

Das Datenbanksystem LADIB

Umweltforschungsprojekte erfordern ein koordiniertes Daten- und Informationsmanagement. Die zentrale Rolle spielt dabei ein Datenbanksystem. Dessen Aufbau, die Qualität seiner Daten und den projektweiten Datenfluß stellt dieser Artikel am Beispiel des Sonderforschungsbereiches „Landnutzungskonzepte für periphere Regionen“ vor.

1 Entstehung und Merkmale von LADIB

Im Sonderforschungsbereich (SFB) 299 „Landnutzungskonzepte für periphere Regionen“ der Deutschen Forschungsgemeinschaft (DFG) wurde von Beginn an ein projektweites Daten- und Informationsmanagement geführt. Erfahrungen anderer interdisziplinärer Grossforschungsprojekte zeigten, dass ein Datenmanagement frühzeitig und konsequent erfolgen muss, um die effiziente Zusammenarbeit der einzelnen Forschergruppen zu fördern. Mitentscheidend für die Einrichtung des SFB war daher die überzeugende Darstellung und Ausführung der Kooperation im teilprojektübergreifenden Datenverbund.

Das Datenmanagement sollte in erster Linie den multilateralen Datenaustausch erleichtern sowie die Verwendungsmöglichkeiten von Forschungsergebnissen dauerhaft gewährleisten. Um diese Ziele zu erreichen wurde das Datenbanksystem LADIB entwickelt. Dieses Akronym steht für LAhn-DIill-Bergland, der Bezeichnung für den Naturraum der Untersuchungsregion des SFB 299.

2 Nutzwert von LADIB

Zu Beginn des Datenbankprojekts LADIB war der Nutzen eines gemeinsamen Datenbanksystems (DBS) umstritten, da der Einsatz eines zentralen Datenmanagements in Grossforschungsprojekten sich bisher kaum etablieren konnte. Es hat sich mittlerweile jedoch gezeigt, dass neben der Förderung der Zusammenarbeit und der langfristigen Datensicherung weitere Aspekte für ein übergreifendes Datenmanagement sprechen, die im Folgenden näher erläutert werden (Tab. 1).

Datentransfer: Über das zentrale Datenbanksystem werden alle Teilprojekte von Verwaltungsaufgaben entlastet, die durch die lokale Koordinierung des Datentransfers zwischen den Arbeitsgruppen entstehen würden. Darüber hinaus lässt sich ein projektweites Metadaten system zentral besser koordinieren als wenn es auf Teilprojekte verteilt ist. Auch der Austausch von Daten mit anderen Forschungsprojekten oder administrativen Einheiten wird mit Hilfe von LADIB erleichtert, so dass der Ausbau des Datentransfers zu einem Forschungsextranet möglich wird. Die genannten Vorteile des zentralen Datenbanksystems bewirken, dass in den 20 Teilprojekten die Personalressourcen den eigentlichen Forschungsaufgaben zur Verfügung stehen.

Durch das einheitliche Datenformat und -management werden projektspezifische Ergebnisse für die Einbindung in

Nutzwert	Beispiele
Datentransfer	#Datenflußzwischen Teilprojekten #Datentransfer mit externen Stellen
Archivierung	#langfristiger Erhalt von Daten #Dokumentation von Daten
Sicherheit	#Schutz vor Datenverlust #Kontrollierter Zugriff auf LADIB
Effizienz	#Wartungsaufwand wird verringert #Programmservice
Projektplanung	#verbesserte Zielorientierung #gezielte Projektdarstellung
Innovation	#neue Auswertungsverfahren #Data Mining

Tab. 1: Nutzwerte von LADIB für den Sonderforschungsbereich 299.

das künftige Umweltinformationssystem (UIS) des SFB vorbereitet. Werden die Ergebnisse und Schlussfolgerungen des SFB 299 auf andere Untersuchungsgebiete übertragen, so brauchen entsprechende Partner nicht mit unterschiedlichen Dateiformaten oder -typen zu arbeiten, sondern erhalten standardisierte Daten. Der Datentransfer ist die kurzfristigste Aufgabe von LADIB.

Archivierung: Indem Datenbankobjekte (DBO) aus den Teilprojekten zentral gesammelt werden, lässt sich deren Verfügbarkeit langfristig sicherstellen. Forschungsergebnisse verschwinden nicht in einer Schublade, wo sie dann vergessen und Jahre später unter Umständen nicht mehr interpretierbar sind. In diesem Zusammenhang ist insbesondere die Verfügbarkeit verbindlicher Dateitypen wichtig, um die Lesbarkeit der Datenbank SFB DAT zu erhalten. Die in LADIB auftretenden Dateitypen werden gegenwärtig sechs Formatklassen zugeordnet (Abb. 1).

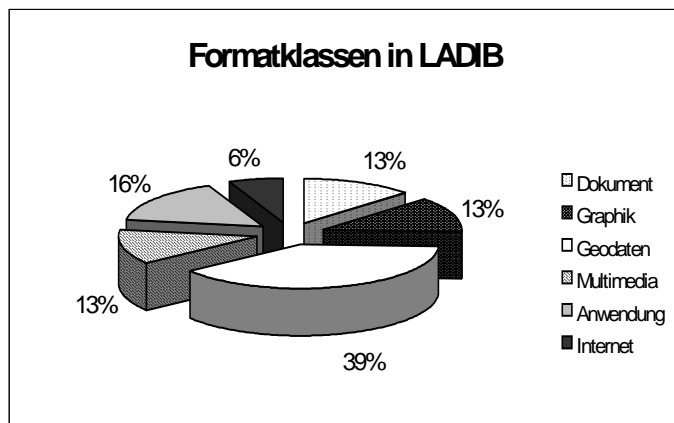


Abb. 1: Prozentualer Anteil der in LADIB auftretende Dateiklassen.

Zur Klasse Dokument gehören Dateien, die vorwiegend Textinformationen enthalten. Dazu zählen unter anderem Dokumente, die in MS-Word verfasst wurden, oder ASCII-Texte, die mit einem einfachen Texteditor erstellt wurden. Im Gegensatz dazu werden Dateien, welche grafische Informationen enthalten, zur Dateiklasse Graphik gezählt. Im Augenblick sind die meisten Datenbankobjekte in LADIB Geodaten, welche mit einem GIS (ERDAS, ArcInfo, ArcView) erzeugt wurden. Schließlich existieren auch Geodaten im PC-Map Format. Als Multimediadaten werden Formate bezeichnet, deren Ziel die anschauliche Präsentation eines Sachverhalts ist. Beispielsweise liegt eine mit ArcView erstellte Animation vor, die einen virtuellen Überflug der Untersuchungsregion zeigt. Zu den Anwendungsdateien zählen Programmdateien, die für die Verarbeitung von Benutzerdaten verwendet werden können. Die Klasse Internetformate steht für Dateien im Hypertext-Format und stark komprimierte Binärdateien für die Darstellung im Internet. Dabei wird in Kauf genommen, dass sich Überschneidungen zur Klasse Dokument und Graphik ergeben können. Normalisierte Daten werden hier nicht gesondert angeführt, da sie mit Oracle verwaltet werden.

