßstab angepasste Generalisierung der Datenbasis durchzuführen. Dies hätte weiterhin den Vorteil, dass mit einer etwas größeren Cluster Toleranz gearbeitet werden könnte, was wiederum effektivere Editierarbeiten zur Folge

hätte (vgl. Abschnitt 4.3.1). Eine weitere Folge wäre, dass die Anzahl der Vertizes in der Datenbasis herabgesetzt werden würde, was positive Auswirkungen auf den benötigten Speicherplatzbedarf und auf die Performance beim Darstellen und Editieren haben kann. Des Weiteren könnte das Datenmodell in Bezug auf die geologischen Linientypen noch ein wenig vereinfacht werden (zu den Vorteilen eines einfachen Datenmodell siehe Abschnitt 3.4, Punkt 1). Mit den Linientypen *Verbreitung kalkhaltiger Schichten im Quartär* und *Schichtgrenze im Mesozoikum, unter mehr als 2m Quartär* (vgl. Anhang A3) liegen z. B. Linientypen vor, die aus historischen Gründen nur auf Einzelblättern der GK 25 realisiert sind. Es ist fraglich, ob es sinnvoll ist, derartige Linientypen in der Datenbasis zu führen. Dies stellt einen Bruch mit einem durchgängigen, standardisierten Datenmodell, das für die gesamte Datenbasis die gleiche Gültigkeit besitzt. Insbesondere für Kunden ist es nicht ohne Weiteres nachzuvollziehen, warum derartige Linientypen auf einigen Kartenblättern existieren und auf anderen nicht. Andere Linien- und Störungstypen treten ebenfalls extrem selten in der Datenbasis auf. In diesen Fällen könnte geprüft werden, ob die geologische Zusatzinformation zu Gunsten eines einfacheren Datenmodells geopfert werden kann. Es wäre denkbar, diese besonderen Linientypen aus dem generellen Datenmodell zu entfernen und für evtl. auftretende Spezialfragestellungen in separaten Feature Classes zu archivieren.

5.3 Allgemeine Schlussfolgerungen

In Abschnitt 3.1 wurde dargelegt, dass die Inhalte traditioneller geologischer Karten einen starken interpretativen Charakter aufweisen. Schon die faktischen Ausgangsdaten, wie Bohrungen oder Beobachtungen, sind in der Geologie – und in anderen Geo- oder Umweltwissenschaften – mit Unsicherheit behaftet. Die Unsicherheit der Ausgangsdaten und deren notwendige Interpretation durch den Kartierer führen dazu, dass auch die kartographische Repräsentation der Geologie mit fehlerbehaftet ist. Weiterhin enthält eine geologische Karte mehr oder weniger viel implizites Wissen, welches vom Betrachter u. U. mit Hilfe von Zusatzinformationen (wie z. B. Profilschnitten) selbstständig erschlossen werden muss. Diese eher ungünstigen Aspekte sind z. T. darauf zurückzuführen, dass zu Zeiten der analogen Kartographie die Karte auch als Datenspeicher genutzt werden musste. Dabei war es schlicht nicht möglich, alle Bohrerergebnisse, Beobachtungen oder Eigenschaften der geologischen Objekte auf einer Karte sinnvoll darzustellen. Digitale raumbezogene geologische Datenbasen wurden oft mit dem Ziel erstellt, die Kartenproduktion durch die Computertechnologie effizienter zu gestalten. Dabei war es erwünscht, die analogen Karten möglichst gut zu reproduzieren. Eine digitale Speicherung der raumbezogenen geologischen Daten kann jedoch mehr bieten als ein bloßes Faksimile einer analogen geologischen Karte, denn sie kann dazu genutzt werden, die oben genannten Schwächen geologischer Karten zu mildern. Es können Konzepte entwickelt werden, die eine schärfere Trennung von faktischen und interpretierten Informationen zulassen. Dies würde die Nachvollziehbarkeit der Kartiererergebnisse verbessern und die Aktualisierung und Revision von kartierten Gebieten erleichtern. Auch für das in der Karte implizit vorhandene Wissen, könnte geprüft werden, ob es nicht in explizit gespeicherte Informationen umgesetzt

werden kann. Für den Umgang mit unsicherem (raumbezogenen) Wissen lassen sich in der Literatur verschiedene Beispiele finden (vgl. FOODY & ATKINSON 2002, HEUVELINK 1998, HEUVELINK 1993). Entsprechende Algorithmen haben jedoch bisher selten Einzug in kommerzielle GIS erhalten, so dass hier weiterhin hoher Forschungsbedarf besteht. Die Anwendung entsprechender Methoden könnte dazu führen, den Interpretationsspielraum geologischer Karten zu verkleinern. Ferner wäre es sinnvoll, die Standardisierung von geologischen Datenmodellen auf nationaler und internationaler Ebene voranzutreiben. Nicht nur weil auf diese Weise der Datenaustausch verbessert und Doppelarbeit vermieden wird, sondern auch weil so generell der Umgang mit den Daten vereinfacht wird und Synergieeffekte entstehen können.

