



Hochschule Neubrandenburg
University of Applied Sciences

Hochschule Neubrandenburg
Studiengang Geoinformatik

**Entwicklung eines Speicherkonzeptes für Geodaten
in Unternehmen**

Bachelorarbeit

vorgelegt von: Ulrike Blank

Zum Erlangen des akademischen Grades
„Bachelor of Engineering“ (B.Eng.)

Erstprüfer: Prof. Dr.-Ing. Andreas Wehrenpfennig

Zweitprüfer: Dipl.-Ing. Matthias Nittel

Bearbeitungszeitraum: 28.07.08 - 22.09.08

URN: urn:nbn:de:gbv:519-thesis2008-0261-9

Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Bachelorarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Neubrandenburg, den 22.09.2008

Danksagung

An dieser Stelle möchte ich mich bei meinen Betreuern, Herrn Prof. Dr.-Ing. Andreas Wehrenpfennig und Herrn Dipl.-Ing. Matthias Nittel, bedanken. Sie unterstützten mich in allen Phasen meiner Bachelorarbeit und standen mir jederzeit mit Rat und Tat zur Seite.

Kurzfassung

Im Rahmen der Bachelorarbeit wird ein Konzept zur Speicherung der anfallenden Geodaten für das Unternehmen GeoSense erarbeitet. Das Unternehmen beschäftigt sich mit Bildflug- und GIS-Projekten. Die dabei erzeugten Geodaten sollen in geeigneter Form gespeichert werden. Je nach Bearbeitungsstand und der Anzahl der am Projekt tätigen Mitarbeiter müssen die Daten an verschiedenen Arbeitsplätzen mit unterschiedlichen Programmen vorgehalten werden. Die Daten sollen während der Bearbeitungszeit von mehreren Bearbeitungsprogrammen abrufbar und nutzbar sein. Des Weiteren sollen die Daten so abgelegt werden, dass sie für ein Geoportal nutzbar sind. Für die Entwicklung des Speicherkonzeptes werden zunächst der Arbeitsablauf und die bei der derzeitigen Datenhaltung auftretenden Probleme anhand eines Beispiels analysiert, um daraus die Anforderungen an das Speichersystem zu definieren. Anschließend werden verschiedene Speichermöglichkeiten vorgestellt. Auf Basis dieser Untersuchungen wird ein Speicherkonzept erstellt. Im Anschluss werden verschiedene Produkte für eine Umsetzung des Konzeptes miteinander verglichen. Nach der Entwicklung des Speichersystems werden noch verschiedene Backup-Strategien und Backup-Medien betrachtet.

Abstract

The purpose of this bachelor thesis is to design a concept for the storage of accumulating geodata for the GeoSense company. GeoSense deals with GIS-projects. The thereby generated geodata shall be stored in a suitable form. The data has to be accessible on various workstations with different programmes according to the state of process and the number of on the project working employees. During the process time the data should be recallable and usable for several application programmes. Furthermore the stored data should be usable for a geoportal. At first the working process and problems will be analyzed with the help of an example for the design of the storage concept. From this analysis the requirements are defined for the storage system. Afterwards different storage possibilities are introduced to develop a suitable storage concept from it. Then various products for the implementation of the storage concept are compared. In addition to the design of the storage concept different backup strategies and backup media are considered.

Inhaltsverzeichnis

1 Einführung	1
2 Ausgangssituation im Unternehmen - Analyse	2
2.1 Technische Ausstattung	2
2.2 Arbeitsabläufe	3
2.3 Datenformate und Speicherbedarf	4
2.4 Probleme	7
3 Entwicklung des Speicherkonzepts	8
3.1 Anforderungen an das System	8
3.2 Speichermöglichkeiten	8
3.2.1 Dateisystem und Datenbanksystem	8
3.2.2 Versionskontrollsystem	11
3.3 Metadaten	15
3.4 Nutzergruppen	17
3.5 Speichermodell	18
4 Umsetzung	23
4.1 Software	23
4.1.1 Server	23
4.1.2 Datenbank	25
4.1.3 Versionskontrollsystem	26
4.2 Nutzung von Subversion	28
5 Backup	30
5.1 Backup-Strategien	30
5.2 Backup-Medien	33
6 Zusammenfassung und Ausblick	34
Literaturverzeichnis	36
Abbildungsverzeichnis	38
Tabellenverzeichnis	39

Abkürzungsverzeichnis

ArcGIS	GIS-Software der Firma ESRI
ASCII	American Standard Code for Information Interchange
DBMS	Datenbankmanagementsystem
DCMI	Dublin Core Metadata Initiative
DXF	Drawing Exchange Format
EPSG	European Petroleum Survey Group
ESRI	Environmental Systems Research Institute
GeoTIFF	Spezielle Form von TIFF
GIS	Geoinformationssystem
GUI	Graphical User Interface
FTP	File Transfer Protocol
ISO	Internationale Organisation für Normung
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
PNG	Portable Network Graphics
Polygis	GIS-Software der Firma ICA
RAID	Redundant Array of Independent Disks
SHP	Shapefile
SVN	Subversion
TFW	World File for TIFF Image
TIFF	Tag Image File Format
WMS	Web Map Service

1 Einführung

Die Grundlage jeder GIS-Anwendung bilden Daten. Diese Daten werden als Geodaten bezeichnet. Sie weisen einen festen Raumbezug auf, über den ein Lagebezug zur Erdoberfläche hergestellt werden kann. Geodaten beschreiben Objekte der Realität durch geometrische und inhaltliche Attribute. Typische Beispiele für Geodaten sind Landkarten mit administrativen Einheiten und anderen als punkt-, linien- oder flächenförmig vorliegenden Objekten (z.B. Städte, Flüsse, Naturschutzgebiete), Satelliten- und Luftbildaufnahmen oder Messungen z.B. geologische Bodenproben.

„Studien beziffern den Anteil von Geodaten auf 80 - 85% bezogen auf alle digital vorliegenden/erhobenen Daten.“¹ Vielerorts stehen Unternehmen und Organisationen vor der Aufgabe, ihre wachsenden Geodaten-Bestände auf kostengünstige Weise intern und auch öffentlich bereitzustellen und anzubieten. So auch das Bachelorarbeit begleitende Unternehmen GeoSense. Nicht nur die Erfassung, Verarbeitung und Darstellung der Geodaten, sondern auch die optimale Speicherung und Verwaltung der bei der Verarbeitung entstehenden Daten sind für eine effiziente Arbeitsweise des Unternehmens wichtig.

Auf Grund der Bedeutung der Speicherung soll im Folgenden ein Speicherkonzept für das Unternehmen GeoSense erarbeitet werden.

Um die anfallenden Geodaten und den dafür benötigten Speicherbedarf im Unternehmen GeoSense zu ermitteln, wird zunächst der Arbeitsablauf analysiert. Anschließend werden die Probleme des derzeitigen Speicherkonzepts aufgezeigt und die Anforderungen an das neue Speichersystem definiert. Im nächsten Schritt werden verschiedene Speichermöglichkeiten vorgestellt. Auf Basis dieser Untersuchungen wird ein Speicherkonzept erstellt und auf die daraus resultierenden Änderungen für das Unternehmen eingegangen. Als Nächstes werden verschiedene Produkte für die exemplarische Umsetzung miteinander verglichen. Im Anschluss daran wird auf das Arbeiten mit Subversion näher eingegangen. Abschließend werden noch unterschiedliche Backup-Strategien und Backup-Medien vorgestellt, da nicht nur die Speicherung von Daten wichtig ist, sondern auch deren Sicherung. Am Ende wird die Arbeit kurz zusammengefasst und ein Ausblick hinsichtlich möglicher Erweiterungen gegeben.