Obschon sechs Datenklassen vorliegen, bleibt die weitere Normierung innerhalb jeder Klasse eine wichtige Aufgabe. Beispielsweise finden sich in der Klasse Dokumente die Formate .txt, .doc, .ps und .pdf. Zukünftig wird angestrebt, LADIB Dokumente auf das pdf- und ASCII-Format zu beschränken, um den Nutzer auf diese Weise mögliche Konvertierungsprobleme zu ersparen.

Neben der Reduktion akzeptierter Datenformate gewährleistet die Dokumentation der Daten, dass Ergebnisse aber auch Rohdaten über das Teilprojekt hinausgehend verständlich und dauerhaft nutzbar bleiben. Beides sind Aspekte der Archivierung, die mit Hilfe des Datenbanksystems erleichtert werden.

Sicherheit: Bei der Frage der Sicherheit von Forschungsergebnissen wird häufig an Missbrauch von Daten oder Datendiebstahl gedacht. Zu Beginn der ersten Projektphase des SFB wurden Bedenken dieser Art diskutiert. Seither erfolgt der Zugriff auf SFB-Daten passwortgeschützt. Darüber hinaus ist eine Differenzierung der Zugriffsrechte möglich. Ein anderer Sicherheitsaspekt besteht darin, unerwünschte Daten aus LADIB fern zuhalten, um so die Leistungsfähigkeit und Qualität des DBS aufrecht zu

erhalten. Aus diesem Grund erfolgt eine Datenprüfung bevor die Daten in LADIB übertragen werden. Schließlich unterliegen die Daten in LADIB einer kontinuierlichen Datensicherung, sobald sie in LADIB vorliegen.

Effizienz: Der Wartungsaufwand für ein projektweit verfügbares Datenbanksystem wie LADIB ist erheblich niedriger als für mehrere, verschiedene Datenbanksysteme, die auf unterschiedliche Teilprojekte verteilt sind. Das zentrale DBS verhindert auch, dass teilprojekt-spezifische Datenformate entstehen, deren Verwendung und Archivierung dem jeweiligen Teilprojekt Verwaltungszeit abfordert. Des weiteren bleiben durch die zentrale Bereitstellung selten benötigter Daten, die aber viel Platz beanspruchen, Speicherplatzressourcen auf den lokalen Rechnern erhalten. Schliesslich vermag das Datenmanagement den Datentransfer während experimenteller Arbeitsspitzen in den Teilprojekten besser aufrechtzuerhalten als eine verteilte Datenhaltung.

Projektplanung: Ein zentrales Datenbanksystem wie LADIB erlaubt die Zusammenschau der Ergebnisse des Gesamtprojekts. Defizite werden erkannt und können frühzeitig in Projektbereichstreffen thematisiert werden. Dadurch kann die Zielrichtung des gesamten Vorhabens laufend evaluiert und bei Bedarf rechtzeitig vor einer neuen Antragstellung optimiert werden. Letztlich erlaubt der gemeinsame Datenpool eine überzeugende Aussendarstellung der Projektfortschritte.

Innovation: Während im betrieblichen Data Mining die Kapitaleffizienz gesteigert werden soll, kann das Data Mining in Forschungsprojekten bis dahin übersehene wissenschaftlich relevante Zusammenhänge aufdecken. Mit der zentralen Datenbasis aus LADIB lässt sich daher Wissen finden, welches verborgen bliebe, wenn die Projektdaten unkoordiniert auf mehrere Teilprojekte verteilt wären. Neue Auswerteverfahren, die erst zukünftig entstehen, oder deren Entwicklung durch Datenbanksysteme wie LADIB angestossen werden können, sind weitere Chancen, die ein zentrales Datenbanksystem bietet.

Die genannten sechs Nutzwertaspekte von LADIB zeigen Effizienzsteigerungen für die Forschungsprozesse im Umweltbereich auf, die sich durch ein konsequentes Datenmanagement ergeben.

3 Die Struktur von LADIB

Ein Datenbanksystem (DBS) besteht in der einfachsten Konfiguration aus einer Datenbank (DB) und einem Datenbankmanagementsystem (DBMS) (Heuer und Saake, 1995). LADIB verwendet ORACLE unter dem Betriebssystem SOLARIS als Datenbankmanagementsystem und nutzt die Möglichkeiten dieses UNIX-Derivats zum benutzer- und datenspezifischen Management der Projektdaten in einem Filesystem mit einer hierarchischen Verzeichnisstruktur.

In der ORACLE-Datenbank werden Daten abgelegt, die schnell verfügbar sein müssen, einer unmittelbaren Auswertung bedürfen und die besonders hohen Sicherheitsanforderungen unterliegen. Die bisherige Erfahrungen zeigen jedoch, dass die Erfordernisse des SFB 299 derzeit noch weitgehend mit dem unter SOLARIS erzeugten Filesystem (SFB DAT) und der darin angelegten Verzeichnisstruktur befriedigt werden können. Dies erklärt

sich aus der Heterogenität der auftretenden Dateitypen, der gegenwärtig noch überschaubaren Zahl von Datenobjekten sowie der geringen Zugriffshäufigkeiten der Teilprojekte auf die Datenbanken von LADIB.

In LADIB werden grundsätzlich Datenbestände vorgehalten, die für mehr als zwei Teilprojekte von Interesse sind. Dies erweist sich als ein effizienter Weg mit großen Datenobjekten zu arbeiten, ohne die Netzressourcen unnötig zu belasten (Braun et al., 2000). Die Datenbanken von LADIB bilden eine für SFB Angehörige plausible Struktur, welche das Auffinden von Daten erleichtert und Erweiterungen des Verzeichnisbaums problemlos zulässt (Abb. 2).

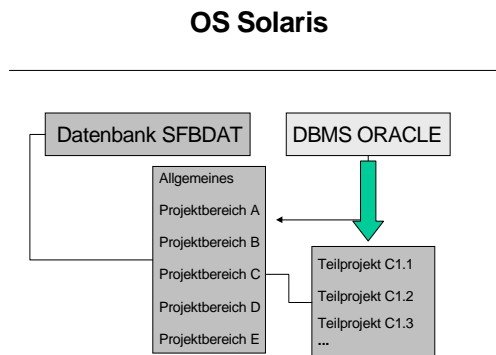


Abb. 2: Struktur des Datenbanksystems LADIB im Sonderforschungsbereich 29.9

Die Datenbank SFB DAT ist in Unterdatenbanken gegliedert. So befindet sich beispielsweise in LADIB eine Unterdatenbank Projektbereich C, die wiederum zwei Teildatenbanken enthält. Es handelt sich um die Teildatenbanken C1 und C2, welche die teilprojektspezifischen Dateien bzw. Datenobjekte enthalten. Diese Benennung entspricht dem internen Aufbau des SFB 299 und ist daher den Nutzern unmittelbar einsichtig. Sie ist überdies leicht um neue Teildatenbanken zu erweitern. Die genannten Datenbankebenen (Unterdatenbank, Teildatenbank) von SFB DAT sind Verzeichnisse mit Daten aus den Teilprojekten. Diese Projektdaten werden nach Maßgabe der Teilprojekte in die jeweilige Datenbank aufgenommen und mit einem Mindestmass an Metainformationen versehen. Projektdaten gelten als Datenobjekt, sofern sie wenigstens eine Datei formen.