Die Arbeit hat weiterhin gezeigt, dass es für geographische Daten nur schwer möglich ist, einen Datenmodellierungsprozess durchzuführen, der dem bewährten Schema des Entwurf „herkömmlicher“ Datenbanken folgt. Dieses beinhaltet, mit abnehmendem Grad der Abstraktion, den Entwurf eines konzeptuellen, eines logischen und eines physischen Modells (vgl. Abschnitt 1.2). Bei geographischen Daten ist es bei den abstrakten Modellierungsschritten schwierig, sich von den speziellen Anforderungen zu lösen, die die zur Implementierung gewählte Software an das Datenmodell stellt. Ebenso ist das in dieser Arbeit benutzte Werkzeug zur Datenbankmodellierung, die UML, nicht auf die spezielle Anforderungen von geographischen Daten zugeschnitten. Es wäre wünschenswert, UML-Diagramme für die Modellierung von geographischen Datenbasen entwickeln zu können, die bei der Implementierung automatisch in die interne Datenstruktur der gewählten Datenbank bzw. des gewählten GIS umgesetzt werden. Um dies zu verwirklichen, muss zunächst ein allgemein akzeptiertes Datenmodell auf konzeptueller Ebene entwickelt werden, wie es z. B. PARENT, SPACCAPIETRA & ZIMÁNYI (1999) zeigen. Auch die voranschreitende Standardisierung von geographischen Datenstrukturen, wie z. B. mit Hilfe der Geography Markup Language (GML, vgl. z. B. OGC 2003a), lässt hoffen, dass die zukünftige Entwicklung bessere Möglichkeiten zur Datenmodellierung von Raumdaten bereithält.

6 Zusammenfassung

Diese Arbeit befasst sich mit der Migration eines großen Bestandes von raumbezogenen Daten zu einer integrierten Speicherung in einem relationalen Datenbankmanagementsystem (DBMS). Bei dem Datenbestand handelt es sich um die Flächendaten der Geologischen Karte von Niedersachsen im Maßstab 1:25.000 (GK 25) des Niedersächsischen Landesamtes für Bodenforschung (NlfB). Der Datenbestand wurde bisher in einem dualen Speichersystem verwaltet, welches aus dem Geographischen Informationssystem (GIS) Arc/Info der Firma Environmental Systems Research Institute (ESRI) und dem DBMS SQL-Server der Firma Microsoft besteht. Die Geometriedaten wurden dabei mit Arc/Info und die Attributdaten mit SQL-Server gespeichert. Die Migration des Datenbestandes zu einem integrierten Speichersystem wird geplant, da sich hierdurch für den Nutzer wesentliche Vorteile ergeben, die im Kapitel 2 der Arbeit erläutert werden. Zu diesen gehören u. a. der Multi-User-Betrieb, die blattschnittfreie Datenbasis, die Verein-

fachung der Datenverwaltung durch Integration der Raum- und Attributdaten und die Nutzung von ausgereiften DBMS-Technologien zur Gewährleistung der Datenintegrität und –sicherheit.

Ziel der Arbeit ist es, einen Weg aufzuzeigen, auf dem die Migration des Datenbestandes zu einer topologisch fehlerfreien Datenbasis in einer Software-Umgebung bestehend aus ArcGIS 8.3, ArcSDE 8.3 und SQL-Server 2000 durchzuführen ist. Der Schwerpunkt der Arbeit liegt dabei auf der Entwicklung und Implementierung eines physischen Datenmodells für eine ArcGIS Geodatabase. Um die Ziele zu erreichen, wurde ein Arbeitsprogramm festgelegt, welches sich an der gängigen Praxis zur Datenmodellierung für DBMS orientiert. Zu den aufeinander aufbauenden Arbeitsschritten gehören:

- 1) Definitionsphase
- 2) Entwurf eines konzeptuellen Datenmodells
- 3) Umsetzung des konzeptuellen Datenmodells in ein physisches Datenmodell
- 4) Implementierung des physischen Datenmodells
- 5) Migration des Datenbestandes
- 6) Bereinigung topologischer Fehler
- 7) Bewertung der Ergebnisse und Identifikation weiterführender Arbeitsschritte