¹ Quelle: [Göbel 02] (Seite 7)

2 Ausgangssituation im Unternehmen - Analyse

Das Unternehmen GeoSense ist im Dienstleistungsbereich angesiedelt. GeoSense ist ein Ingenieurbüro für Datenverarbeitung und Geoinformation und hat sich dabei auf die Bereiche

- Bildflug (Planung, Durchführung und Auswertung)
- Luftbildauswertung
- Geoinformationssysteme (Einrichtung, Betreuung sowie Pflege, Bearbeitung und Anpassung von Datenbanken)
- Thermographie

spezialisiert. Weitere Tätigkeitsfelder des Unternehmens sind die Herstellung von Karten, Passpunktmessungen sowie die Triangulation von Luftbildern.

Für alle oben genannten Bereiche ist der Einsatz von Geodaten unabdingbar. Es werden Geodaten abgerufen, bearbeitet und abgelegt. Für eine effektive Arbeitsweise des Büros ist eine bestmögliche Speicherung der bei der Bearbeitung entstehenden Daten wichtig. Um ein optimales Speicherkonzept zu entwickeln, werden zunächst die technische Ausstattung, der Arbeitsablauf sowie verwendete Datenformate und der benötigte Speicherbedarf analysiert.

2.1 Technische Ausstattung

Für die Bearbeitung der Projekte stehen an jedem Arbeitsplatz verschiedene Bearbeitungsprogramme zur Verfügung, unter anderem ArcGIS [ArcGIS 05] von der Firma ESRI und Polygis [Polygis] von der Firma IAC. Die Computer des Unternehmens sind in ein lokales Netzwerk eingebunden. Über dieses Netzwerk wird der Datenaustausch zwischen den Computern sowie der Zugriff auf einen Server gewährleistet. Zurzeit werden alle Projektdaten auf einem Server gespeichert, der als Fileserver dient. Die Projektdaten des Unternehmens werden auf 3 Festplatten (à 500 GB) gesichert, die zu einem RAID 5 (Redundant Array of Independent Disks) zusammengefasst sind. Der Zugriff auf die Daten erfolgt über das File Transfer Protokoll (FTP). Als Software für den FTP-Server ist der FileZilla-Server [FileZilla] im Einsatz. Über FileZilla werden auch die Rechte von Gruppen und den einzelnen Nutzern der Firma geregelt.

2.2 Arbeitsabläufe

Stellvertretend für die durchzuführenden Arbeiten im Ingenieurbüro GeoSense wird der Arbeitsablauf für die Orthofotoberechnung näher beschrieben.

Das Unternehmen erhält die auszuwertenden Luftbilder auf DVD oder Festplatte. Die Luftbilder werden für die Verarbeitung auf den Server übertragen. Voraussetzung für eine photogrammetrische Auswertung von Luftbildern ist die Aerotriangulation mit Bündelblockausgleichung. Dabei werden die einzelnen Luftbilder zu einem festen Bildverbund verknüpft und mit Hilfe von Passpunkten innerhalb eines Koordinatensystems orientiert. Im nächsten Schritt wird je nach Auftrag für die Orthofotoberechnung ein digitales Geländemodell erzeugt oder gekauft. Anschließend werden die einzelnen Orthofotos berechnet. Orthofotos sind über ein Geländemodell entzerrte Luftbilder. Die Verzerrungen, die bei der fotografischen Aufnahme eines Bildes durch die unterschiedlichen Abstände des Geländes zur Kamera entstehen, werden rechnerisch unter Verwendung eines digitalen Geländemodells beseitigt. Durch die Korrektur erhält jedes Bild einen einheitlichen Maßstab und kann mit anderen Geodaten überlagert bzw. kombiniert werden. Aus den einzelnen digitalen Orthofotos wird durch das Erzeugen von Schnittlinien ein Mosaikbild des Gebietes hergestellt. Diesen Vorgang nennt man Mosaikieren. Die Gebiete beiderseits der Schnittlinien weisen oft starke Helligkeits- und Kontrastunterschiede auf. Dies kommt im Wesentlichen daher, dass die Blickrichtung und Belichtung und damit der Sonneneinfallswinkel in den jeweiligen Bildern unterschiedlich ist. Durch radiometrische Anpassung werden diese Helligkeitsunterschiede reduziert. Im letzten Schritt wird das Mosaikbild in gleich große Kacheln zerlegt. Je nach Auftrag werden nach der Luftbildauswertung noch entsprechende Kartierungen durchgeführt. Der Kunde erhält das Endprodukt entweder auf DVD oder Festplatte.

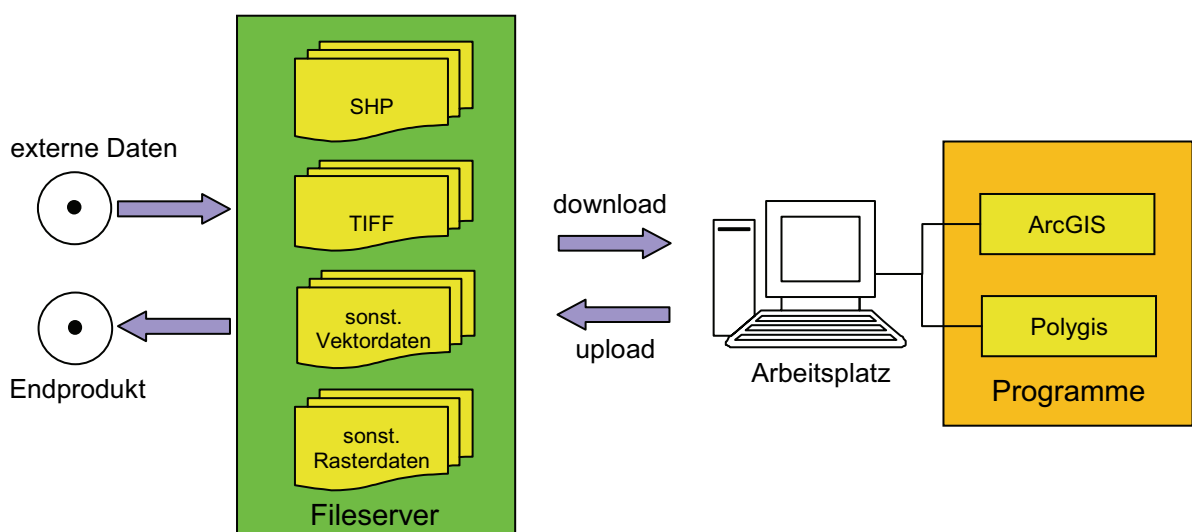


Abbildung 1: Datenfluss

In der Abbildung 1 ist der Datenfluss dargestellt. Externe Daten, wie zum Beispiel Luftbilder vom Kunden oder gekaufte Höhenmodelle, werden auf den Server übertragen. Für die jeweiligen Arbeitsschritte werden die benötigten Daten vom Server auf den Arbeitsplatz geladen. Nach Beendigung des Arbeitsschrittes werden die neuen bzw. aktualisierten Daten auf den Server hochgeladen. Um zu verhindern, dass mehrere Mitarbeiter, die an einem Arbeitsplatz arbeiten, die gleichen Daten mehrmals herunterladen bzw. in unterschiedlichen Ordnern hochladen, existiert auf jedem Arbeitsplatz die gleiche Verzeichnisstruktur wie auf dem Server. Die Verzeichnisstruktur sieht wie folgt aus:

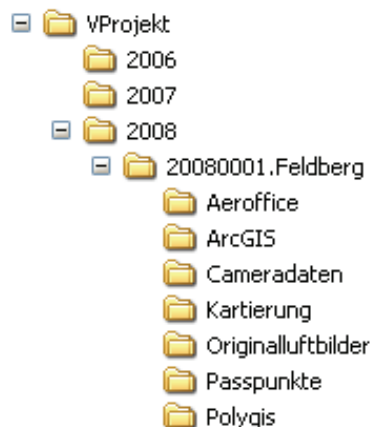


Abbildung 2: Verzeichnisstruktur

Nach Beendigung eines Projektes werden alle Dateien, die für zukünftige Projekte nicht mehr verwendet werden bzw. das Endprodukt für den Kunden auf DVD oder externe Festplatte ausgelagert.