Metadaten und Datenqualität: Für die Beschreibung und Einordnung der Daten des SFB sind Metainformationen zu den einzelnen Datenobjekten in LADIB notwendig. Diese werden in einer standardisierten Metadatei bzw. einem Metadatenkopf für jedes Datenobjekt durch das entsprechende Teilprojekt erzeugt. In Analogie zum Begriff Datenbankobjekt wird nachfolgend das Wort Metadatenobjekt verwendet. Gemeinsam mit dem Datenobjekt bildet es ein reguläres Datenbankobjekt. Im Hinblick auf die längerfristige Nutzbarkeit der Datenobjekte oder deren Verwendung für Projekte andere Arbeitsgruppen ist eine solche, einheitliche Dokumentation in einem Metadatenobjekt unumgänglich. Das Metadatenobjekt besteht

aus einem Kern, der von allen Teilprojekten ausgefüllt wird und einem variablen Bereich, der datenobjektspezifisch ist (Braun, 2000).

Nicht alle Objekte der Datenbank erfüllen die Anforderung eines regulären Datenbankobjektes. Aufgrund der bisherigen Erfahrung wurde daher ein Qualitätsmodell für irreguläre und reguläre Datenbankobjekte formuliert (Abb. 3). Die irregulären Datenbankobjekte in LADIB unterscheiden sich in die drei Stufen Datei, nicht erläutertes Datenbankobjekt, und erläutertes Datenbankobjekt. Die vierte Stufe bilden die regulären Datenbankobjekte.

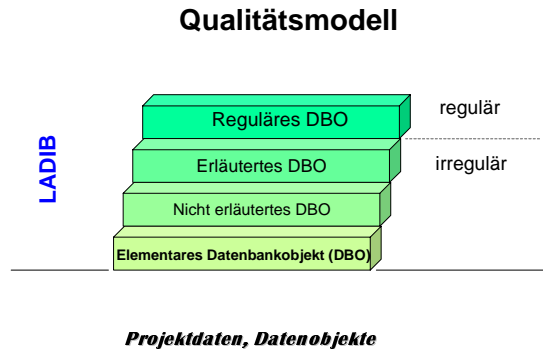


Abb. 3: Qualitätsmodell für die Daten des Sonderforschungsbereiches 299.

Auf der untersten Stufe der irregulären Datenbankobjekte (DBO) befindet sich die Datei ohne weitere Erläuterung als grundlegendes Element der Datenbank. Derartige DBO gelangen in LADIB, wenn beispielsweise mehrere Dateien nur eine gemeinsame Textdatei mit Metainformationen besitzen. Sie bilden jedoch keinen Verbund, im Sinne eines Informationsobjekts, wie es z.B. das Zwischenergebnis floristischer Untersuchungen innerhalb einer Vegetationsperiode darstellt.

Das sogenannte nicht erläuterte DBO ist eine Einheit, welche selbsterklärend ist und aus diesem Grund keine Metainformation benötigt. Ein solches Objekt kann aus mehreren Dateien bestehen, die erst gemeinsam betrachtet die relevante Information vermitteln, wie z.B. Karten, die von einem Geographischen Informationssystem (GIS) erzeugt wurden und in denen die Informationen für Geometrien und Sachdaten meist getrennt vorgehalten werden (z.B. ArcInfo). Das nicht erläuterte DBO kann aber auch aus nur einer Datei - beispielsweise einem ausführbaren Programm - bestehen, das kompiliert vorliegt.

Die dritte Stufe der Datenqualität in LADIB bilden die erläuterten DBO, welche im Gegensatz zur zweiten Stufe wenigstens einen erläuternden ASCII-Text aber noch kein Metadatenobjekt besitzen. Gerade bei stark gegliederten Datenobjekten wie etwa dem Geostatistikprogramm GSLIB würde die Dokumentation jeder Routine bzw. Datei anhand eines Metadatenobjektes einen erheblichen Aufwand verlangen und dennoch keine Verbesserung der Objektbeschreibung bedeuten. Daher wird nur ein beschreibender ASCII-Text für alle GSLIB-Programmdateien erzeugt.

Die vierte Stufe des Qualitätsmodells ist das reguläre Datenbankobjekt, bestehend aus Daten- und Metadatenobjekt. Eine wichtige Aufgabe des Datenmanagements stellt die Kontrolle und Evaluierung dieses vierstufigen Qualitätsmodells dar. Um die Datenqualität sicherzustellen, bedarf es der Mitarbeit der Nutzer. Daher wird LADIB kooperativ geführt, um so die Motivation der Teilprojekte zur Qualitätssicherung zu erhöhen.

Zugriffsregeln sind notwendig, damit die Datenstruktur von LADIB vor nicht autorisierten Veränderungen geschützt bleibt. Der Zugriff auf LADIB erfolgt über WWW-Browser. Mit Hilfe des FTP-Protokolls kann jeder autorisierte Nutzer Daten von LADIB auf seinen Rechner herunterladen. Daneben sind weitere Werkzeuge zum Download einsetzbar, wie SAMBA, WS_FTP und einfache UNIX Kommandos (Braun et al., 2000). Der Upload von Datenobjekten nach LADIB wird dagegen vom Datenbank- bzw. Systemadministrator gesteuert. Auf diese Weise erfolgt eine Qualitätskontrolle der DBO vor deren Übernahme in das Datenbanksystem des SFB 299.

4 Entwicklung von LADIB

Eine im Vorfeld des Sonderforschungsbereichs durchgeführte Pilotstudie zeigte, dass relationale Daten eine dominierende Bedeutung in der zentralen Datenbank des SFB haben würden. Diese Ergebnisse entsprechen bislang nicht der tatsächlichen Entwicklung von LADIB. Da sich aber in der Studie seinerzeit auch abzeichnete, dass nichtrelationale Daten anfallen könnten, entschied sich das Datenmanagement für ein objektrelationales DBMS. Seitens der Teilprojekte wurde darauf hingewiesen, dass eine zu stark formal fixierte Daten- und Methodenbank zu Beginn der Projektlaufzeit kontraproduktiv wirken könnte. Aufgrund dieser Einwendungen erfolgte zunächst nur eine Spezifizierung der Metadatenstruktur, die im Laufe des Projektfortgangs weiter entwickelt wurde (Braun, 2000). Erste SFB-relevante Daten trafen bereits im zweiten Quartal 1997 beim zentralen Datenmanagement ein und wurden zunächst in einer Filesammlung gelagert. Von dort gelangten sie in das im November 1997 freigegebene Datenbanksystem LADIB. Weitere Meilensteine der Entwicklung von LADIB waren dessen Internetanbindung 1998 und seine Vorbereitung für ein Umweltinformationssystem (UIS) im Jahre 1999. Gegenwärtig wird der Ausbau des UIS mit LADIB als eine von mehreren Komponenten vorangetrieben. Eine Volltextsuchmaschine wurde Ende 2000 entwickelt. Die

Eingabe der Metadaten erfolgt seit dem Frühjahr 2001 über ein strukturiertes Formular.

