

Integrierter Zugriff auf heterogene paläoökologische Datenbestände

Andreas Bergmann¹, Holger Gärtner², Martin Breunig¹,
Armin B. Cremers¹ und Richard Dikau²

Abstract

Datenbestände in den Geowissenschaften zeichnen sich durch unterschiedliche Anforderungen der einzelnen Fachrichtungen aus. Dies bezieht sich sowohl auf verwendete Datentypen als auch auf Formate. Aus der fortschreitenden Technisierung der Datenerfassung und Entwicklung neuer Methoden resultiert eine schnell wachsende Datenmenge, die in sehr heterogener Form vorgehalten wird. Dieser Beitrag beschreibt wesentliche Komponenten eines Systems, das den Zugriff auf heterogene, verteilte Datenbestände unterstützt. Die vorgestellte Client/Server Architektur unterstützt dabei zwei Dimensionen von Erweiterbarkeit. Auf der einen Seite besteht die Möglichkeit, neue Datenquellen mittels objektorientierter Modellierung der jeweils vorhandenen Daten zu erschließen. Auf der anderen Seite können dem Anwender geeignete interaktive Explorationswerkzeuge für neu hinzukommende Datentypen zur Verfügung gestellt werden.

1 Einleitung

In Forschung, Industrie und Verwaltung wird die Notwendigkeit gesehen, zur Entscheidungsfindung und zur Erkennung übergeordneter Zusammenhänge auf verschiedenste Datenbestände zurückgreifen zu können. Derzeit existiert in vielen Bereichen eine sehr große und schnell wachsende Datenmenge, die für unterschiedliche Zwecke erhoben und - entsprechend modelliert - digital vorgehalten wird. Dabei kommen viele sehr unterschiedliche Systeme und Rechnerplattformen zum Einsatz. Dies führt zu einer Isolation der Datenbestände, die in einigen Fällen bis in kleine Organisationsstrukturen durchgreift.

Dieser Zustand ist insbesondere auch für die Geowissenschaften nicht zufriedenstellend. Die Erhebung und Analyse von Daten, gefolgt von fachspezifischen Interpretationen dieser Analysen mit dem Ziel, neue Erkenntnisse über natürliche Prozesse zu gewinnen, ist häufig mit extrem hohen Aufwand an technischer Ausrüstung, Arbeitskraft und Fachwissen verbunden. Darüber hinaus wird der steigende Einsatz

¹ Universität Bonn, Institut für Informatik III, email: andreas@informatik.uni-bonn.de

² Universität Bonn, Geographische Institute, email: holger@slide.giub.uni-bonn.de

neuester Techniken von einer wachsenden Methodenvielfalt wie z.B. dendrogeomorphologische Untersuchungen, hydrologische Modellierung, Kartierungs- oder Datierungsmethoden begleitet (Dikau 1992; Vinken 1992). Dieses führt zu riesigen und schnell anwachsenden Datenmengen, die in verschiedenen Datenbanken in unterschiedlichsten Formaten verwaltet werden. Seit einigen Jahren werden große Anstrengungen unternommen zentrale Datenbanken einzurichten, um Datenverlust zu vermeiden und die Möglichkeit zu schaffen, einmal erhobene Daten für weitere Untersuchungen in verschiedenen Forschungsrichtungen wiederzuverwenden. Weiterhin wurden diese Datenbanken mit spezifischen Anfragemöglichkeiten versehen. Dieser Ansatz vereinfacht zunächst einmal den Zugang zu bestimmten Daten, birgt aber auch einige Probleme, die im folgenden Kapitel näher erläutert werden.

Seit 1996 fördert die Deutsche Forschungsgemeinschaft das Projektbündel IOGIS (Interoperable Geoinformationssysteme), in dem sechs Forschergruppen aus verschiedenen Disziplinen wie Geographie, Geologie, Klimatologie, Fernerkundung und Informatik zusammengeschlossen sind. Die Projekte untersuchen verschiedene Aspekte bezüglich der Heterogenität von Daten und Methoden im Kontext von GIS (IOGIS 1998). Die hier beschriebenen Arbeiten werden im Kontext des IOGIS Teilprojektes OPALIS³ an der Universität Bonn durchgeführt (Bergmann et al. 1998). Ziel dieses Projektes ist zum einen die Modellierung von geowissenschaftlichen Datenbeständen mit Hilfe objektorientierter Techniken und zum anderen die Entwicklung einer offenen GIS Architektur, die den integrierten Zugriff auf autonome, heterogene Datenbestände ermöglicht. Als Pilotanwendung für OPALIS soll ein System entwickelt werden, das Geowissenschaftler bei der Charakterisierung der zeitlichen Entwicklung von Ökosystemen und der Untersuchung der Bedeutung verschiedener Prozesse für heutige Umweltbedingungen unterstützt.

Dieser Aufsatz ist wie folgt gegliedert: der zweite Abschnitt beschreibt beispielhaft einige Szenarien für den integrierten Zugriff auf verschiedene Datenbestände und analysiert, ausgehend von der oben beschriebenen heterogenen Umgebung, die Anforderungen, die sich für den Entwurf von OPALIS ergeben.

Im dritten Abschnitt diskutieren wir dann die für OPALIS entwickelte komponentenbasierte Client/Server Architektur. Neben der Beschreibung der wesentlichen Systemkomponenten gehen wir hier auch auf die Möglichkeiten der Server- und Client-Erweiterung ein.