2.3 Datenformate und Speicherbedarf

Bei der Bearbeitung von Projekten fallen unterschiedliche Raster- und Vektordaten [RaVe] an.

Rasterdaten beziehen sich nur direkt auf Flächen, wobei das geometrische Grundelement der Bildpunkt (Pixel) ist. Das Bild entsteht durch spalten- und zeilenweise angeordnete Pixel mit entsprechenden farblichen Flächenfüllungen. Die Größe eines solchen Pixels hängt von der Auflösung der Ausgabegeräte bzw. von den Erfassungsmethoden ab. Rasterzellen enthalten Informationen als Zahlenwert über Farbwerte, Höhenwerte oder verschiedenartige Eigenschaften der Erdoberfläche. Herkömmliche Rasterdaten weisen meist keinen räumlichen Bezug auf. Erst durch die Georeferenzierung werden dem Rasterbild Koordinaten eines geodätischen Bezugssystems zugewiesen. Bei der Georeferenzierung wird der Mittelpunkt des oberen linken Pixels der zu Grunde liegenden

Matrix über eine geographische Koordinate definiert. Die Informationen zum räumlichen Bezug werden entweder im Header des Rasterformates (z.B. GeoTIFF) oder in einer zusätzlichen ASCII-Datei (*.tfw beim Rasterformat TIFF) gespeichert. Typische Beispiele für Rasterdaten sind Luftbildaufnahmen und Satellitenbilder. Gebräuchliche Formate sind ESRI Grid, GeoTIFF, TIFF, JPEG und PNG.

Vektordaten erlauben gegenüber den Rasterdaten eine präzisere Darstellung von Geodaten. Für die Darstellung von Geo-Objekten werden geometrische Primitive (Punkte, Linien, Flächen) verwendet. Jeder Punkt besteht aus einem Koordinatenpaar. Linien setzen sich aus zwei Punkten zusammen und Polygone bestehen aus einer Aneinanderreihung von Linien, welche eine geschlossene Fläche ergeben. Die einfachen Grundstrukturen ermöglichen die Herstellung einer logischen Datenstruktur und eines Objektbezugs. Bevorzugte Datenformate sind Coverages und Shapefiles.

Der Speicherbedarf für ein Projekt wird am Beispiel der Orthofotoberechnung (siehe Kapitel 2.2) dargestellt. Von folgenden Parametern wird ausgegangen:

- 500 Bilder
- Maßstab: 1:7000
- Auflösung: 10 cm/Pixel
- Gebiet: 1,5 km x 1,5 km pro Bild.

In der Tabelle 1 sind die anfallenden Bilddateien mit ihrem jeweiligen Speicherbedarf aufgelistet. Bei den Bilddaten handelt es sich um Rasterdaten im TIFF Format. Bedingt durch ihre Datenstruktur benötigen sie den meisten Speicherplatz.

Bilddateien	Speicherbedarf pro Bild	Speicherbedarf insgesamt
Originalbilder	645 MB	315 GB
Orthofotos	645 MB	315 GB
2 km x 2 km Kacheln	1,2 GB	90 GB
radiometrisch angepasste Kacheln	1,2 GB	90 GB
500 m x 500 m Kacheln	75 MB	90 GB

Tabelle 1: zu verwaltende Bilddateien

Die Passpunkt- und Aerotriangulationsdaten liegen jeweils im ASCII-Format vor. Weitere anfallende Geodaten sind nachfolgend aufgelistet.

Geodaten	Datenformat	Datentyp
Höhenmodell	DXF, ASCII	Vektor, Raster
Schnittlinien	SHP, DXF	Vektor
Kartierungen	OP3D ² , SHP, DXF	Vektor

Tabelle 2: weitere zu verwaltende Geodaten

Die oben genannten Daten sind bis auf das Höhenmodell nur einige Kilobyte, maximal 2,5 Megabyte groß. Das Höhenmodell kann bis zu 350 MB groß werden.

Somit ergibt sich ein Speicherbedarf von ca. 900 GB pro Projekt. In der Regel werden immer ein bis zwei Projekte gleichzeitig bearbeitet. Je nach Auftragslage können es aber auch noch mehr Projekte sein. Somit ist die Speicherplatzkapazität vom Server (1,5 TB) fast vollständig verbraucht. Auch in den nächsten Jahren ist mit dem vorgenannten Speicherbedarf für die zu bearbeitenden Projekte zu rechnen.

Die Daten sollen während der Bearbeitungszeit von den Bearbeitungsprogrammen der Luftbildauswertung sowie Polygis und ArcGIS abrufbar und nutzbar sein. Des Weiteren soll über ein Geodatenportal auf die Daten zugegriffen werden können. Das Geoportal ist eine spezielle Form des Webportals. Über eine Internetseite wird der zentrale Zugang zur Information mit Raumbezug ermöglicht. Es werden sowohl Raster- als auch Vektordaten im Geoportal angezeigt und bearbeitet. Das Geoportal wurde mit Open Source Java-Framework deegree [deegree] realisiert. Deegree dient der Verwaltung und Darstellung von Geodaten. Für die Erzeugung der Karten wird der Web Map Service v2.1 von deegree verwendet und für die Darstellung der Karten wird iGeoPortal - Standard edition v2.1 von deegree benutzt. Die Daten können in Form eines Imagefiles, eines Shapefiles oder in einer Datenbank (z.B. PostgreSQL/PostGIS und Oracle Spatial) vorliegen.

Im Unternehmen liegen die Geodaten vorwiegend im SHP-, DXF- und TIFF-Format vor. Die nachfolgende Tabelle zeigt, welche Formate von welchen Programmen unterstützt werden. ArcGIS und Polygis unterstützen aber noch weitere Formate als die hier aufgelisteten. Zusätzlich sind noch die unterstützenden Datenbanken, Oracle und PostgreSQL, aufgelistet.

² OP3D: Ein programmspezifisches Speicherformat von tridicon 3D.

Programme	SHP	DXF	TIFF	Oracle	PostgreSQL
ArcGIS	x	x	x	x	x
Polygis	x	x	x	x	x
Geodatenportal	x	-	x	x	x

Tabelle 3: Datenformate

2.4 Probleme

Das Speichersystem, wie es derzeit existiert, hat einige positive und negative Aspekte. Positiv zu bemerken sind die feste Verzeichnisstruktur, wodurch die Suche nach bestimmten Dateien erleichtert wird und die Vergabe von Zugriffsrechten.

Probleme treten dann auf, wenn Benutzer eine Datei mit gleichem Namen oder eine Datei mit unterschiedlichem Bearbeitungsstand auf den Server laden wollen. Die vorhandene Datei kann dabei versehentlich überschrieben werden. Sollte diese Datei für spätere Zwecke noch verwendet werden, ist ein Wiederherstellen dieser Datei nur schwierig bzw. gar nicht möglich.