5 Der Datenbestand in LADIB

Die Evaluierung von LADIB im Herbst 2000 zeigt, dass die Möglichkeiten des DBS durch die Arbeitsgruppen unterschiedlich stark genutzt werden. Aus Sicht des Datenbankadministrators lassen sich Besonderheiten im Datenfluss erkennen, die eine Interpretation der Geschäftsprozesse in den Teilprojekten zulassen. Durch diese Betrachtung erlangt der Datenbankmanager die Kompetenz, um auf die Bedürfnisse der Anwender einzugehen. Dies ist eine Voraussetzung für die aktive Betreuung der Benutzer von LADIB. Betrachtet man die Nutzung von LADIB unter dem Aspekt des Speicherplatzbedarfs, so heben sich wenige Teilprojekte hervor (Abb. 4).

Lediglich fünf Teilprojekte stellen rund 95% des Datenvolumens von LADIB. Von daher ist das Datenmanagement stark durch die Anforderungen einer Minderheit im SFB geprägt. In dieser Situation dürfen die restlichen 15 Teilprojekte nicht vernachlässigt werden. Die Dominanz eines Projektbereichs ist nicht zu erkennen. Betrachtet man diese Daten aus dem Blickwinkel des Qualitätsmodells (Kap. 3), so relativieren sich die Beiträge der einzelnen Teilprojekte (Abb. 5). Führt man die Aussagen der Abbildungen 4 und 5 zusammen, ist zu erkennen, dass die Teilprojekte A2, C1.1 und D3 offenbar nur wenige, aber dafür sehr große Dateien in LADIB einbrachten. Im Teilprojekt A2 sind dies rasterbasierte Geodaten aus einem Vorläuferprojekt des SFB, welches inzwischen abgeschlossen ist. Bei den anderen beiden Teilprojekten sind es speicherintensive Tabellen- und Grafikelemente, die einen erheblichen Teil der Systemressourcen beanspruchen. Entsprechend wird der Betreuungsaufwand für die Teilprojekte unterschiedlich bemessen.

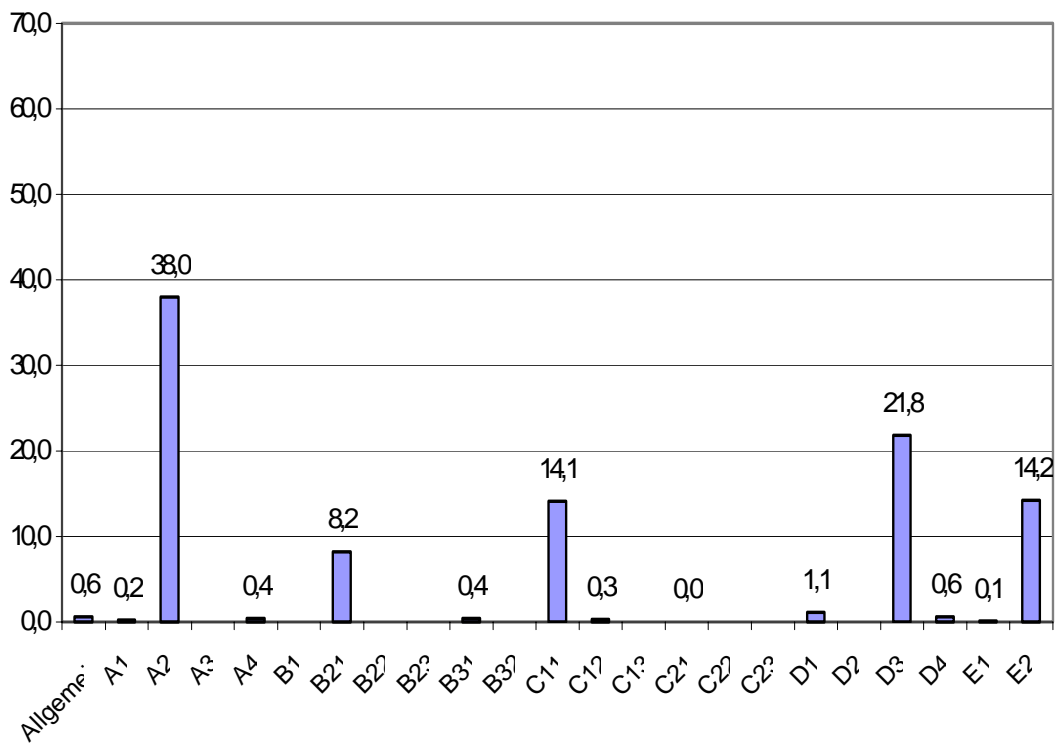


Abb. 4: Speicherplatzbedarf der Teilprojekte des SFB an LADIB in Prozent (100% = 252 MB).

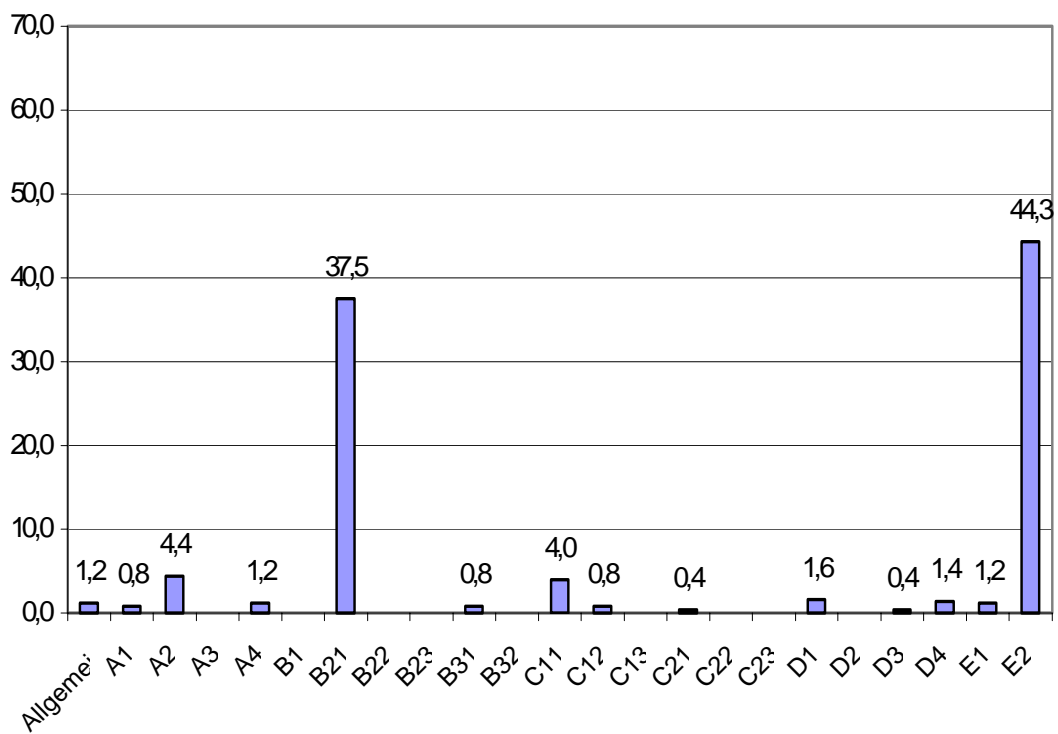


Abb. 5: Prozentanteile der SFB-Teilprojekte in LADIB gemessen an der Dateizahl.