Der vierte Abschnitt skizziert eine prototypische Anwendung des Systems in den Geowissenschaften. Der Schwerpunkt der Darstellung liegt hier auf den bisher auf der Basis des in Abschnitt 3 beschriebenen Baukastens realisierten Komponenten. Schließlich stellen wir in einem Ausblick zukünftigen Entwicklungsmöglichkeiten von OPALIS vor.

³ Das Projekt OPALIS (Offenes Paläoökologisches Informationssystem) wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert.

2 Anforderungen an die Datenverwaltung

Als Grundlage für den Aufbau des OPALIS-Systems wurde zu Beginn der Arbeiten in Abstimmung mit Fachgruppen des DFG-Schwerpunktprogrammes "Wandel der Geo-Biosphäre während der letzten 15000 Jahre - Kontinentale Sedimente als Ausdruck sich verändernder Umweltbedingungen" (SPP) eine Anforderungsanalyse aus geowissenschaftlicher Sicht erarbeitet.

Die Entwicklung zentraler Datenbanken und daran gekoppelter Anfragesysteme zielt darauf ab, Datenbestände, die oft mit großem technischen und nicht zuletzt finanziellen Aufwand erhoben wurden, zentral, allgemein verfügbar und bezüglich der Datenformate homogenisiert vorzuhalten. Diese Systeme leisten einen großen Beitrag zum Abbau der sogenannten Datenfriedhöfe. Eines der größten Probleme zentraler Datenhaltung ergibt sich allerdings daraus, daß viele potentielle Datenanbieter aus unterschiedlichen Gründen nicht bereit sind, die mit hohem Aufwand erhobenen Daten uneingeschränkt (wenn überhaupt) zur Verfügung zu stellen. Ein wichtiger Punkt ist hierbei die Tatsache, daß alle Daten, die an eine zentrale Datenbank übergeben werden vorgegebenen Formaten und Strukturen entsprechen müssen. Dies bedeutet oft einen erheblichen Aufwand für die Eigentümer der Daten, da die eigenen Strukturen nur sehr selten den Vorgaben entsprechen. Aus diesem Grund ist die Wahrscheinlichkeit des Datenverlustes aufgrund der Tatsache, daß man keine Informationen über die Existenz der Daten erhält weiterhin relativ hoch.

Es wird daher ein System benötigt, das es ermöglicht, mit Hilfe von für den jeweiligen Anwendungsbereich geeigneten interaktiven Zugriffsmöglichkeiten Daten aus verschiedenen heterogenen Datenquellen zu explorieren und so für eigene Projekte alle relevanten Daten einfach zu lokalisieren. Die Datenbestände sollen dabei weiterhin verteilt vorgehalten und autonom verwaltet bleiben. Der Zugriff auf die Datenbestände erfolgt über speziell modellierte Metadaten, also allgemeinen Informationen über die im Verbundsystem vorhandenen Daten, wobei der Eigentümer der Daten von Fall zu Fall über die Freigabe der eigentlichen Daten selbst entscheiden kann. Das System wird folglich nicht weltweit offen, sondern auf verschiedenen Organisationsebenen einsetzbar sein.

Um dies zu verdeutlichen, soll hier beispielhaft die Möglichkeit einer "integrierten" Anfrage auf die verteilten Datenbestände des Systems aufgezeigt werden. Eine Anfrage nach Schwellenwerten zur Auslösung von Massenbewegungen in den (räumlich getrennten) Gebieten A, B und C erfordert, sofern keine direkten Schwellenwertinformationen vorgehalten werden, eine Aufsplittung der komplexen Anfrage in Einzelfragen. Unter anderen wären dies z.B. "Gibt es aus dem Bereich eines Gebietes (oder aller Gebiete) Untersuchungen und Daten über Massenbewegungen; von wem wurden sie wann und wo durchgeführt; um welche Art(en) der Massenbewegung handelt es sich; Häufigkeit des Auftretens (Frequenz/Magnitude); vorhandene Archivinformationen; Bohrinformationen, bodenmechanische und/oder geophysikalische Daten ; Temperatur- und Niederschlagszeitreihen, um nur die wichtig-

sten zu nennen. Die Möglichkeit, diese unterschiedlichen Datentypen (Textinformation, Punkt-, Linien- und Flächendaten, Zeitreihen, etc.) ungeachtet der verwendeten Datenformate anfragen zu können, eröffnet die Möglichkeit, eigene Datenbestände zu erweitern und damit eigene Berechnungen durchzuführen. Werden die entsprechenden Daten nicht direkt freigegeben, erhält man doch zumindest die Information, daß diese Daten existieren und wo sie vorgehalten werden (Metadaten).

Darauf aufbauend wurde die Konzeption einer Client-Server-Architektur mit allgemeiner Server-Funktionalität und fachspezifischen Client-Komponenten entwickelt, die die schnelle und einfache Verwaltung dezentraler Datenbestände erlaubt und die Visualisierung der Daten unterstützt. Die Server-Komponente muß über umfassende Kataloge und Register verfügen, sach- und raumbezogene Recherchen ermöglichen, Zugriffsregelungen sowie Methoden und Anleitungen zur Homogenisierung und Bearbeitung auch themenspezifischer Daten bereithalten.