Wenn mehrere Mitarbeiter an verschiedenen Arbeitsplätzen an einem Projekt arbeiten, ist es wichtig, dass jedem der aktuellste Stand der Bearbeitung zur Verfügung steht. Dazu müssen die Nutzer wissen, dass Dateien verändert bzw. neue Dateien hinzugefügt wurden und welche Dateien davon konkret betroffen sind.

Im jetzigen System sind keine Metadaten zu den Geodaten vorhanden. Die Suche nach einer Datei kann sich als langwierig und schwierig gestalten, vor allem bei ausgelagerten Dateien, die sich auf der Festplatte bzw. DVD befinden. Somit kann es auch passieren, dass eine Datei mehrmals im System vorhanden ist und unnötig Speicherplatz verbraucht, da die Existenz dieser Datei unbekannt war.

Jedes Bearbeitungsprogramm speichert die erstellten Daten in unterschiedlichen Formaten ab. Die jeweiligen Formate werden aber nicht von allen Programmen unterstützt. Für die Luftbildauswertung wird zum Beispiel tridicon 3D der Firma GTA verwendet. Die kartierten Daten werden im binären OP3D-Format gespeichert. Dieses Format wird jedoch weder von ArcGIS noch von Polygis unterstützt. Ohne entsprechende Maßnahmen können die erstellten Dateien somit nicht in anderen Programmen verwendet werden.

3 Entwicklung des Speicherkonzepts

3.1 Anforderungen an das System

Aus der Analyse des derzeitigen Speichersystems und seiner Probleme ergeben sich die Anforderungen an das neue Speichersystem.

Alle Raster- und Vektordaten, die während der Bearbeitung von Bildflug- und GIS-Projekten entstehen, werden zentral auf einem Server gespeichert. Die Daten sollen während der gesamten Bearbeitungszeit von mehreren Arbeitsplätzen aus und von verschiedenen Bearbeitungsprogrammen abrufbar und nutzbar sein, insbesondere von den Programmen ArcGIS und Polygis. Zusätzlich soll das System den lesenden und schreibenden Zugriff auf die Daten mehreren Nutzern gleichzeitig gewährleisten. Des Weiteren sollen die Daten so abgelegt werden, dass sie nach der Bearbeitung für ein Geoportal genutzt werden können. Um unberechtigte Zugriffe auf die Geodaten auszuschließen, soll vor jedem Zugriff auf den Server eine Authentifizierung nötig sein.

Für eine Übersicht über den gesamten Datenbestand im Unternehmen sollen neben den Geodaten zusätzlich noch Metadaten gespeichert werden. Durch die Verwendung von Metadaten wird die Suche nach bestimmten Daten erleichtert.

3.2 Speichermöglichkeiten

3.2.1 Dateisystem und Datenbanksystem

Für die Speicherung der Geodaten gibt es zwei Möglichkeiten: ein Dateisystem oder ein Datenbanksystem.

Das Dateisystem entspricht der bisherigen Speicherung der Daten. Die Daten werden hierbei in Dateien gespeichert, die in Verzeichnissen abgelegt und verwaltet werden. Datenbanksysteme setzen sich aus Datenbankmanagementsystem (DBMS) und einer oder mehreren Datenbanken zusammen (siehe Abb. 3). Das DBMS dient der Verwaltung des Datenbestandes, hält ihn auf der logischen Ebene konsistent und macht ihn berechtigten Nutzern zugänglich. In der Datenbank werden die in Beziehung zueinander stehenden Daten gespeichert. Nutzer bzw. Programme greifen nicht direkt auf die Datenbank zu, sondern über eine definierte Schnittstelle des DBMS.

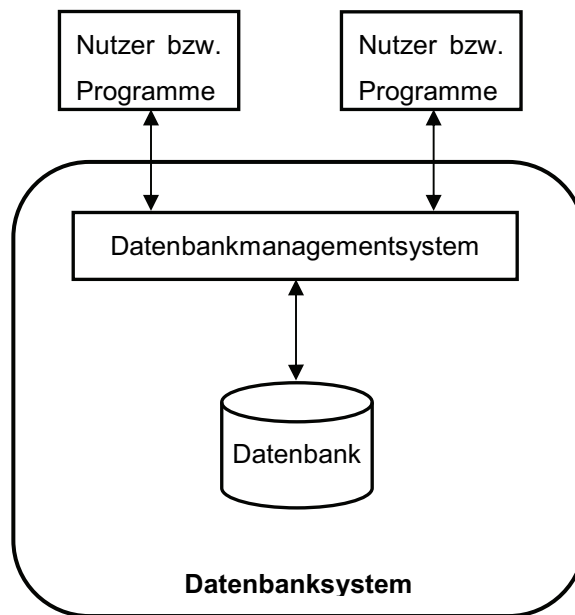


Abbildung 3: Datenbanksystem

Im Folgenden wird auf einige Vorteile von Datenbanksystemen gegenüber Dateisystemen eingegangen. [DB]

Inhaltliche Strukturierung großer Datenmengen

Daten mit ähnlichen Inhalten können relativ einfach vom Datenbanksystem gesucht werden.

Datenunabhängigkeit

Datenbanksysteme erlauben durch die Unterstützung der Drei-Ebenen-Schema-Architektur eine weitgehende Datenunabhängigkeit. Beim Dateisystem hingegen hat jede Änderung im Datenmodell oder in der physischen Speicherung der Daten Auswirkungen auf die Anwendungssoftware.

Anfragebearbeitung

Mittels einer Abfragesprache besteht die Möglichkeit beim Datenbanksystem, den Datenbestand vielfältig abzufragen. Bei Dateisystemen ist nur eine eingeschränkte Suche möglich. Über den Explorer ist zum Beispiel eine Suche nach Dateinamen durchführbar. Bei Textdateien kann auch der Inhalt durchsucht werden.

Persistenz

Das Datenbanksystem schützt den Nutzer besser vor Datenverlust, auch in extremen Situationen wie z.B. bei Stromausfall. Die Grundlage hierfür ist das sogenannte Transaktionskonzept. Das ist eine ununterbrochene Folge von Datenmanipulationsbefehlen, die die Datenbank von einem alten logischen konsistenten in einen neuen logischen konsistenten Zustand überführt. Kann ein Eintrag in einer Datenbank erfolgreich abgeschlossen werden, so hat sich ihr Zustand geändert. Gelingt der Eintrag nicht, so wird die Datenbank wieder in den alten konsistenten Zustand zurückversetzt. Trotzdem sollte auf ein regelmäßiges Backup (siehe Kapitel 5) auch bei Datenbanksystemen nicht verzichtet werden.

Integrität

Das Datenbanksystem prüft bei Änderungen der Datenbank (spätestens am Transaktionsende), ob die Datenbank noch konsistent ist, d.h. ob die Daten noch so sind, wie es die Integrationsregeln der Datenbank verlangen. Bei der Anwendung von Dateien sind solche Überprüfungen nur durch Programmierungen möglich.

Mehrbenutzerbetrieb und Datensicherheit

Mehrere Anwender greifen gleichzeitig auf die Datenbank zu, daher muss der Mehrbenutzerbetrieb sichergestellt werden. Dies erfolgt wiederum mit dem Transaktionskonzept. Dateisysteme können keinen zeitgleichen und konfliktfreien Mehrbenutzerbetrieb auf einer zentralen Basis gewährleisten. Dies trifft auch auf die Datensicherheit zu. Bei Datenbanksystemen wird die Datensicherheit durch Maßnahmen wie Zugriffskontrolle, Überwachung bzw. Protokollierung von Aktionen auf der Datenbank erreicht.