Die Definitionsphase dient im Wesentlichen dazu, die Sicht des Nutzers auf die Daten kennen zu lernen und die Anforderungen an das zu entwickelnde Datenmodell zu definieren. Hierzu wurde die Natur geologischer Karten, das bestehende Arc/Info-Datenmodell und die Rahmenbedingungen, unter denen die GK 25 am NLFb in die Arbeitsabläufe integriert ist, analysiert und ein Anforderungskatalog aufgestellt. Dieser beinhaltet u. a., dass das Datenmodell möglichst einfach konzipiert und auf weitere Kartenserien des NLFb übertragbar sein soll, dass die Datenbasis keinen fachlichen Veränderungen unterworfen werden soll, dass der Arbeitsaufwand möglichst gering gehalten werden muss und dass sämtliche auf die GK 25 angewiesenen Arbeitsabläufe zu jedem Zeitpunkt des Projektes unterstützt werden müssen.

Der Entwurf des konzeptuellen und des physischen Datenmodells wurde mit Hilfe der *Universal Modeling Language* (UML) durchgeführt. Dieses Werkzeug ist ein De-facto-Standard für den Datenbankentwurf und begünstigt einen strukturierten Modellierungsprozess, dessen Ergebnisse mit UML-Klassendiagrammen visualisiert werden können. Diese Diagramme können direkt dazu benutzt werden, die Datenbankstruktur zu implementieren, wenn die DBMS-Software diesen Vorgang unterstützt, was auch bei der Geodatabase von ArcGIS 8.3 der Fall ist. Die Ergebnisse der Datenmodellierung sind in den Anhängen B1-B3 und B5 dargestellt. Die Geometriedaten in die leere Datenbankstruktur der entworfenen Geodatabase zu laden, ist ein einfacher Vorgang, der mit dem Simple Data Loader von ArcGIS durchgeführt wurde. Die auch im Arc/Info-Modell schon in SQL-Server Tabellen gespeicherten Attributdaten wurden direkt in

die Geodatabase kopiert und die Beziehungen zwischen den verschiedenen Tabellen wurden in der SDE-Datenbank als Relationship Classes angemeldet.

Nach dem Laden der Daten zeigte sich, dass die Datenbasis an den ehemaligen Blatträndern der Arc/Info-Coverages zahlreiche Topologie-Fehler aufwies. Diese Positionsfehler äußerten sich als Überlappungen und Lücken von bzw. zwischen den Polygonen, welche die geologischen Kartiereinheiten repräsentieren. Die Positionsfehler lassen sich auf ungenaue Georeferenzierung und auf die Bildung der Topologien in Arc/Info zurückführen. In der Arbeit werden Ansätze diskutiert, die zur Behebung bzw. Vermeidung dieser Topologie-Fehler beitragen können. Für die Geometriedaten der GK 25 erweist es sich als sinnvoll, die Topologie-Fehler mit Hilfe eines von ArcGIS bereitgestellten Werkzeuges (*Zipper Task* genannt) zu beheben. Dieses Werkzeug hat gegenüber anderen Methoden, die Auswirkungen auf die gesamte Datenbasis haben, den Vorteil, dass es im Allgemeinen nur die Polygone verändert, die tatsächlich am ehemaligen Blattrand liegen. Dadurch, dass dieses Werkzeug interaktiv gesteuert wird, hat es den weiteren Vorteil, dass es den lokalen Gegebenheiten der Fehler flexibel angepasst werden kann. Andererseits läuft es nicht voll automatisch ab, wodurch ein gewisser Zeitaufwand für Editierarbeiten eingeplant werden muss, der aber für die Datenbasis der GK 25 vertretbar erscheint.

Die Arbeit schließt mit einer Diskussion der Arbeitsergebnisse, in der u. a. Schritte ausgearbeitet werden, die zur Weiterführung der Systemumstellung notwendig sind. Weiterhin werden Ideen präsentiert, die zu einer generellen Verbesserung von raumbezogenen geologischen Datenbasen und zum Modellierungsprozess von geographischen Daten beitragen können.

7 Summary

This thesis is about the migration of a large spatial database into an integrated storage in a relational database management system (DBMS). The spatial database is the geological map in the scale 1:25,000 (GK 25) Geological Survey of Lower Saxony (NLfB). Until now this database was stored in a dual system, comprised of the geographic information system (GIS) Arc/Info from the Environmental Systems Research Institute (ESRI) and the DBMS SQL-Server 2000 from Microsoft. In this system geometry data is stored in Arc/Info and the attribute data is stored in SQL-Server. The migration is planned because of many vital advantages it brings for the user. Among these advantages are multi-user access, a sheetless database, simplification of data handling, integration of geometry and attribute data, and the use of mature DBMS-technology for assuring data integrity and data security (further discussion in chapter 2 of this thesis).