Desweiteren sollen Methodenmodule Interpolationsverfahren, Raster/Vektorverschneidungen, Ableitungen von Attributen aus Primär- und Proxydaten sowie die Modellierung zeitlicher Veränderungen des Reliefs (3D) oder der Vegetation ermöglichen. Das heißt, die räumliche Differenzierung von Entwicklungen (Klima, Vegetation, geomorphologische Prozesse, etc) in der Zeit muß ermöglicht werden (3D/4D Erweiterung).

Um die Integration heterogener Datenbestände zu ermöglichen, werden die jeweiligen Strukturen der Datenbestände zunächst in objektorientierte Datenmodelle überführt. Diese Objektmodelle bilden die Grundlage für die Repräsentation der Semantik geowissenschaftlicher Daten (siehe Kap.3.1) und werden in enger Zusammenarbeit mit den Verwaltern der einzelnen Datenbestände durch die Systementwickler erstellt. Dadurch entfällt der für die jeweiligen Arbeitsgruppen aufwendige Schritt, die eigenen Datenstrukturen und -formate gemäß spezieller Vorgaben zu verändern.

3 OPALIS Architektur

Um den im vorherigen Abschnitt diskutierten Anforderungen bzgl. autonomer Datenverwaltung und Einsatz in verschiedenen Projekten und Fachrichtungen gerecht werden zu können, ist für OPALIS eine offene und verteilte Architektur entworfen worden.

Wir schlagen eine objektorientierte Architektur vor, die durchgängig aus unabhängig realisierbaren und damit austauschbaren Komponenten zusammengesetzt ist. Dabei stützen wir uns auf den CORBA® Standard (Object Management Group 1997), einer objektorientierten Architektur für den verteilten Zugriff auf heterogene Softwaresysteme. Die Verwendung eines CORBA® kompatiblen Object Request Brokers (ORB) als Kommunikationsinfrastruktur für OPALIS ermöglicht sowohl eine plattform- und programmiersprachenunabhängige Realisierung der einzelnen

Komponenten als auch eine ortstransparente Kommunikation mittels standardisierter Protokolle über das Internet (Balovnev et al. 1998). In folgenden geben wir einen Überblick über die wichtigsten Bausteine des OPALIS Systems, wie sie in Abbildung 1 dargestellt sind.

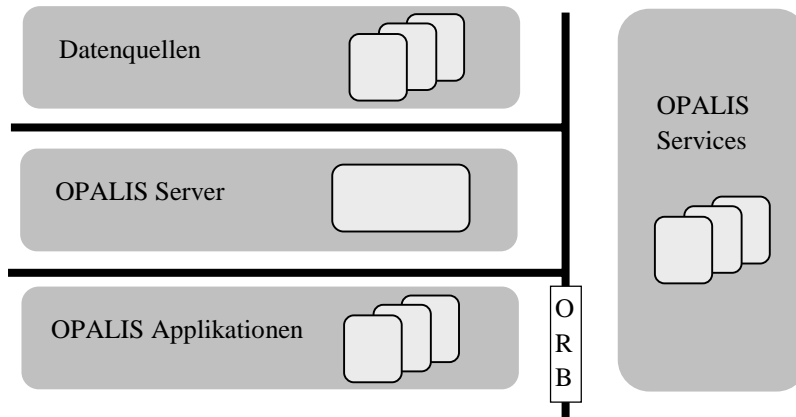


Abbildung 1
Die Hauptkomponenten der OPALIS Architektur

Datenquellen sind die Systeme, die an den verschiedenen Standorten der an einem Informationsverbund beteiligten Organisationen gepflegt werden. Um sie an den OPALIS Server anzuschließen, wird jeweils eine Wrapperkomponente entwickelt, die es ermöglicht die benötigten Informationen zu exportieren. Dabei handelt es sich in erster Linie um Metainformationen der verschiedenen Quellsysteme. Zu diesen zählen u.a. eine Beschreibung der Quelle inklusive des für diese Quelle entwickelten Objektmodells, Informationen über verwendete Standards zur Datenerfassung und die durch die Daten abgedeckte räumliche und/oder zeitliche Ausdehnung. Die aktuell angeforderten Daten werden aus Performanz und Verfügbarkeitsgründen am OPALIS Server temporär in einer Cache-Komponente gespeichert.

Der **OPALIS Server** besteht aus mehreren Komponenten, die zusammen für OPALIS Applikationen den integrierten Zugriff auf die angeschlossenen Datenquellen realisieren. Die Anfrageschnittstelle wird mit Hilfe einer Query-Komponente realisiert, die die in Kapitel 2 beschriebenen Anfragen durchführt und die Ergebnisse in einer für den Client interpretierbaren Form bereitstellt. Die Query-Komponente stützt sich dabei auf einen vorintegrierten Datenbestand, der in der Warehouse-Komponente des OPALIS Servers verwaltet wird (siehe Kap 3.1). Eine Integrationskomponente sorgt schließlich für die Extraktion der benötigten Daten aus den angeschlossenen Quellsystemen und die Generierung der zur Anfrageunterstützung

eingesetzten integrierten Information. Der OPALIS Server soll so gestaltet werden, daß jede Instanz wiederum als OPALIS Datenquelle genutzt werden kann.