Zugriffskontrolle

Ähnlich wie Dateisysteme unterscheiden Datenbanksysteme zwischen verschiedenen Benutzern und Benutzergruppen. Es besteht die Möglichkeit, einzelnen Benutzern bzw. Benutzergruppen nach inhaltlichen Gesichtspunkten abgestufte Zugriffsrechte zu gewähren. Datenbanksysteme kontrollieren die Einhaltung dieser Rechte. Im Gegensatz zum Dateisystem können Zugriffsrechte nicht nur für physisch zusammen gespeicherte Daten vergeben werden, sondern auch für logisch zusammengehörige Daten.

Konkret würde der Einsatz eines Datenbanksystems für das Unternehmen Folgendes bedeuten:

- vereinfachte Suche nach Daten
- vielfältige Abfragemöglichkeiten des Datenbestandes
- einheitlicher Datenbestand für alle Nutzer
- Mehrbenutzerbetrieb
- mehr Datensicherheit
- besserer Schutz vor Datenverlust (Persistenz)
- großer Arbeitsaufwand bei der Überführung der Daten vom Dateien- ins Datenbanksystem auf Grund bereits großer vorhandener Datenmengen
- alle Datenbank unterstützenden Programme können auf alle Daten zugreifen.

3.2.2 Versionskontrollsystem

Das größte Problem bei der Bearbeitung der Projekte ist, dass mehrere Anwender gleichzeitig lesend und schreibend auf dieselben Daten zugreifen. Um dieses Problem zu lösen, bietet sich ein sogenanntes Versionskontrollsystem an. Es wird in der Softwareentwicklung, Backup- und Archivierungssystemen, Aktualisierungssystemen und Protokollsystem verwendet. Die Versionskontrolle ermöglicht das gemeinschaftliche Bearbeiten und Benutzen von Daten. Nachfolgend sind einige Eigenschaften eines Versionskontrollsystems aufgelistet [Wassermann 06]:

Versionierung

Dateien und Verzeichnisse werden versioniert. Somit können zu jeder Zeit alte Versionen der Dateien wiederhergestellt werden. Bei der manuellen Verwaltung ist dies nur durch Erstellung eines Backups für jede Version möglich, was auf Dauer zu einer unübersichtlichen Menge an Sicherungsdateien führt.

Übernahme von Neuerungen

Änderungen, die nachträglich in älteren Versionen durchgeführt werden, können automatisch in nachfolgende Versionen übernommen werden.

Parallelität

Mehrere Benutzer können gleichzeitig an denselben Daten arbeiten. Die jeweiligen Änderungen werden dann zusammengeführt. Sollte es bei der Zusammenführung zu Konflikten kommen, können diese relativ leicht behoben werden.

Zu den Hauptaufgaben eines Versionskontrollsystems gehören:

- Protokollierung aller Änderungen
- Archivierung verschiedener Versionen
 - Wiederherstellung alter Versionen
 - Verwaltung verschiedener Entwicklungslinien
- Koordination von konkurrierenden Zugriffsversuchen.

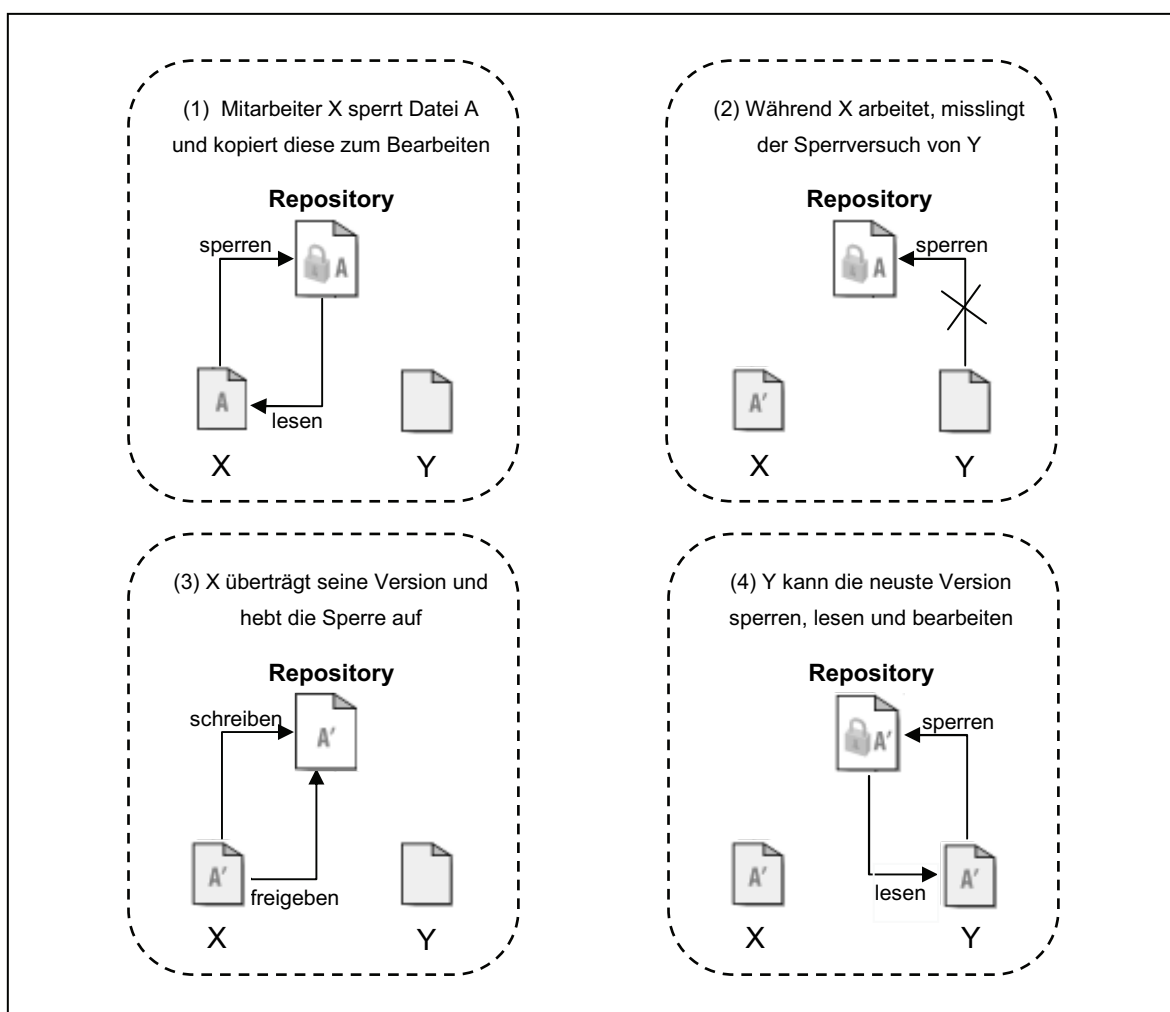
Ein Versionskontrollsystem besteht aus einem oder mehreren Repositories und einem Versionsmanagementsystem. Beim Repository handelt es sich um einen zentralen Datenspeicher, der eine Art Fileserver ist. Alle Versionen, Änderungen und Zugriffsdaten werden in einer Datenbank oder in speziellen Dateien gespeichert. Der Vorteil von Repositories gegenüber einem Fileserver liegt darin, dass sich ein Repository jede Änderung merkt, die jemals an den Dateien und Verzeichnissen vorgenommen wurde. Somit besteht die Möglichkeit, ältere Versionen der Daten wiederherzustellen oder zu untersuchen, in welcher Weise sich die Daten im Laufe der Zeit verändert haben. Das Versionsmanagementsystem dient der Verwaltung der Repositories.