The aim of this study is to work out a way that permits the migration of the geological data into a software environment comprised of ArcGIS 8.3, ArcSDE 8.3 und SQL-Server 2000. Additionally the database should be free of topology errors after the migration process. The main focus of this thesis is on the development of a conceptual and physical data model for an

ArcGIS Geodatabase. Based on the common practice of database modelling for DBMS a working schedule was designed, which included the following tasks:

- 8) Definition phase
- 9) Design of a conceptual data model
- 10) Translating of the conceptual data model into a physical data model
- 11) Implementation of the physical data model
- 12) Migration of the database
- 13) Fixing topology errors
- 14) Reflection on the results and identification of necessary future tasks

In the definition phase the user's view on the data is analysed and requirements for the data model that is to be developed are defined. To achieve this, the nature of geological maps, the current Arc/Info data model and the work flow which integrates the GK 25 were analysed and a list of requirements was formulated. This list includes, among others things, the demand that the data model should be simple and transferable to the other map series of the NLFb, that the data should not be modified in its geological meaning and that all work flow processes that depend on the GK 25 should be supported at any time of the project.

The Universal Modelling Language (UML) was used to design both the conceptual and the physical models. This tool is a de facto standard for database design, additionally it encourages a structured modelling process whose results can be visualised by UML class diagrams. These diagrams can be used to directly implement the database if the DBMS software supports this process, as is the case with ArcGIS 8.3. The results of the database modelling are depicted in appendix B1-B3 and B5 of this thesis.

Loading of the geometry data is a straightforward process in which the Simple Data Loader of ArcGIS was used. The attribute data, which was already stored in SQL-Server tables in the Arc/Info data model, was directly copied into the Geodatabase and the relations between the various tables were registered in the SDE-database via Relationship Classes.

After data loading it became obvious that there were many topology errors in the database. These errors were located at the former sheet edges of the Arc/Info Coverages and manifested themselves as gaps and overlaps between adjacent polygons representing the geologic mapping units. The cause of this positional errors can be tracked down to imprecise georeferencing and the topology building process in Arc/Info. This thesis discusses several approaches that can help avoiding or correcting the topology errors. An appropriate method for correcting the topology errors of the GK 25 data is demonstrated, which involves a functionality of ArcGIS, called Zipper Task. This method modifies only the geometries of polygons that actually lie at the former sheet edges, which is an advantage over other methods that affect the whole database. As Zipper

Task is an interactively controlled tool it has the additional advantage of being flexibly adjustable to the local conditions of errors. On the other hand, a correction of errors can not be fully automated through this tool. Thus, a certain time span must be invested into editing the database, which is of an acceptable duration in the case of the GK 25 data.

The thesis closes with the discussion of the results and necessary future tasks are derived, that are needed to accomplish the system's re-engineering. Furthermore, some thoughts about improving spatial geological databases and the modelling process for spatial data are presented.

8 Literaturverzeichnis

- ADAM, N. R. & A. GANGOPADHYAY (1997): *Database Issues in Geographic Information Systems*. Kluwer Academic Publishers, Boston, 135 S.
- ALBRECHT, J. (1999): *Geospatial Information Standards. A Comparative Study of Approaches in the Standardisation of Geospatial Information*. In: COMPUTERS & GEOSCIENCES, Band 25, S. 9-24.
- ALLEN, P. M. (1996): *Standardization of Mapping Practices in the British Geological Survey*. In: COMPUTERS & GEOSCIENCES, Band 23, Nr. 6, S. 609-612.
- BAIN, K. A. & J. R. A. GILES (1997): *A Standard Model for Storage of Geological Map Data*. In: COMPUTERS & GEOSCIENCES, Band 23, Nr. 6, S. 613-620.
- BATTY, P. (1992): *Exploiting Relational Database Technology in a GIS*. In: COMPUTERS & GEOSCIENCES, Band 18, Nr. 4, S. 453-462.
- BATTY, P. M. & R. G. NEWELL (1996): *GIS Databases are different*. Smallworld GIS White Paper.
- BÉDARD, Y. (1999): *Principles of Spatial Database Analysis and Design*. In: MAGUIRE, D. J., GOODCHILD, M. F. & D. W. RHIND (Hrsg.): *Geographical Information Systems*. 2. Aufl., Longman, London, S. 413-424.
- BLASCHKE, R. et al. (1989): *Interpretation geologischer Karten*. 2. Aufl., Ferdinand Enke, Stuttgart, 159 S.
- BOESS, J., MÜLLER, U. & J. SBRESNY (1999): *Erläuterungen zur digitalen Bodenkundlichen Übersichtskarte 1:50.000 (BÜK 50) von Niedersachsen*. In: NIEDERSÄCHSISCHES LANDESAMT FÜR BODENFORSCHUNG (Hrsg.), Arbeitshefte Boden, Heft 1999/1, Hannover, S. 3-60.
- BOMBIEN, H. & R. LIES (1991): *Die Verknüpfung konventionell-kartographischer und digitaler Arbeitsweisen bei der Bearbeitung geowissenschaftlicher Karten*. In: BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE (Hrsg.): *Geologische und bodenkundliche Landesforschung in Niedersachsen*. Geologisches Jahrbuch, Reihe A, Heft 126, Hannover.
- BOMBIEN, H. & R. STOLZ (1993): *Die digitale Erfassung und Speicherung geologischer Flächendaten im Niedersächsischen Bodeninformationssystem (NIBIS)*. In: BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE (Hrsg.): *Wissenschaftliche Berichte zum Niedersächsischen Bodeninformationssystem (NIBIS) (1. Folge)*. Geologisches Jahrbuch, Reihe F, Heft 27, Hannover.
- BRODARIC, B. & M. GAHEGAN (2002): *Distinguishing Instances and Evidence of Geographical Concepts for Geospatial Database Design*. In: EGENHOFER, M. & D. MARK (Hrsg.): *Geographic Information Science*. 2nd Intern. Conference, GIScience 2002, Lecturer Notes in Computer Science 2478, S. 55-68.

- BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE (Hrsg.) (2002): *Geologische Kartieranleitung*. Geologisches Jahrbuch, Reihe G, Heft 9, Hannover, 135 S.
- BURROUGH, P. A. (1992): *Are GIS Data Structures Too Simple Minded?* In: COMPUTERS & GEOSCIENCES, Band 18, Nr. 4, S. 395-400.
- CASPERS, G., ELBRACHT, J. & E. SCHNEIDER (2004): *Hochwassergefährdungskarte von Niedersachsen – Ergebnis einer methodischen Auswertung geologischer Fachdaten*. In: ZEITSCHR. F. ANGEW. GEOLOGIE (in Vorbereitung).
- CHEN, P. P. (1976): *The Entity-Relationship Model: Toward a Unified View of Data*. In: COMMUNICATIONS OF THE ACM, Band 13, Heft 6, S. 377-387.
- CHRISMAN, N. R. (1990): *Deficiencies of Sheets and Tiles: Building Sheetless Databases*. In: INT. J. GEOGRAPHICAL INFORMATION SYSTEMS, Band 4, Nr. 2, S. 157-167.
- CODD, E. F. (1990): *The Relational for Database Management, Version 2*. Addison-Wesley, Reading, 536 S.
- CONNOLLY, T. & C. BEGG (2002): *Database Systems*. 3. Aufl., Addison-Wesley, London, 1236 S.
- CONNOLLY, T., BEGG, C. & A. STRACHAN (1996): *Database Systems: A Practical Approach to Design, Implementation and Management*. Addison-Wesley, London.
- DATE, C. (1986): *An Introduction to Database Systems*. 4 Aufl., Addison-Wesley, Reading.
- DAVID, R. ET AL. (1993): *GeO2: Why Objects in a Geographical DBMS?* In: ABEL, D. & B. C. OOI (Hrsg.): *Advances in Spatial Databases*, Proceedings of the 3rd Intern. Symposium, SSD'93, Singapore, June 1993, Springer, S. 264-276.
- DEMERS, M. N. (1997): *Fundamentals of Geographic Information Systems*. John Wiley & Sons, New York, 486 S.
- DULLER, P. R. (1995): *The Quality Assurance of Geological Data*. In: GILES, J. R. A. (Hrsg.) (1995): *Geological Data Management*. Geological Society Special Publication Nr. 97, S. 5-11.
- EGENHOFER, M. J. (1994): *Deriving the Composition of Binary Topological Relations*. In: J. VISUAL LANGUAGES AND COMPUTING, Band 4 Heft 2, S. 133-149.
- EGENHOFER, M. J., FRANK, A. U. & J. P. JACKSON (1989): *A Topological Data Model for Spatial Databases*. In: BUCHMANN, A. et al. (Hrsg.): *Design and Implementation of Large Spatial Databases*, Lecturer Notes in Computer Science 409, S. 271-286.
- ELMASRI, R. & S. B. NAVATHE (1994): *Fundamentals of Database Systems*. 2 Aufl., Benjamin/Cummings, Redwood City.
- ERLER, T. (2002): *UML*. Moderne Industrie Buch, Bonn, 336 S.