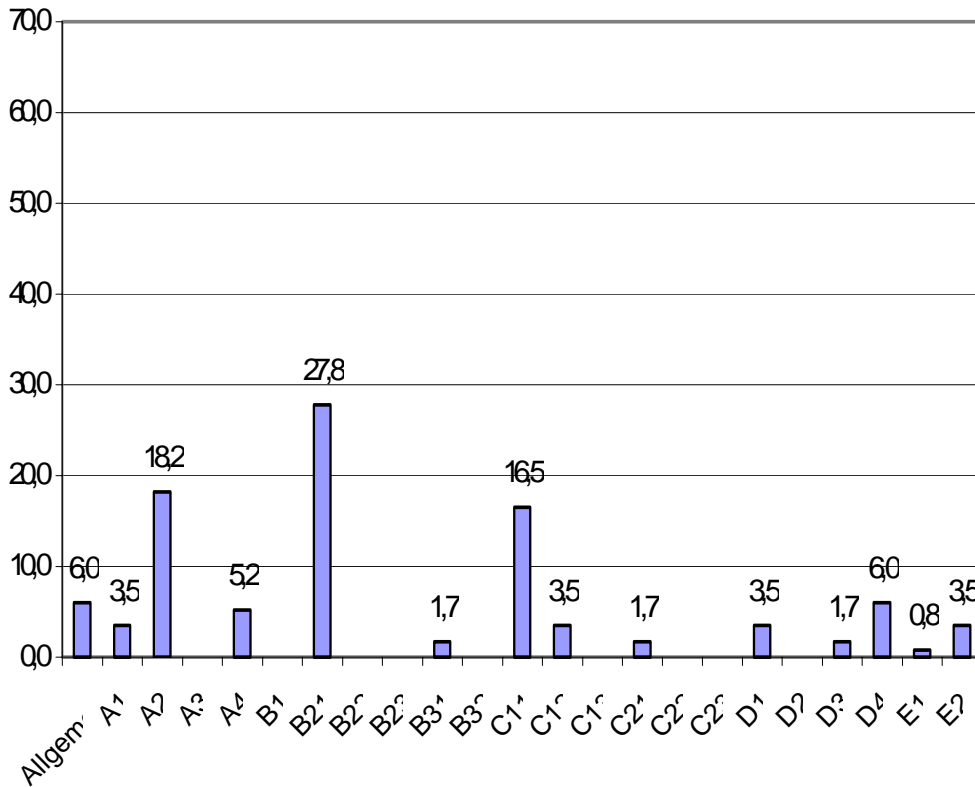


Abb. 6: Der prozentuale Anteil nicht erläuterter Datenbank-objekte in LADIB, aufgliedert nach Teilprojekten.

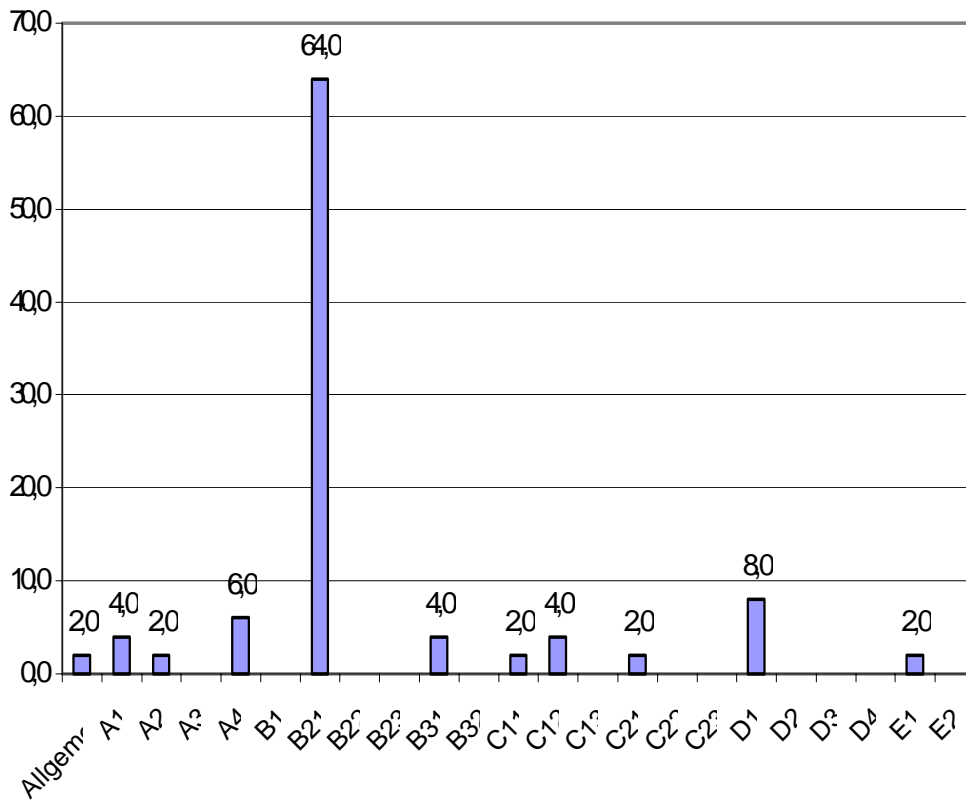


Abb. 7: Darstellung des prozentualen Anteils erläuterter Datenbankobjekte in LADIB.

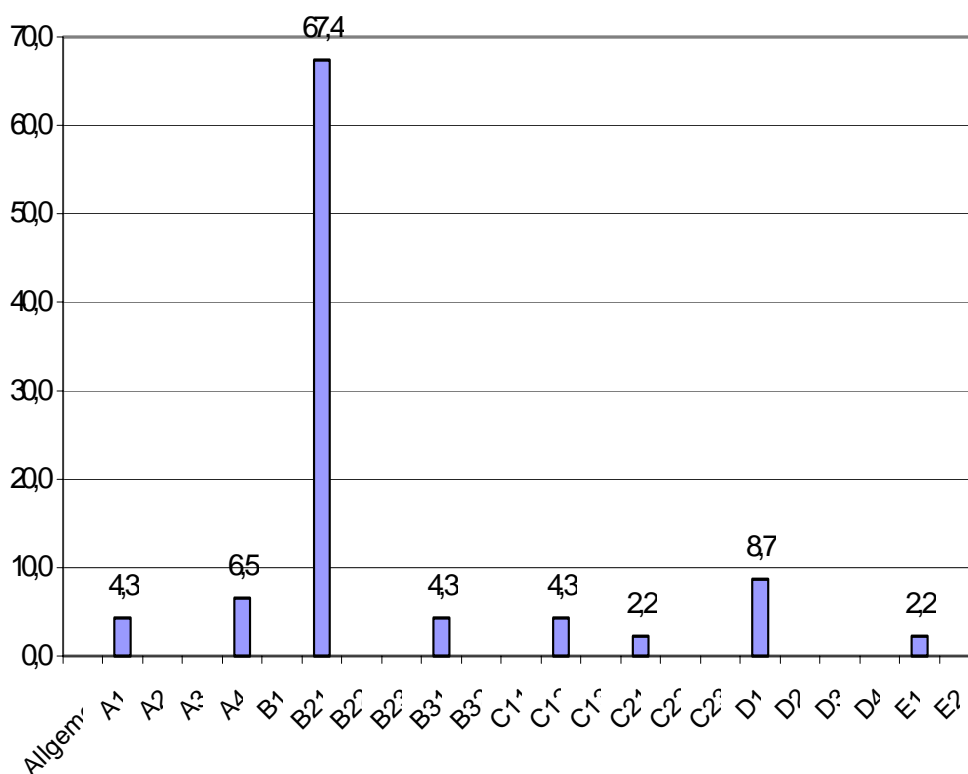


Abb. 8: Darstellung des prozentualen Anteils regulärer Datenbankobjekte in LADIB.