Versionskontrollsysteme werden zwar überwiegend in der Softwareentwicklung zur Verwaltung von Quellcodes eingesetzt, sind aber für die Verwaltung aller Daten geeignet, für die eine Versionierung sinnvoll ist. Auch für die Projekte im Unternehmen ist die Speicherung von verschiedenen Arbeitsständen sinnvoll.

Um das Problem der gemeinsamen Datennutzung (Filesharing) zu lösen, gibt es zwei Modelle: Sperren-Ändern-Freigeben und Kopieren-Ändern-Zusammenführen. [Wasserman 06]

Sperren-Ändern-Freigeben

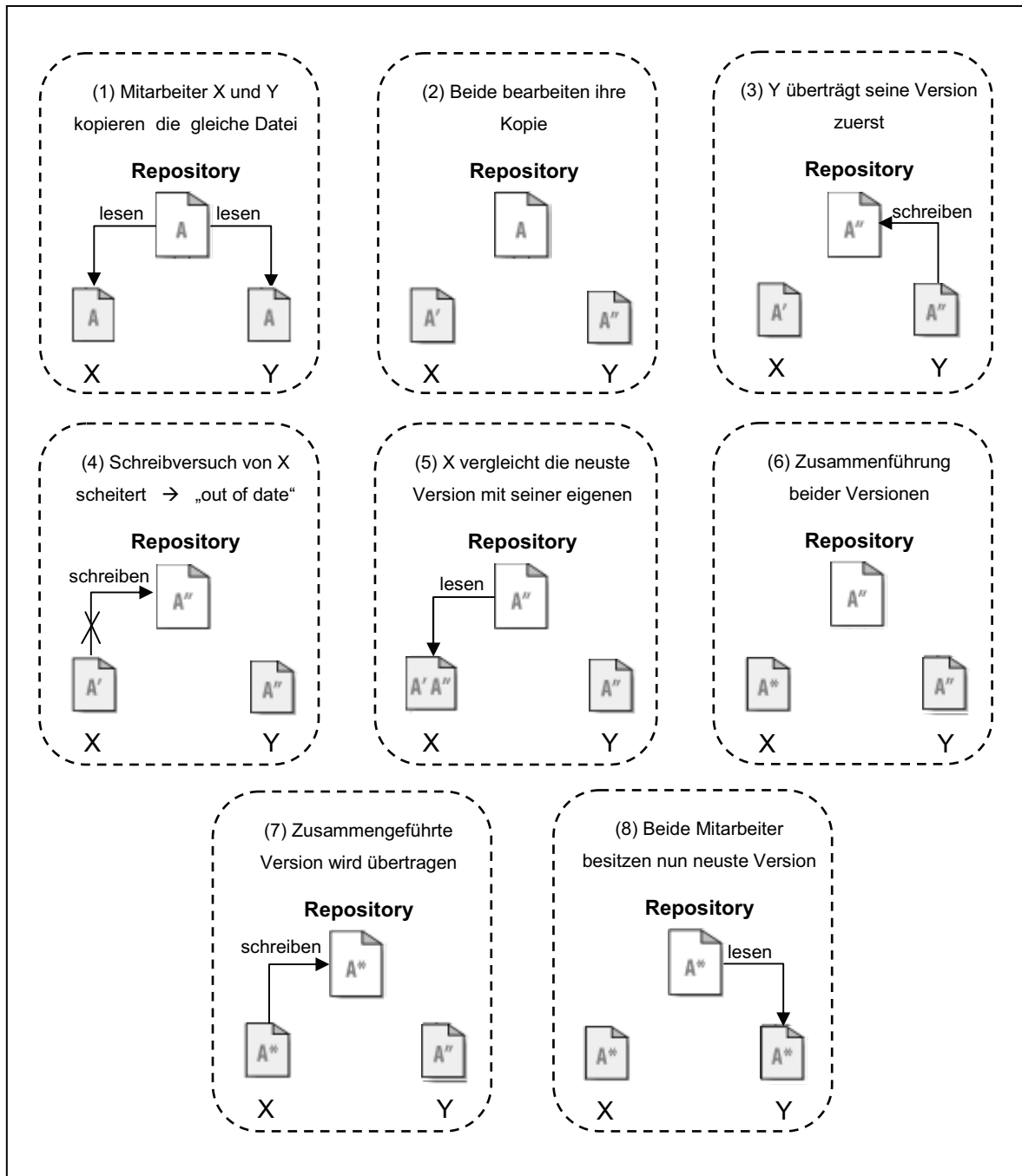
In diesem Modell lädt der Mitarbeiter X die gewünschte Datei aus dem Repository und sperrt somit die Datei, damit sie nicht zwischenzeitlich von anderen Mitarbeitern verändert werden kann. Der Mitarbeiter X kann nun die Datei lokal bearbeiten und verändern. Andere Mitarbeiter können nur noch lesend auf die Datei zugreifen. Erst nachdem der Mitarbeiter X die Datei wieder freigibt, kann ein anderer Mitarbeiter die Datei sperren und Änderungen vornehmen. Somit kann nur eine einzige Person eine Datei zu einem Zeitpunkt verändern. In der Abbildung 4 ist dieses Modell vereinfacht dargestellt.

Abbildung 4: Modell Sperren-Ändern-Freigeben³

Kopieren-Ändern-Zusammenführen

Bei diesem Modell kann jeder Benutzer zu jeder Zeit Dateien vom Versionskontrollsystem laden, unabhängig davon, wer noch mit den Dateien arbeitet. Im Gegensatz zum Sperr-Modell erhält jeder Mitarbeiter nur eine Kopie der Datei. Solange die Version, an der der Benutzer seine eigenen Änderungen vorgenommen hat, mit der entsprechenden Originalversion aus dem Versionskontrollsystem übereinstimmt, wird die neue Datei als neue Version ohne Weiteres übernommen und der Vorgang ist damit beendet. Dabei werden nur Änderungen gegenüber dem Original übernommen. Checkt ein zweiter Mitarbeiter seine geänderten Dateien ein, kommt es zum Konflikt. Die Originalversion aus dem Versionskontrollsystem stimmt nicht mit der aktuellen Version aus dem Kontrollsystem überein. Der zweite Mitarbeiter muss den Konflikt lösen, um seine Änderungen ins Repository zu übertragen. Dies geschieht zum Beispiel durch das Zusammenführen beider Versionen. In der nachfolgenden Abbildung ist dieses Modell dargestellt.

³ Quelle: [SVN]

Abbildung 5: Modell Kopieren-Ändern-Zusammenführen⁴⁴ Quelle: [SVN]

3.3 Metadaten

Bei der Arbeit mit umfangreichen Geodatenbeständen sind Metadaten unbedingt notwendig. Am einfachsten lassen sich Metadaten als „Daten über Daten“ beschreiben.

In roher Form sind digitale Geodaten praktisch unverständlich. Erst mit Metadaten lässt sich ein effektiver und nachhaltiger Betrieb gewährleisten. Die Erstellung von Metadatenbanken ist ein notwendiger und unverzichtbarer Schritt, um die Transparenz zu vorhandenen Geodaten herzustellen und den Zugang zu Geodatenbeständen wesentlich zu erleichtern.

Auf Grund mangelnder Koordination werden Geodaten oft mehrfach erhoben. Datenbestände bleiben vielfach ungenutzt, da über Umfang, Qualität, Aktualität und Verfügbarkeit nur unzureichend Informationen vorhanden sind.

Für die Beschreibung der Metadaten wird ein Metadatenmodell verwendet. Im Metadatenmodell wird festgelegt, mit welchen Angaben die Daten zu beschreiben sind. Die erforderlichen Angaben lassen sich in fünf Gruppen einteilen, die sich auf folgende Punkte beziehen [Joos 00]:

- Modell der Daten (Objektklassen und Struktur der Daten)
- Herkunft der Daten
- Qualität
- Verfügbarkeit
- Referenzanwendungen.