Die **OPALIS Services** stellen eine erweiterbare Menge von Komponenten dar, die eine Reihe von Funktionen, geowissenschaftlichen Methoden und Datenservices zur Verfügung stellen. Diese werden vom OPALIS Server vor allem genutzt, um Informationen für die notwendige Konvertierungen während des Integrationschrittes zu bekommen. Zur Zeit stehen u.a. Funktionen zur Umrechnung von Maßeinheiten und verschiedenen kartographischen Projektionen zur Verfügung. Ein Service der Angaben über die Einteilung Deutschlands in topographische Karten im Maßstab 1:25000 verwaltet, ermöglicht z.B. die Abbildung des Namens oder der Nummer einer Karte auf die zugehörige räumlichen Ausdehnung oder umgekehrt bei Angabe einer Koordinate das Auffinden der Informationen über die entsprechende Karte. Die Service Komponenten sind darüberhinaus ohne weiteren Aufwand auf gleiche Weise von einer OPALIS Applikation ausführbar. Hier ist aber in erster Linie der Zugriff auf geowissenschaftliche Methoden von Nutzen.

OPALIS Applikationen sind Anwendungen, die auf spezielle Bedürfnisse von Forschergruppen oder Projekten hin entwickelt werden können. Dabei ist es nicht notwendig einen Neuentwicklung durchzuführen. Vielmehr gibt es die Möglichkeit, wichtige Komponenten, die u.a. die Kommunikation von Client und Server, den Informationsaustausch zwischen Anwendungskomponenten auf dem Clientrechner und die angemessenen interaktive Exploration komplexer Datenstrukturen unterstützen, vom OPALIS Server zu beziehen (siehe Kap. 3.2). Diese können dann im Sinne einer generischen OPALIS Browserapplikation oder als Grundbaustein für eine Eigenentwicklung eingesetzt werden. Diese Vorgehensweise hat mehrere Vorteile. Neben der Entwicklungsunterstützung können OPALIS Applikationen durch Fernwartung ohne jeden weiteren Aufwand von neuen oder verbesserten Komponenten profitieren. Neue Datentypen und Methoden sind, wenn vom Anwender gewünscht, direkt zu benutzen. Um die für die Umsetzung dieser Möglichkeiten benötigte Plattformunabhängigkeit zu erzielen, haben wir Java als Entwicklungsplattform gewählt.

Im allgemeinen steht beim Einsatz von OPALIS die Frage im Vordergrund, wer hat welche Daten, die für eine zu bearbeitende geowissenschaftliche Frage relevant sind, und inwieweit werden diese zur Verfügung gestellt. Deshalb interessieren wir uns bei der Realisierung integrierter Anfragen in erster Linie für Metainformationen (Informationen über Daten in Anlehnung an die Definition des Open GIS Consortium - Open GIS Project No. 97-110R1). Darüber hinaus soll bei entsprechender Verfügbarkeit die Exploration der konkreten Rohdaten ermöglicht werden.

Daraus ergeben sich verschiedene Anforderungen an die Datenverwaltung auf dem OPALIS Server. Zum einen muß der Server um neue Datenquellen erweiterbar sein. Die entsprechenden Metainformationen sollen in angemessener Form und Zeit verfügbar sein. Andererseits muß es möglich sein, neue Datentypen adäquat zu unterstützen und zukünftigen oder bestehenden Anwendungen entsprechende Explora-

tionswerkzeugen anzubieten. Im folgenden gehen wir daher detaillierter auf den OPALIS Server und die Unterstützung, die OPALIS für die Entwicklung von Client-Applikationen bietet, ein.

3.1 Datenverwaltung auf dem OPALIS Server

In diesem Abschnitt werden die Möglichkeiten zur Verwaltung der Metainformationen auf dem OPALIS Server und die Unterstützung, die das OPALIS System für den Anschluß einer weiteren Datenquelle bietet, diskutiert

3.1.1 Die Warehouse-Komponente im OPALIS Server

Für die Speicherung integrierter Daten gibt es im Prinzip zwei unterschiedliche Strategien, die man mit föderierten Datenbanksystemen (Conrad 1997) bzw. Data Warehouse Systemen (Labio et al. 1997) verbinden kann. Föderierte Systeme gehen dabei davon aus, daß alle Daten lokal in den verschiedenen Quellsystemen gespeichert und verwaltet werden. Anfragen werden auf einem Server umformuliert und an die Datenquellen weitergeleitet. Die einzelnen Antworten werden dann gesammelt und entsprechend einem globalen Schema integriert. Auf diese Weise läßt sich die Aktualität der gewonnenen Information garantieren. Bei dieser Vorgehensweise muß man aber mit langen Antwortzeiten und unvollständigen Informationen rechnen, da einzelne Datenquellen zur Anfragezeit möglicherweise nicht verfügbar sind. Der Data Warehouse Ansatz verfolgt eine entgegengesetzte Strategie. Die benötigten Daten werden schon vor einer Anfrage integriert und an einer zentralen Stelle gespeichert. Auf diese Weise können die Reaktionszeiten beim Zugriff auf integrierte Information aus allen Datenquellen relativ kurz gehalten werden. Demgegenüber steht aber, daß inzwischen ungültig gewordene Informationen vorgehalten werden. Darüber hinaus ist dieser Ansatz weniger flexibel. Informationen, die zur Integrationszeit nicht modelliert waren, stehen nicht zur Verfügung.