- ESRI (Hrsg.) (2004a): *Tips and Tricks for Using ArcGIS 8.3*. Environmental Systems Research Institute, White Paper, Online:
<http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (09.03.2004).
- ESRI (Hrsg.) (2004b): *Versioning*. Environmental Systems Research Institute, Technical Paper, Online: <http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (09.03.2004).
- ESRI (Hrsg.) (2004c): *Versioning Workflows*. Environmental Systems Research Institute, Technical Paper, Online:
<http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (09.03.2004).
- ESRI (Hrsg.) (2003a): *Building a Geodatabase*. Environmental Systems Research Institute, Software Documentation, Redlands.
- ESRI (Hrsg.) (2003b): *System Design Strategies*. Environmental Systems Research Institute, White Paper, Online:
<http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (29.09.2003).
- ESRI (Hrsg.) (2003c): *ArcGIS: Working with Geodatabase Topology*. Environmental Systems Research Institute, White Paper, Online:
<http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (15.07.2003).
- ESRI (Hrsg.) (2003d): *Geoscience Data Model*. Environmental Systems Research Institute, Online:
<http://support.esri.com/index.cfm?fa=downloads.dataModels.filteredGateway&dmid=30> (11.08.2003).
- ESRI (Hrsg.) (2000): *Multiuser Geographic Information Systems with ArcInfo 8*. Environmental Systems Research Institute, White Paper, Online:
<http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (15.07.2003).
- ESRI (Hrsg.) (1999): *Modelling Our World*. Environmental Systems Research Institute, Software Documentation, Redlands.
- ESRI (Hrsg.) (1998): *Spatial Database Engine*. Environmental Systems Research Institute, White Paper, Online:
<http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (15.07.2003).
- FGDC (1998): *Content Standard for Digital Geospatial Metadata*. Federal Geographic Data Committee, FGDC-STD-001-1998, Washington, Online:
http://www.fgdc.gov/standards/documents/standards/metadata/v2_0698.pdf (09.03.2004)
- FLACKE, W. & B. KRAUS (2003): *Koordinatensysteme in ArcGIS*. Points Verlag, Norden, 255 S.
- FOODY, G. M. & P. M. ATKINSON (Hrsg.) (2002): *Uncertainty in Remote Sensing and GIS*. John Wiles & Sons, New York.

- FRANK, A. (1988): *Requirements for a Database Management System for a GIS*. In: PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING, 54, S. 1557-1565.
- GADIA, S. K. & V. CHOPRA (1993): *A Relational Model and SQL-like Query Language for Spatial Databases*. In: ADAM, N. R. & B. K. BHARGAVA (Hrsg.): *Advanced Database Systems*, Lecturer Notes in Computer Science 759, Springer, Berlin.
- GILES, J. R. A. & K. A. BAIN (1995): *The Nature of Data on a Geological Map*. In: GILES, J. R. A. (Hrsg.) (1995): *Geological Data Management*. Geological Society Special Publication Nr. 97, S. 33-40.
- GOODCHILD, M. F. (1992a): *Geographical Information Science*. In: INT. J. GEOGRAPHICAL INFORMATION SYSTEMS, Band 6, Nr. 1, S. 31-45.
- GOODCHILD, M. F. (1992b): *Geographical Data Modelling*. In: COMPUTERS & GEOSCIENCES, Band 18, Nr. 4, S. 401-409.
- GOODCHILD, M. F. (1999): *Modelling in Distributed Environments*. In: GOODCHILD, M. F. et al. (Hrsg.): *Interoperating Geographic Information Systems*, Kluwer, London, S. 133-134.
- HADZILACOS, T. & N. TRYFONA (1996): *Logical Data Modelling for Geographical Applications*. In: INT. J. GEOGRAPHICAL INFORMATION SYSTEMS, Band 10, Nr. 2, S. 179-203.
- HARDY, P. (2001): *Active Objects and Dynamic Topology for Spatial Data Re-Engineering and Rich Data Modelling*. Online: www.laser-scan.com (18.07.03).
- HARMON, J. E. & S. J. ANDERSON (2003): *The Design and Implementation of Geographic Information Systems*. John Wiley & Sons, New York.
- HEALEY, R. G. (1991): *Database Management Systems*. In: MAGUIRE, D. J., GOODCHILD, M. F. & D. W. RHIND (Hrsg.): *Geographical Information Systems: Principles and Applications*, Longman, Essex, S. 251-267.
- HEINEKE, H. J. et al. (2002): *Das MeMaS im Bodeninformatiionssystem NIBIS*. In: CYFFKA, B. & J. W. HÄRTLING (Hrsg.): *Bodenmanagement*, Springer, Berlin.
- HEUVELINK, G. B. M. (1998): *Error Propagation in Environmental Modelling with GIS*. Tylor & Francis, London, 127 S.
- HEUVELINK, G. B. M. (1993): *Error Propagation in Quantitativ Spatial Modelling: Applications in Geographic Information Systems*. Nederlandse Geografische Studies, Band 163, Utrecht.
- HOEL, E., MENON, S. & S. MOREHOUSE (2003): *Building a Robust Relational Implementation of Topology*. In: HADZILACOS, T. et al. (Hrsg.): *Advances in Spatial and Temporal Databases*, Proceedings of the 8th Intern. Symposium on Spatial and Temporal Databases, SSTD 2003, Santorini Island, July 2003, Lecturer Notes in Computer Science 2750, Springer, Berlin, Online: <http://support.esri.com/index.cfm?fa=knowledgebase.whitePapers.gateway> (15.09.03)