In Abbildung 6 werden die nicht erläuterten Datenobjekte in LADIB nach Teilprojekten gegliedert. Das bedeutet eine weitere Qualifizierung der Datenbestände in LADIB.

Es zeigt sich, dass in der Teildatenbank E2 zwar viele Dateien abgelegt wurden, diese aber nur wenige erläuterte Objekte bilden. In E2 ist ein Datenbankobjekt beispielsweise der Projektantrag, bestehend aus vielen einzelnen Teilprojektanträgen, welche eine große Dateizahl bedingen. Aus der Qualitätsstufe 2 fallen diese jedoch heraus. Die Anteile der Teilprojekte an LADIB nähern sich auf dieser Qualitätsstufe einander.

Beim Anteil erläuterter Datenbankobjekte gewinnt Teilprojekt B2.1 (Abb. 7) eine herausragende Stellung. Dessen Ergebnisse sind von grundlegender Bedeutung für andere Teilprojekte, die auf ihnen aufbauen. Für die entsprechende Arbeitsgruppe erwächst daraus ein hoher Erwartungsdruck auf qualitativ hochwertige und valide Daten.

Diese Wichtigkeit von B2.1 dokumentiert auch Abbildung 8. In ihr sind die regulären Datenbankobjekte von LADIB dargestellt. Diese Datenbankobjekte entsprechen dem Qualitätsniveau, das letztlich alle Daten in LADIB besitzen sollen. Mehr als Zweidrittel des regulären Inhalts von LADIB ist demnach der Teildatenbank der Arbeitsgruppe B2.1 zuzuordnen. Die Teilprojekte A2 sowie E2, die sehr viel Speicherplatz von LADIB belegten, haben keine regulären Beiträge zu LADIB beigesteuert. Ebenso verhält es sich mit C1.1.

6 Zeitliche Dynamik des Datenflusses in LADIB

Das DBS übernahm ab November 1997 alle Dateien mit SFB-Relevanz, die bis dahin im geordneten Dateiverzeichnis des SFB Servers standen. Neben einer kurzfristigen, saisonalen Dynamik, nach der Daten vor allem am Ende des Winterhalbjahres in LADIB eingebracht wurden, ist eine über vier Jahre dauernde Dynamik erkennbar (Abb. 9).

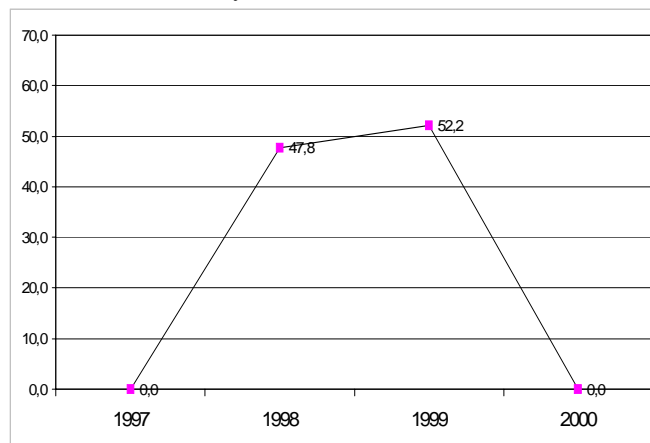


Abb. 9: Die Dynamik der Speicherung regulärer Datenbankobjekte in LADIB (1/1997 – 9/2000).

Zu Beginn der ersten dreijährigen SFB-Projektphase werden keine Datenobjekte in LADIB abgelegt, die aus dem gerade beginnenden Forschungszyklus stammen. Dagegen findet in den beiden Folgejahren eine beständige Datenanlieferung statt. Das erste Jahr einer Projektphase ist dasjenige Jahr, in dem noch keine fertig ausgewerteten Versuchsergebnisse aus dem Projekt selbst vorliegen. Nach Ende der Vegetationsperiode, die im allgemeinen auch gleichzeitig die Versuchsperiode ist, besteht eine Zeitverzögerung von einem halben Jahr für den Datenfluss, so dass zu Beginn des zweiten Quartals 1998 erste reguläre DBO nach LADIB gelangen. Gemäss dieser Annahme hätten aber zu Beginn des Jahres 2000 wieder Daten nach LADIB kommen sollen, die aus der Vegetationsperiode 1999 stammen. Das geschieht nicht. Vermutlich wird dieser Effekt durch den Doktorandenwechsel nach der ersten Projektphase hervorgerufen. Aus den neuen Versuchen liegen noch keine Ergebnisse vor. Die alten Daten aber werden erst im Zusammenhang mit der abgeschlossenen Dissertation für LADIB verfügbar sein. Diese ist nach unseren Beobachtungen mit etwa einem Jahr Verzögerung zum Versuchende - also nach etwa vier Jahren - fertiggestellt.

Ein weiteres interessantes Phänomen wird erkennbar, wenn man von der Dynamik regulärer Datenbankobjekte absieht und ausschliesslich den Datenstrom in Megabyte (MB) betrachtet (Abb. 10).

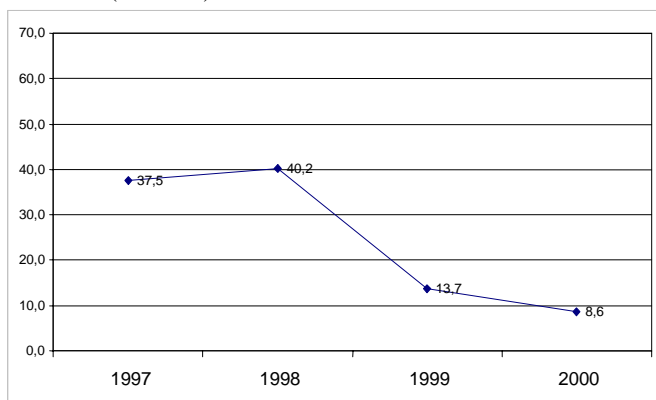


Abb. 10: Prozentualer Datenfluss über die Zeit in MegaByte (MB) nach LADIB, ohne weitere Differenzierung nach Datenqualität.