Für OPALIS haben wir einen Mittelweg gewählt. Aufgrund der sehr großen Datenmenge und bestimmten Zugriffsbeschränkungen, kann der überwiegende Teil der Daten nicht auf einem zentralen Server verwaltet werden. Außerdem ist eine lokale Pflege der eigenen Daten, die für ein Projekt erhoben und analysiert werden aus Verfügbarkeits- und Flexibilitätserwägungen wünschenswert. Dies gilt so nicht für die oben beschriebenen Metainformationen. Ein rein verteilter Ansatz würde hier zu nicht tolerierbaren langen Antwortzeiten führen.

Aktuelle Arbeiten im OPALIS Projekt beschäftigen sich daher mit der Definition und Implementierung von semantischen Zugriffsstrukturen. Diese sollen Anfragen auf im geowissenschaftlichen Kontext interessante Kategorien wie Raum, Zeit und Sachbegriffe unterstützen.

Der Raumbegriff ist beispielsweise in den verschiedenen Datenquellen unterschiedlich modelliert. Dabei finden sich nicht nur Koordinatenwerte, denen unterschiedliche kartographische Projektionen zugrunde liegen, sondern auch Ortsangaben, die sich auf die Nummer oder den Namen eines TK25 Blattes oder bestimmte Gemeinden und Städte beziehen. Der Zeitbegriff kann z.B. je nach Anwendungsgebiet durch Realzeit, unterschiedlich skalierte Zeitintervalle oder mit Angaben die sich nach einem Standard wie dem Symbolschlüssel Geologie richten modelliert sein. Die oben erwähnten semantischen Zugriffsstrukturen sollen durch Ausnutzen von Wissen über Zusammenhänge der vielfältigen Darstellungsmöglichkeiten eine integrierte Sicht auf die Datenquellen unterstützen.

Die Entwicklung dieser semantischen Zugriffsstrukturen setzt eine detaillierte Modellierung der inhaltlichen Struktur eines Datensatzes voraus. Speziell in den Geowissenschaften wird die Modellbildung als Werkzeug genutzt, reale Umweltbedingungen in geometrisch basierten Entitäten darzustellen, um sie in Geo-Informationssystemen bearbeiten zu können. Dieser geometrisch basierte Ansatz, d.h. die Bildung von Entitäten aufgrund ihrer Lagebeziehungen zu einem definierten Ort der Erde, führt zu einer isolierten Betrachtung einzelner Systembereiche, da die Wechselwirkungen mit anderen Systemen ignoriert werden. Folglich kann diese Art der Modellierung nicht als ausreichende Grundlage für die Entwicklung semantischer Zugriffsstrukturen verwendet werden.

Eine stärkere Annäherung an reale Umweltsysteme, und somit auch an die Grundlagen der o.g. Strukturen, wird durch die Beachtung der Semantik der zu definierenden Objekte erreicht, also durch die Beschreibung eines Objektes aufgrund seiner Eigenschaften und Wechselwirkungen mit umgebenden Systemen (Gärtner/Bermann 1999). Die semantisch basierte Objektbildung führt nicht zu einer künstlichen geometrischen Einengung einzeln definierter Entitäten, sie bietet vielmehr neue Möglichkeiten der Verknüpfung derselben durch die Beschreibung systemarer Wechselwirkungen. Sie ermöglicht die Einbeziehung räumlicher und zeitlicher Daten unabhängig von der Geometrie und bildet so die Grundlage für die Modellierung komplexer Prozesse in Raum und Zeit (Raper 1995).

3.1.2 Anschluß weiterer Datenquellen

Um eine neue Datenquelle an den OPALIS Server anzuschließen, muß jeweils eine Wrapper-Komponente realisiert werden. Diese besteht auf der eine Seite aus einer Anfragekomponente, die entsprechende Anforderungen mit Hilfe einer internen Anfragesprache an die zugehörige Quelle weitergibt und die Ergebnismenge als Objekte im entsprechenden Cache ablegt. Als interne Anfragesprache nutzen wir eine Teilmenge der Object Query Language (OQL), die von der ODMG definiert wurde (Cattell/Barry 1997). Die Realisierung dieser Komponente hängt von dem konkret eingesetzten System an der Datenquelle ab. Ein häufig verwendetes Werkzeug für


```

module Source1 {
  class Bohrung {
    relationship Lagedaten lagedaten inverse Lagedaten::bohrung;
    relationship Schichtenverzeichnis schichtenverzeichnis inverse Schichtenverzeichnis::bohrung;
    attribute string Name; ...; attribute double<length/m> Endteufe;
  };
  class Lagedaten ( extent Lagedaten ) {
    relationship Bohrung bohrung inverse Bohrung::lagedaten;
    attribute long<projection/GK> Rechtswert; attribute long<projection/GK> Hochwert;
    attribute short TK25Nr; attribute double HoeheZuNN;
  };
  class Schichtenverzeichnis {
    relationship Bohrung bohrung inverse Bohrung::schichtenverzeichnis;
    attribute string Autor; attribute string Typ;
    relationship list<Schicht> schichten inverse Schicht::verzeichnis;
  };
  class Schicht ( extent Schichten ) {
    relationship Schichtenverzeichnis verzeichnis inverse Schichtenverzeichnis::schichten;
    attribute double<length/m> Untergrenze;
    attribute string Stratigraphie; attribute string Hauptpetrographie; attribute string Nebenpetrographie;
    attribute string Genese; attribute string Farbe; attribute string Zusatz;
  };
};

```

Abbildung 2
Beschreibung einer Datenquelle in erweiterter ODL Notation

die Datenhaltung sind relationale Datenbanken. Diesen Typ eines Quellsystems für OPALIS können wir mit Werkzeugen wie z.B. ObjectDRIVER⁴ anschließen. Zur Zeit wird eine Datenquelle noch vollständig im zugehörigen Cache gespiegelt.