- ISO (2003): *Geographic Information – Metadata*. International Standard 19115:2003(E). International Standards Organisation, Technical Committee ISO/TC 211 (Hrsg.).
- JOHNSON, B. R. et al. (1999): *Geologic Map Data Model Version 4.3*. United States Geological Survey (Hrsg.). Online: <http://geology.usgs.gov/dm/model/Model43a.pdf> (14.09.2003).
- KENNEDY, H. (Hrsg.) (2001): *Dictionary of GIS Terminology*. ESRI Press, Redlands, 116 S.
- KLEINER, C. (2003): *Modeling Spatial, Temporal and Spatio-Temporal Data in Object-Relational Database Systems*. Dissertation, Fachbereich Informatik, Universität Hannover, 202 S.
- KOLOVSON, C. P., NEIMAT, M.-A. & S. POTAMIANOS (1993): *Interoperability of Spatial and Attribute Data Managers: A Case Study*. In: ABEL, D. & B. C. OOI (Hrsg.): *Advances in Spatial Databases*, Proceedings of the 3rd Intern. Symposium, SSD'93, Singapore, June 1993, Springer, S. 239-259.
- LAURINI, R. & D. THOMPSON (1992): *Fundamentals of Spatial Information Systems*. Academic Press, New York.
- LAXTON, J. L. & K. BECKEN (1996): *The Design and Implementation of a Spatial Database for the Production of Geological Maps*. In: COMPUTERS & GEOSCIENCES, Band 22, Nr. 7, S. 723-733.
- MAGUIRE, D. J. & J. DANGERMOND (1991): *The Functionality of GIS*. In: MAGUIRE, D. J., GOODCHILD, M. F. & D. W. RHIND (Hrsg.): *Geographical Information Systems: Principles and Applications*, Longman, Essex, S. 319-335.
- MAGUIRE, D. J., GOODCHILD, M. F. & D. W. RHIND (Hrsg.): *Geographical Information Systems. 2 Aufl.*, Longman, London, Section 2(b): Spatial Databases, S. 371-424.
- MILNE, P., MILTON, S. & J. L. SMITH (1993): *Geographical Object-oriented Databases: A Case Study*. In: INT. J. GEOGRAPHICAL INFORMATION SYSTEMS, Band 7, Nr. 1, S. 39-56.
- MOREHOUSE, S. (1992): *The ARC/INFO Geographic Information System*. In: COMPUTERS & GEOSCIENCES, Band 18, Nr. 4, S. 435-442.
- MÜLLER, U. et al. (1992): *Dokumentation zur Methodendatenbank des Fachinformationssystems Bodenkunde*. NIEDERSÄCHSISCHES LANDESAMT FÜR BODENFORSCHUNG (Hrsg.), Technische Berichte NIBIS, Nr. 3, Hannover.
- NADM (2003): *NADM Conceptual Model 1.0, A Conceptual Model for Geologic Map Information*. North American Geologic-map Data Model Steering Committee, Online: <http://ncgmp.usgs.gov/ngmdbproject/reports/> (23.02.2004).
- NIEDERSÄCHSISCHES LANDESAMT FÜR BODENFORSCHUNG (Hrsg.) (2000): *Ausgabeformen der Flächendatenbank des Niedersächsischen Bodeninformationssystems NIBIS*. Arbeitshefte Boden, Heft 2000/1, Hannover.