Die Metadaten stellen somit die Menge aller zur Beurteilung der Einsetzbarkeit relevanten Informationen dar. Sie beschreiben die Geodaten vollständig. Die genaue Struktur der Metadaten ist von dem konkreten Anwendungszweck abhängig.

Aufgrund der hohen Bedeutung der Metadaten haben sich in der letzten Zeit einige Standards entwickelt. Mit Hilfe solcher Standards soll die Speicherung bzw. der Austausch von Metadaten vereinheitlicht werden. Im Folgenden wird auf zwei verbreitete Standards näher eingegangen.

ISO 19115

Der Metadatenstandard ISO 19115 [ISO] legt eine Mindestmenge an Informationen fest, die zu jeder Geoinformationsressource angeboten werden sollte. Neben diesem Minimalsatz (z.B. Titel, Urheber, Datum der letzten Änderung, räumliche Ausdehnung) sieht der Standard allerdings eine Vielzahl weiterer Elemente vor, die eine sehr detaillierte Beschreibung beispielsweise der Struktur und Qualität von einzelnen Datensätzen oder auch ganzen Datenbeständen ermöglichen.

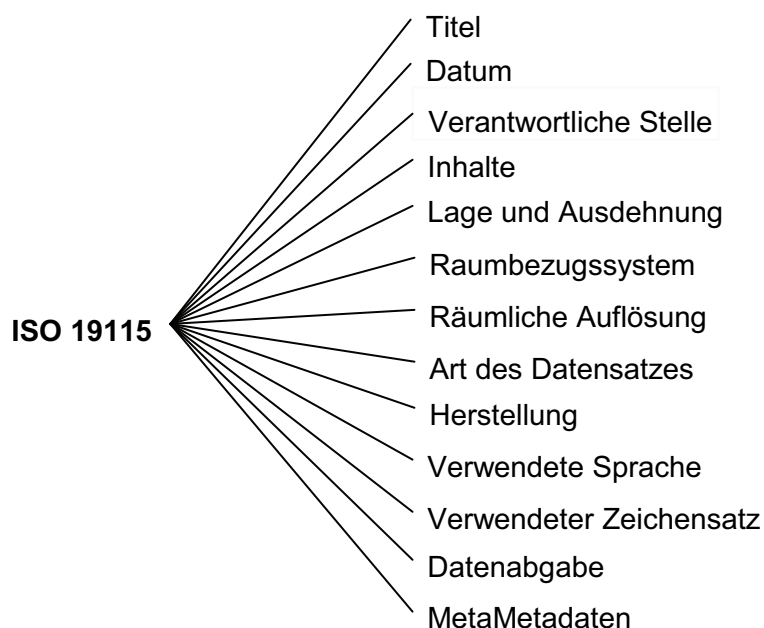


Abbildung 6: Metadaten zur Beschreibung von Geodaten nach ISO 19115 (vereinfachte Darstellung)

Dublin Core

Das Metadaten-Schema Dublin Core [Dublin Core] wurde von der Dublin Core Metadata Initiative (DCMI) entwickelt. Es dient der Beschreibung von Dokumenten und anderen Objekten im Internet. Im Gegensatz zur ISO 19115 beschränkt sich der Dublin Core im Wesentlichen auf 15 Elemente.

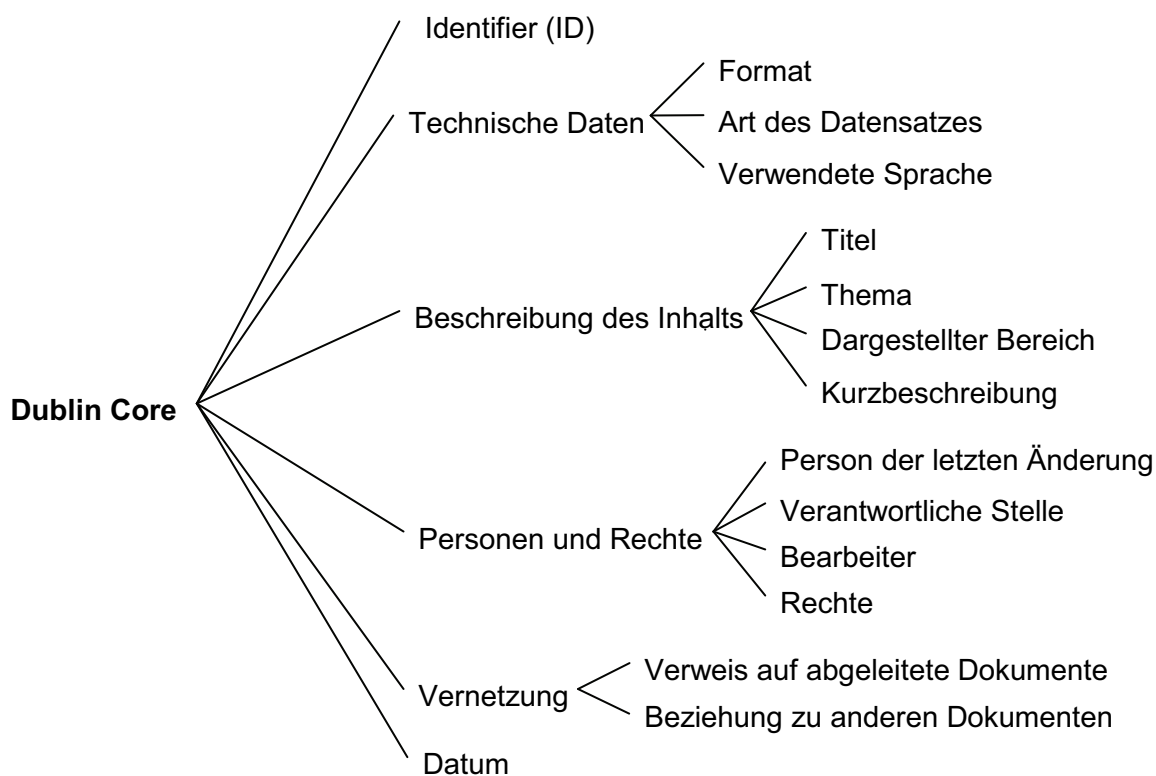


Abbildung 7: Metadaten zur Beschreibung von Geodaten nach Dublin Core

3.4 Nutzergruppen

Insgesamt gibt es zwei Benutzergruppen im Unternehmen, die auf das Speichersystem zugreifen können. Die zwei Benutzergruppen haben jeweils verschiedene Aufgaben. In der nachfolgenden Abbildung sind die einzelnen Nutzergruppen mit ihren Aufgaben dargestellt.