Auf der anderen Seite hängt die Realisierung der Wrapper-Komponente von den in der Quelle abgelegten Datentypen und Formaten ab. Um die Integration der Daten aus verschiedenen Quellsystemen im OPALIS Server durchführen zu können, muß die Semantik der jeweiligen Daten beschrieben sein. Dazu verwenden wir eine Notation, die sich eng an die Object Definition Language (ODL) (Cattell/Barry 1997) anlehnt und einige Erweiterungen aufweist. Zur Zeit besteht vor allem die Möglichkeit, neben der Objektstruktur einer Datenquelle auch Maßeinheiten für Attribute zu beschreiben (s. Abbildung 2). Informationen dieser Art werden von der Integrationskomponente verwendet, um Vergleichswerte zu skalieren.

Für OPALIS wurde ein ODL Parser entwickelt, der eine Objektbeschreibung einer Datenquelle verarbeiten kann. Verschiedene Codegeneratoren erzeugen daraus automatisch die Cache Implementierung und die für die Integrationskomponente geeignete Repräsentation des Quellschemas.

⁴ ObjectDRIVER, an Open Object Wrapper dedicated to Relational Databases Reusing, www.inria.fr/cermics/dbteam/ObjectDriver

3.2 Komponentenbasierter Client-Baukasten

Ein weiteres Ziel von OPALIS besteht darin, den Zugriff auf integrierte Datenquellen so zu gestalten, daß zum einen die Entwicklung von Explorationswerkzeugen unterstützt wird, die den unterschiedlichen Anforderungen verschiedener Anwendungsgebiete gerecht werden. Zum anderen sollen einmal entwickelte Werkzeuge ohne großen Aufwand erweiterbar sein, um vom OPALIS Server neu zur Verfügung gestellte Datentypen und Methoden einbeziehen zu können.

Eine OPALIS Applikation setzt sich aus verschiedenen Komponenten zusammen, die vor allem die Kommunikation mit dem Server, Anfragen sowie Visualisierung und Navigation in der Ergebnismenge realisieren. Auf einem Client Rechner wird nur ein Basissystem installiert, das als Plugin-Host dient und den Daten und Nachrichtenaustausch zwischen den Anwendungskomponenten steuert. Auf dem Server stehen verschiedene Realisierungen der einzelnen Komponenten in einem Baukasten zur Verfügung, die je nach Bedarf einmal auf den Client Rechner heruntergeladen werden und die dort existierende Anwendung ergänzen. Um die hierfür notwendige Plattformunabhängigkeit zu erreichen wurde Java als Implementierungssprache gewählt.

Komponenten existieren auf unterschiedlichen Ebenen im Baukasten. Auf der einen Seite gibt es komplexe Komponenten wie z.B. eine weitgehend konfigurierbare digitale Karte, die sich in vielen Kontexten einsetzen läßt oder ein Werkzeug, das zur Visualisierung von Bohrdaten verwendet werden kann, falls bestimmte Strati-graphieinformationen vorliegen (Abb. 3,4). Auf einer niedrigeren Ebene gibt es sehr einfach strukturierte, aber dafür generische Komponenten, die z.B. die Visualisierung verschiedener Datentypen in der Kartenkomponente steuern. Die für den Anwender transparente Erweiterung der Applikation kann auf diese Weise unterstützt werden. Wird z.B. eine Anfrage gestellt, um alle für ein bestimmtes Gebiet verfügbaren Daten zu finden, können die für die Visualisierung bisher unbekannter Datentypen benötigten Komponenten, wenn gewünscht, automatisch in die Applikation integriert werden.

Aufgrund der engen Integration von Java und Internet, wäre es auch möglich eine OPALIS Applikation in Form von Java-Applets anzubieten. Damit könnte OPALIS ohne zusätzliche Installation von jedem Rechner, der einen Internetanschluß und einen Browser wie z.B. Netscape oder Internet Explorer besitzt, genutzt werden. Hier-von haben wir zunächst Abstand genommen, da die Ladezeiten für Applets in der Regel noch relativ groß sind. Dieser Ansatz kann aber bei der Verwendung des von SUN entwickelten Java-Plugins ohne großen Aufwand realisiert werden.

4 Geowissenschaftliches Anwendungsbeispiel

An dieser Stelle soll beispielhaft eine Anfrage auf Bohrinformationen aufgezeigt werde, die zur Rekonstruktion von Paläooberflächen herangezogen werden können. Die im folgenden abgebildeten Screenshots (Abb. 3, 4) beinhalten konkret benötigte Client Komponenten (georeferenzierte Karte, OPALIS-WellTool, etc.), die unter anderem die Visualisierung der Ergebnisse unterstützen.