- OOSTEROM, P. VAN et al. (2002): *The Balance Between Geometry and Topology*. In: OOSTEROM, P. VAN & D. E. RICHARDSON (Hrsg.): *Advances in Spatial Data Handling*, 10th Intern. Symposium on Spatial Data Handling, Ottawa, July 2002, Springer, Berlin.
- OGC (2003a): *The Importance of Going "Open"*. Open GIS Consortium, Online: <http://opengis.org/press/?page=papers> (05.03.2004).
- OGC (2003b): *Data Models and Interoperability*. Open GIS Consortium, Online: <http://opengis.org/press/?page=papers> (05.03.2004).
- PARENT, C., SPACCAPIETRA, S. & E. ZIMÁNYI (1999): *Spatio-Temporal Conceptual Models: Data Structures + Space + Time*. In: MEDEIROS, C. B. et al (Hrsg.): *Advances in Geographic Information Systems*, Proceedings of the 7th ACM Symposium, GIS'99, Kansas City, November 1999, S. 26-33.
- PREUSS, H., VINKEN, R. & H.-H. VOSS. (1991): *Symbolschlüssel Geologie*. 3. Aufl., NLFb/BGR, Hannover.
- PREUSS, H. & H.-H. VOSS (1983): *Die Bohrdatenbank von Niedersachsen*. NLFb/BGR, Hannover, 13 S. (Unveröff.).
- RASMUSSEN, K. (1995): *An Overview of Database Analysis and Design for Geological Systems*. In: GILES, J. R. A. (Hrsg.) (1995): *Geological Data Management*. Geological Society Special Publication Nr. 97, S. 5-11.
- RIGAUX, P., SCHOLL, M. & A. VOISARD (2002): *Spatial Databases. With Application to GIS*. Morgan Kaufman, San Francisco, 410 S.
- RUMBAUGH, J., JACOBSON, I. & G. BOOCH (1999): *The Unified Modeling Language – Reference Manual*. Addison-Wesley, Reading.
- SAUER, H. (1992): *Relationale Datenbanken. Theorie und Praxis. Inklusive SQL-2*. 2. Aufl., Bonn.
- SAUNDERS, M. R., SHIELDS, J. A. & M. R. TAYLOR (1995): *Improving the Value of Geological Data: a Standardized Data Model for Industry*. In: GILES, J. R. A. (Hrsg.) (1995): *Geological Data Management*. Geological Society Special Publication Nr. 97, S. 41-54.
- SBRESNY, J. (1997): *Fehlerquellen in Raumbezogenen Informationssystemen*. BUNDESANSTALT FÜR GEOWISSENSCHAFTEN UND ROHSTOFFE (Hrsg.) (1997), Geologisches Jahrbuch, Reihe F, Heft 33, Hannover.
- SCHMIDT, M. & T. DEMMIG (2001): *SQL GE-PACKT*. mitp, Bonn, 1431 S.
- SHETH, A. P. (1999): *Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics*. In: GOODCHILD, M. et al. (Hrsg.): *Interoperating Geographic Information Systems*, Kluwer, London, S. 5-29.
- SOLLER, D. R. et al. (2002): *The Central Kentucky Prototype: An Object-Oriented Geologic Map Data Model for the National Map Database*. U. S. Geological Survey Open-file Report 02-202, Online: <http://ncgmp.usgs.gov/ngmdbproject/reports/> (23.02.2004).

- THEOBALD, D. M. (2001): *Topology Revisited. Representing Spatial Relations*. In: INT. J. GEOGRAPHICAL INFORMATION SCIENCE, Band 15, Nr. 8, S. 689-705.
- UBEDA, T. & M. J. EGENHOFER (1997): *Topological Error Correcting in GIS*. In: SCHOLL, A. & A. VOISARD (Hrsg.): *Advances in Spatial Databases*, 5th Intern. Symposium, SSD'97, Berlin, July 1997, Lecturer Notes in Computer Science 1262, Springer, S. 283-297.
- VOISARD, A. (1999): *Abduction and Deduction in Geologic Hypermaps*. In: GÜTING, R. H., PAPADIAS, D. & F. LOCHOVSKY (Hrsg.): *Advances in Spatial Databases*, 6th Intern. Symposium, SSD'99, Hong Kong, July 1999, Lecturer Notes in Computer Science 1651, S. Springer, 311-329.
- VOSSMERBÄUMER, H. (1991): *Geologische Karten*. 2. Aufl., Schweizerbart, Stuttgart.
- WORBOYS, M. F. (1999): *Relational Databases and Beyond*. In: MAGUIRE, D. J., GOODCHILD, M. F. & D. W. RHIND (Hrsg.): *Geographical Information Systems*. 2 Aufl., Longman, London, S. 373-385
- WORBOYS, M. F. (1994): *A Unified Model for Spatial and Temporal Information*. In: THE COMPUTER JOURNAL, BAND 37, HEFT 1, S. 26-34
- WORBOYS, M. F., HEARNshaw, H. M. & D. J. MAGUIRE (1990): *Object-Oriented Data Modeling for Spatial Databases*. In: INT. J. GEOGRAPHICAL INFORMATION SYSTEMS, Band 4, Nr. 4, S. 369-383.
- WATSON, P. (2002): *Topology and ORDBMS Technology*. Laser-Scan White Paper. Online: www.radius.laser-scan.com (12.07.03).

Anhang A

Anhang B