So reisst der Datenstrom über vier Jahre nie ab, verringert sich aber im Verlauf dieses Zeitraums stetig. Allerdings muss dabei berücksichtigt werden, dass gerade zu Beginn des SFB speicherintensive Geodaten in LADIB eingestellt wurden (1997). Im Jahre 1997 beispielsweise genügen die Geodaten aus A2 noch nicht den Anforderungen eines standardisierten Metadatenobjektes, weshalb sie in Abbildung 9 nicht berücksichtigt werden. Im darauffolgenden Jahr gelangen erste Datenobjekte der Qualitätsstufe reguläres Datenbankobjekt nach LADIB, da sie die Grundlage der Arbeit für weiterer Teilprojekte waren. So steigt 1998 sowohl das Datenvolumen als auch die Datenqualität an. Zieht man dagegen die Zahl der gespeicherten Dateien heran, so ist eine stete Zunahme an Lieferungen aus den Teilprojekten zu verzeichnen (Abb. 11). Der Grund dafür findet sich in der aussergewöhnlichen Aktivität des Teilprojektes E2, welches die SFB-Berichte und SFB-Anträge des Jahres 1999 sowie

Karten des Landesvermessungsamtes Hessen in LADIB ablegt.

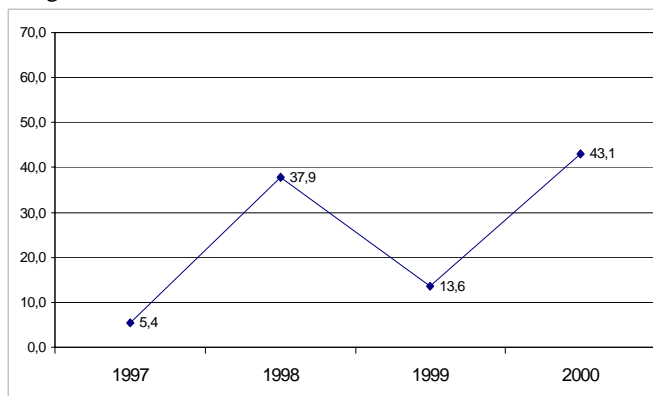


Abb. 11: Prozentualer Datenfluss ausgedrückt in der nach LADIB übertragenen Dateizahl ohne weitere Differenzierung nach Datenqualität.

Unterzieht man die Datenflussdynamik der fünf aktivsten Teilprojekte einer eingehenden Betrachtung so fallen weitere Besonderheiten auf (Abb. 12).

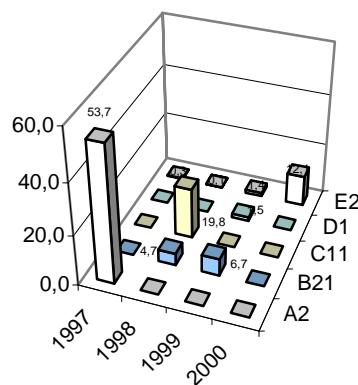


Abb. 12: Zeitliche Dynamik des Datenflusses ausgewählter Teilprojekte in Prozent.

Die Teilprojekte A2 und C1.1 dominieren zunächst die zeitlichen Dynamik des Datenflusses nach LADIB. Der Beitrag von A2 rührt aus dem schon genannten Vorläuferprojekt des SFB 299. Das Teilprojekt C1.1 hingegen sandte einmalig ein großes Datenvolumen. Da es im zweiten Projektjahr eintraf, kann es sich noch nicht um die Daten aus der abgeschlossenen Dissertation handeln, wie im Falle des Teilprojektes D1. Vielmehr waren es Zwischenergebnisse der Qualitätsstufe eins. Das Teilprojekt B2.1 weist dagegen einen kontinuierlichen Datenfluss auf. Pulsierende Datenströme wie aus A2, C1.1 oder D1 sind in LADIB schlecht planbar. Sie deuten nach unserer Erfahrung auf nicht reguläre Datenobjekte hin.

7 Schlussfolgerung und Ausblick

Aus der beobachteten Zusammensetzung und Entwicklung der Datenbestände in LADIB, lassen sich Gesetzmäßigkeiten der Datendynamik sowie bei den Projektdaten ableiten.

Datendynamik: Aufgrund der Datendynamik in LADIB ist es sinnvoll, freie Speicherkapazitäten vorzuhalten, da der

Bedarf an Speicher nicht gleichmäßig ansteigt. Die beobachtete Datendynamik ist für im Freiland arbeitenden Gruppen typisch und jahreszeitlich bedingt. Nach den Geländeerhebungen und der Datenerfassung nimmt der Datenstrom im Winterhalbjahr zu (z. B. B3.1, B2.1). Teilprojekte, die unabhängig von Jahreszeiten arbeiten, neigen dazu Daten mit dem Abschluss einer Dissertation zu senden (z.B. D1). Unabhängig von diesen Zyklen sind Teilprojekte, die externe Daten verfügbar machen (z. B. A2), übergeordnete Daten produzieren (z. B. E2) oder aufgrund spezifischer Anforderungen aus dem SFB besonders nachgefragt sind (z. B. B2.1). Die Entwicklung des Datenbestandes in LADIB spiegelt demnach die Arbeitspraxis der Teilprojekte wider.

Projektdateien: Zwar besteht Konsens zwischen allen Teilprojekten über die Notwendigkeit eines Datenmanagements, doch erzeugen einzelne Projekte größere Datenmengen als andere. Dabei ist die Qualität der Daten unterschiedlich. Vermutlich konnten manche Arbeitsgruppen aus Zeitgründen ihre Daten nicht transferieren. Dies unterstreicht die Vorteile eines zentralen DBS, weil es Teilprojekte von Datenmanagementaufgaben entlastet. Zum anderen ist davon auszugehen, dass der Datentransfer in bestimmten Arbeitsgruppen möglicherweise aufgrund interner Vorbehalte nur eine niedrige Priorität besitzt. Solche Zurückhaltung lässt sich nur schwer aufheben, und ist ein Problem des Wissensmanagements: wird Wissen freigegeben, so ist es im SFB allgemein verfügbar, kritisierbar und bleibt doch wenig entlohnt. Eine Furcht besteht möglicherweise auch darin, dass andere Arbeitsgruppen besseren Gebrauch von den eigenen, mühsam erworbenen Daten machen könnten. Initiativen, welche der Veröffentlichung von Daten einen den Fachpublikationen vergleichbaren Stellenwert einräumen, könnten hier einen beschleunigten Datenschub auslösen. Generell muss dieser Bereich der öffentlichen Wissenserzeugung modernisiert werden, um auf diese Weise Kooperationen zwischen Forschung und Anwendung zu beschleunigen. Letztlich gewinne alle Beteiligten aus einer solchen Public Private Partnership.

Datenqualität: Wie die Erfahrungen des Datenmanagements zeigen, sind der Umfang der gelieferten Daten und deren Qualität nicht immer kongruent zueinander. Mit anderen Worten: Man unterliegt einer Täuschung, wenn man die Leistungsfähigkeit einer Gruppe am Datenstrom in MB misst. Um aber den Datenstrom nicht zu hemmen oder durch hohe Qualitätsnormen abzuwürgen, wurde ein vierstufiges Qualitätsmodell eingeführt. Das für LADIB verwendete Qualitätsmodell erlaubt den Einstieg in die Datenproduktion und den Datentransfer bei geringen Hürden. Zugleich kann das Datenmanagement auf Zielvorgaben verweisen und so die Nutzer von LADIB anregen, ihr Datenqualitätsniveau zu steigern. Langfristig wertvoll sind die Daten auf der Stufe vier.