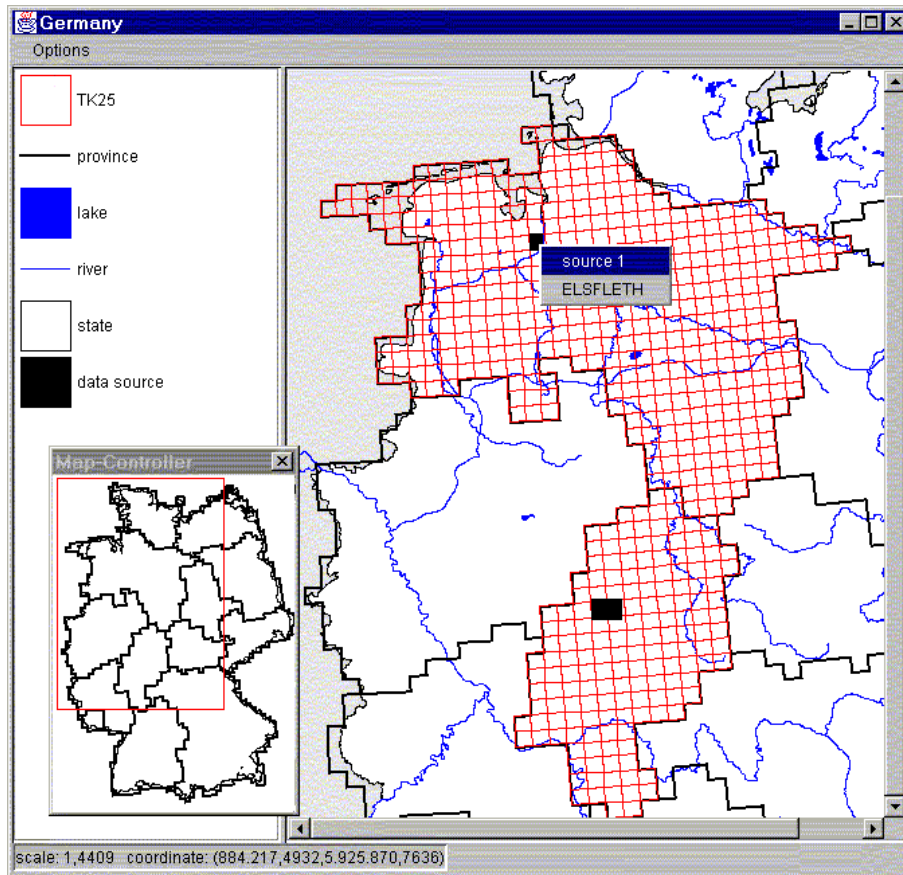


Abbildung 3

Ergebnis einer Anfrage nach räumlicher Verteilung von Bohrungen in den angeschlossenen Datenquellen

Der erste Schritt der Anfrageauswertung bezieht sich auf das Vorhandensein von Bohrdaten in für die Fragestellung interessanten Gebieten. Die Darstellung der Ergebnisse (Abb. 3) erfolgt zunächst in Form einer Karte, in der neben der Markierung

der laut Analyse der Datenbestände relevanten Verteilungsbereichen der Bohrdaten und den entsprechenden Datenquellen weitere Informationen georeferenziert abgreifbar sind (z.B. Name + Nummer einzelner TK25, Koordinaten, Städte, Flüsse, etc.).

Im zweiten Schritt besteht die Möglichkeit, aus der Auflistung der beinhalteten Bohrungen einzelne (oder auch mehrere) Bohrungen zu selektieren (Abb. 4). Das OPALIS-WellTool bietet die Möglichkeit, alle neben der Stratigraphie verfügbaren

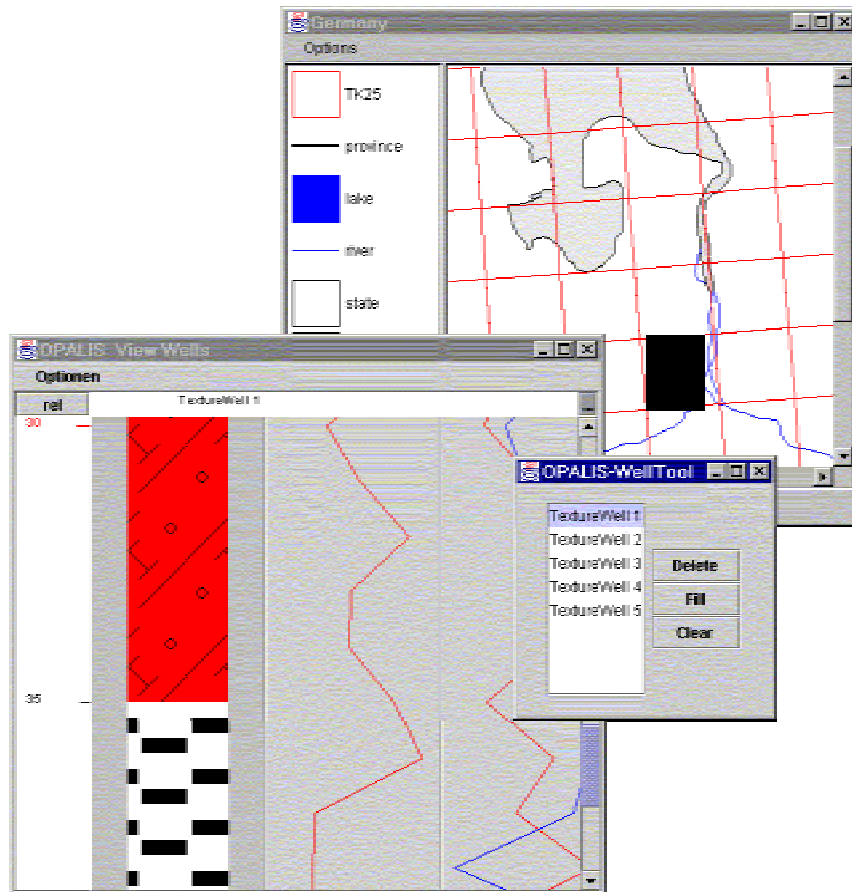


Abbildung 4

Anfrage auf eine bestimmte Bohrung und Darstellung der Daten

Daten zu visualisieren. Desweiteren besteht über die Nutzung des Map-Controllers die Möglichkeit, über die Karte in Bereiche einzelner TK25 zu zoomen und, soweit

in den Datenbeständen vorhanden, die Bohrlokalitäten mit digitalen Karten in Verbindung zu bringen.