Sehr wichtig ist die frühzeitige Formulierung der Metadatenpezifikation. So wurde den SFB Angehörigen unmittelbar nach dessen Einrichtung vermittelt, dass bezüglich ihrer Datenproduktionszyklen Mindestanforderungen bestehen. Offenbar ist es von Vorteil, wenn das Datenmanagement früh präsent ist und dabei flexibel bleibt. Im Gegensatz dazu braucht das Management

bestehender Datenbanken keine dieser beiden Vorgaben zu erfüllen.

Vorteile/Nachteile von LADIB: Neben den schon genannten Nutzen einer zentralen Datenhaltung sind weitere Vorteile aus dieser Form der Datenhaltung zu nennen. So sieht sich der Datenproduzent in die Pflicht genommen, seine Ergebnisse bzw. Daten frühzeitig zu evaluieren und die Produktion redundanter Daten zu vermeiden. Daraus folgt eine Daten-konsistenz, die zweifelhafte, unnötige oder falsche Information von der Projektgemeinschaft fernhält. Aktive Datengenerierung wird insoweit belohnt, als das Ansehen einer Gruppe mit der Qualität und Validität ihrer Produkte in LADIB steigt. Weitere Möglichkeiten zur Anerkennung produktiver Gruppen sollten entstehen.

Neben den vielen Vorteilen zentraler Daten- und Informationshaltung finden sich aber auch Nachteile. Einmal publizierte Daten lassen sich zwar wieder aus LADIB zurücknehmen, ziehen aber einen Prestigeverlust nach sich. Daher neigen Forschergruppen im SFB dazu, ihre Daten für LADIB erst dann freizugeben, wenn sie sich der Qualität ihrer Datenprodukte gänzlich sicher sind. Diesem Problem könnte das Datenmanagement entgegenzutreten, indem ein inoffizieller Datenbereich errichtet wird. Doch bliebe dann immer noch die Schwierigkeit bestehen, dass die Entfernung von Daten aus dem inoffiziellen Datenpool, die bereits in den Arbeitsprozess einer anderen Forschergruppe eingebunden waren, zu erheblichen Problemen führen könnte.

Ein weiterer Nachteil besteht darin, dass für die einzelnen Arbeitsgruppen die Möglichkeit entfällt zu kontrollieren, wer ihre Daten nutzt bzw. anfordert. Verbunden damit ist eine Einschränkung des eigenbestimmten Datenaustauschs zwischen den Teilprojekten.

Hinsichtlich der Datensicherheit, insbesondere Datenverlust und Diebstahl, besitzt der zentrale Datenserver allerdings Vorteile aufgrund eines regelmäßigen Backups bzw. der Datenspiegelung. Der Verlust oder gar Diebstahl von Daten ist bei einem professionellen Datenmanagement deutlich geringer, als bei einer verteilten Datenhaltung. Diese Vorteile eines zentralen Datenmanagements treten aber bei bestimmten Aufgabenstellungen in den Hintergrund, so beispielsweise bei einem dezentralen Clearinghouse (Lochter und Knauer, 2000, Wächter et al., 2000).

Mit dem vorliegenden Artikel wurde der Versuch unternommen, die Aufmerksamkeit auf den Datenfluss in einem Datenmanagementprojekt zu lenken. Naturgemäß tritt dabei der softwaretechnische hinter den beschreibend-statistischen Blickwinkel des Datenbanksystems zurück. Dieser statistische Aspekt eines Datenprojekts darf auch in einem Forschungsprojekt des Umweltbereiches nicht unterschätzt werden. Er schafft die Grundlage für anwenderorientiertes Informationsmanagement, wie es auf erfolgreichen Informationsseiten des Internets praktiziert wird.

8 Literatur

- BRAUN, P. (2000): Zur Dynamik von Metadaten. In: Umweltdatenbanken im Web. (Hrsg.: Kramer, R. und Hosenfeld, F.), 52-61. - Umweltbundesamt Wien GmbH, Wien, Österreich.
- BRAUN, P., Becker, H. & Köhler, W. (2000): Die Entwicklung des Netzsystems ARTUS als Voraussetzung für

ein Umweltinformationssystem. - Zeitschrift für Agrarinformatik 8, 10-15.

HEUER, A. und Saake, G. (1995): Datenbanken: Konzepte und Sprachen. – International Thompson Publishing, Bonn.

LOCHTER, F. & Knauer, S. (2000): ICDP Clearinghouse. - <http://dc.gfz-potsdam.de/dc/>

WÄCHTER, J., Merz, B. & Lochter, F. (2000): Das Deutsche Forschungsnetz Naturkatastrophen. – GIS 5/00 (5), 5-6.

Das Datenbanksystem LADIB (P. Braun, M. Szibalski und W. Köhler)

Dr. Martin Szibalski ist tätig am Statistischen Bundesamt in Wiesbaden (Fon: 0611/753676, Gruppe IT-Nutzerservice, Gustav-Stresemann-Ring 11, 65189 Wiesbaden, Deutschland).

Prof. Dr. Wolfgang Köhler arbeitet an der Justus-Liebig-Universität in Giessen (Fon. 0641/9937540, Abteilung Biometrie und Populationsgenetik, Heinrich-Buff-Ring 26-32, 35392 Giessen, Deutschland).

Zusammenfassung

Für die Unterstützung des Transfers und die langfristige Nutzbarkeit von Ergebnissen der Umweltforschung wurde im Sonderforschungsbereich (SFB) 299 das zentrale Datenbanksystem LADIB aufgebaut. Die fachübergreifende und dauerhafte Verständlichkeit der Ergebnisse wurde durch eine frühzeitige Festlegung der Metadaten sowie ein Qualitätsmodell für Datenobjekte sichergestellt.

Seitens der Teilprojekte im SFB wurden die Chancen des zentralen DBS in unterschiedlichem Maße genutzt. Grundlegende Teilprojekte erzeugten frühzeitig qualitativ hochwertige Daten, die von anderen Teilprojekten als Eingang für eigene Untersuchungen genutzt wurden. Dagegen produzierten einige Teilprojekte keine oder qualitativ minderwertige Daten. Eine Datendynamik ergab sich aus den fachspezifischen Methoden unterschiedlicher Teilprojekte. Insgesamt erwies sich die frühe Aktivität des Datenmanagements als vorteilhaft für die Etablierung eines kooperativen Wissensmanagements.

Stichworte: Datenbanksystem, Umweltforschung, Metadaten, Qualitätsmodell, Datendynamik

The Data Base System LADIB (P. Braun, M. Szibalski und W. Köhler)

Summary

To support the information transfer and the long term usability of environmental research results, the central Data Base System LADIB has been built up as a part of the Collaborative Research Center (CRC) "Landuse concepts for peripheral regions". For a transdisciplinary and lasting understandability of the results Metadata and quality model for data objects were defined.

The possibilities of the central DBS were used to a different extent by the CRC subprojects. Basic subprojects generated data of high quality, which other subprojects used as input for their own works. However some subprojects produced no or qualitatively bad data for LADIB. A data dynamic resulted from subject specific methods. The early introduction of LADIB proofed to be a valuable step towards cooperative knowledge management.

Key words: data base system, environmental research, metadata, quality model, data dynamic

Dr. Paul Braun arbeitet am GeoForschungsZentrum in Potsdam (Fon: 0331/2881694, Daten- und Rechenzentrum, Telegrafenberg A3, 14473 Potsdam, Deutschland).