5 Ausblick

Bei der Weiterentwicklung von OPALIS werden die Möglichkeiten für Anfragen die Raum-, Zeit- und Sachkriterien umfassen sukzessive weiter ausgebaut.

Bisher haben wir uns in Bezug auf die Datenquellen auf Bohrungsdaten beschränkt. In dem Maße, in dem in Zukunft auch 2D-Daten (z.B. Karten) und 3D-Daten (z.B. Rutschungskörper) eine größere Rolle spielen, werden sowohl die Visualisierungs- und Interaktionsmöglichkeiten durch entsprechende Client-Komponenten als auch der OPALIS Server um eine 3D Datenverwaltung erweitert (Balovnev et al. 1999). Die laufenden Arbeiten konzentrieren sich unter anderem auf eine 3D Darstellung der Bohrungen aus angeschlossenen Datenquellen. Die unter Nutzung von VRML durchgeführte 3D-Beschreibung erlaubt in Zusammenhang mit der für die wichtigsten Webbrowser verfügbaren Software in dargestellten Szenen zu navigieren. So wird es möglich sein, die Bohrungsdaten im einzelnen zu untersuchen und diese direkt mit anderen, spezialisierten Client-Komponenten weiter zu verarbeiten. Interessanter wird diese Nutzung der Darstellung dann, wenn weitere echte 3D Datentypen wie ein digitales Höhenmodell oder in 3D modellierte Prozesse in die Darstellung integriert werden.

Danksagung

Für die Unterstützung unserer Arbeiten möchten wir uns insbesondere bei den IGBP Pilotgruppen W. Andres und J. Wunderlich (Universitäten Frankfurt/Marburg), H. Streif und Ch. Hoselmann (NLF Hannover) sowie G. Wagner und A. Lang (MPI und Universität in Heidelberg) bedanken. Unser Dank gilt weiter Marcus Assmann, Christian Michels und Uwe Radetzki für die Unterstützung bei der Implementierung und Aufbereitung der Daten.

Literaturverzeichnis

- Balovnev, O., Bergmann, A., Breunig, M., Cremers, A.B., Shumilov, S. (1998): A CORBA-Based Approach to Data and Systems Integration for 3D Geoscientific Applications, in: Proceedings of International Conference on Spatial Data Handling - SDH'98, Vancouver, Canada
- Balovnev, O., Bergmann, A., Breunig, M., Cremers, A.B., Shumilov, S. (1999): Remote Access to Active Spatial Data Repositories, in: Proceedings of the 1st International Workshop on Telegeoprocessing, Lyon, France

- Bergmann, A., Gärtner, H., Breunig, M., Cremers, A.B., Dikau, R. (1998): Design and First Steps in the Development of OPALIS, in: Fritsch, D., Englich, M., Sester, M. (Hrsg.): Proceedings of the ISPRS Commission IV Symposium on "GIS - Between Visions and Applications", Stuttgart, Germany, IAPRS, 32, Part 4
- Conrad, S. (1997): *Föderierte Datenbanksysteme – Konzepte der Datenintegration*, Berlin et al.
- Dikau, R. (1992): Aspects of constructing a digital geomorphological base map, in: *Geol. Jahrbuch*, A122, S. 357-370
- Gärtner, H., Bergmann, A. (1999): Object-oriented modeling of Geo data as a basic tool for the integration of heterogeneous paleoecological information, in: *Computers and Geosciences* (in review)
- IÖGIS (1998): *IÖGIS – Vision einer neuen Generation interoperabler GIS*, Bericht der teilnehmenden Gruppen der Universitäten Berlin, Bonn, Berlin/Bonn, Freiburg, Münster und Stuttgart
- Labio, W.J., Zhuge, Y., Wiener, J.L., Gupta, H., Garcia-Molina, H., Widom, J. (1997): The WHIPS Prototype for Data Warehouse Creation and Maintenance, in: Proceedings of the ACM SIGMOD Conference, Tuscon, Arizona
- Cattell, R.G.G., Barry, D.K. (1997): *The Object Database Standard: ODMG 2.0*, San Mateo/CA
- Object Management Group (1997): *CORBA 2.0/IIOP Specification*, OMG formal document 97-09-01, <http://www.omg.org/corba/c2indx.htm>
- Raper, J.F., Livingstone, D. (1995): The development of a spatial data explorer for an environmental hyperdocument, in: *Environment and Planning*, B 22, 6, S. 679 -687
- Vinken, R. (Hrsg.) (1992): *From Geoscientific Map Series to Geo Information Systems*, in: *Geol. Jahrbuch*, A 122, Hannover