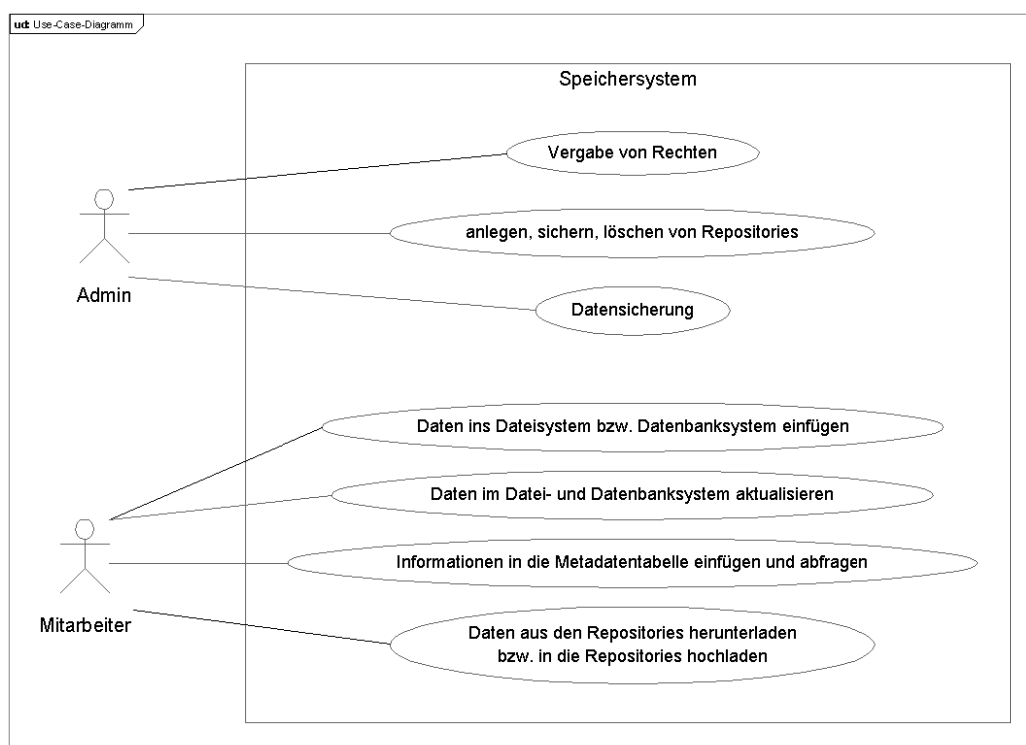


Abbildung 8: Use-Case Diagramm

Administrator

Der Administrator kümmert sich um die Vergabe von Rechten. Er legt fest, mit welchen Rechten die Mitarbeiter und die Geoportalnutzer auf die Geodaten zugreifen können. Des Weiteren ist er für das Anlegen, Sichern und Löschen der Repositories verantwortlich. Eine weitere Aufgabe ist die Datensicherung. Der Administrator kontrolliert die Datensicherung hinsichtlich der Durchführung, Korrektheit und Vollständigkeit.

Mitarbeiter

Die Mitarbeiter haben im Wesentlichen vier Aufgaben. Zum einen haben sie die Aufgabe, die Daten nach der Beendigung eines Projektes in das Dateisystem bzw. in das Datenbanksystem des Servers zu überführen und dort auf einen aktuellen Stand zu halten. Des Weiteren können die Mitarbeiter Informationen in die Metadatentabelle einfügen und nach bestimmten Inhalten abfragen. Während der Bearbeitung der Projekte können sich diese Nutzer aus den Repositories Daten herunterladen und neu erzeugte bzw. aktualisierte Daten hochladen.

Neben diesen zwei internen Nutzergruppen gibt es auch noch die Nutzer des Geoportals. Die Geoportalnutzer verfügen nur über lesende Rechte. Ihre Tätigkeiten haben somit keinen Einfluss auf das Speichersystem.

3.5 Speichermodell

Aus den Anforderungen an das Speichersystem (siehe Kapitel 3.1) und der Analyse der Speichermöglichkeiten (siehe Kapitel 3.2) für Geodaten ergibt sich für das Unternehmen GeoSense folgendes Speichermodell.

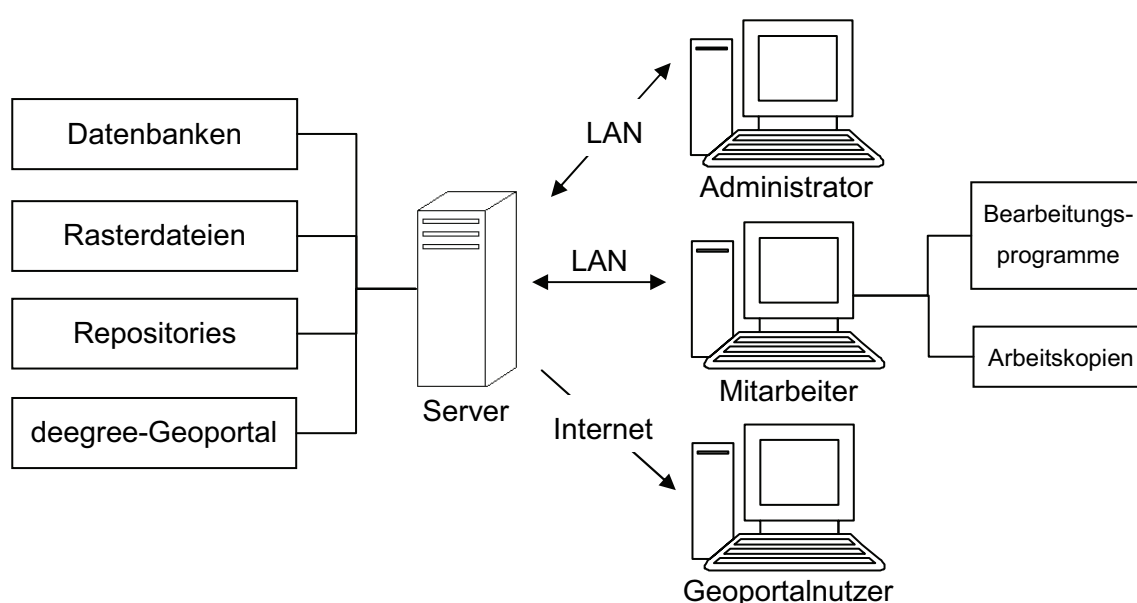


Abbildung 9: Gesamtmodell des Speicherkonzeptes

Auf dem Server werden alle relevanten Geodaten abgelegt. Für das Speichermodell kommt eine Kombination aus Datei-, Datenbanken- und Versionskontrollsystem zum Einsatz. So werden alle für die Projektbearbeitung benötigten Daten im Repository vorgehalten. Rasterdaten, die nach der Bearbeitung für das Geoportal bzw. für nachfolgende Projekte Verwendung finden, werden im Dateisystem gespeichert. Die Vektordaten hingegen werden in einer Datenbank verwaltet. Zusätzlich befindet sich noch die Software deegree für die Darstellung des Geoportals auf dem Server. Über das Internet können die externen Nutzer auf das Geoportal zugreifen. Der Administrator und die Mitarbeiter sind über LAN mit dem Server verbunden. Auf den einzelnen Arbeitsplätzen stehen den Mitarbeitern für die einzelnen Projekte die jeweiligen Bearbeitungsprogramme und die Arbeitskopien der Repositories zur Verfügung.

Auf Grund der veränderten Datenhaltung ändert sich der Datenfluss. Zunächst wird für ein neues Projekt ein Repository angelegt. Anschließend werden die vom Kunden bereitgestellten Daten ins Repository übertragen. Jeder Mitarbeiter lädt sich eine Arbeitskopie des Projektes herunter. Veränderungen in der Arbeitskopie bzw. neue Dateien werden nach der Beendigung eines Arbeitsschrittes ins Repository übertragen. Nach Fertigstellung des Projektes gehen die Bilddateien vom Repository ins Dateisystem über. Alle weiterhin benötigten Vektordaten werden in der Datenbank abgelegt. Das Speichermodell sieht vorerst eine manuelle Übertragung der Daten vom Repository ins Dateisystem und Datenbanksystem vor. Nach Projektabschluss wird eine Sicherheitskopie des Repository erstellt und auf einen externen Datenträger übertragen. Auf dem Server wird das Repository gelöscht. Der neue Datenfluss ist in der nachfolgenden Abbildung dargestellt.

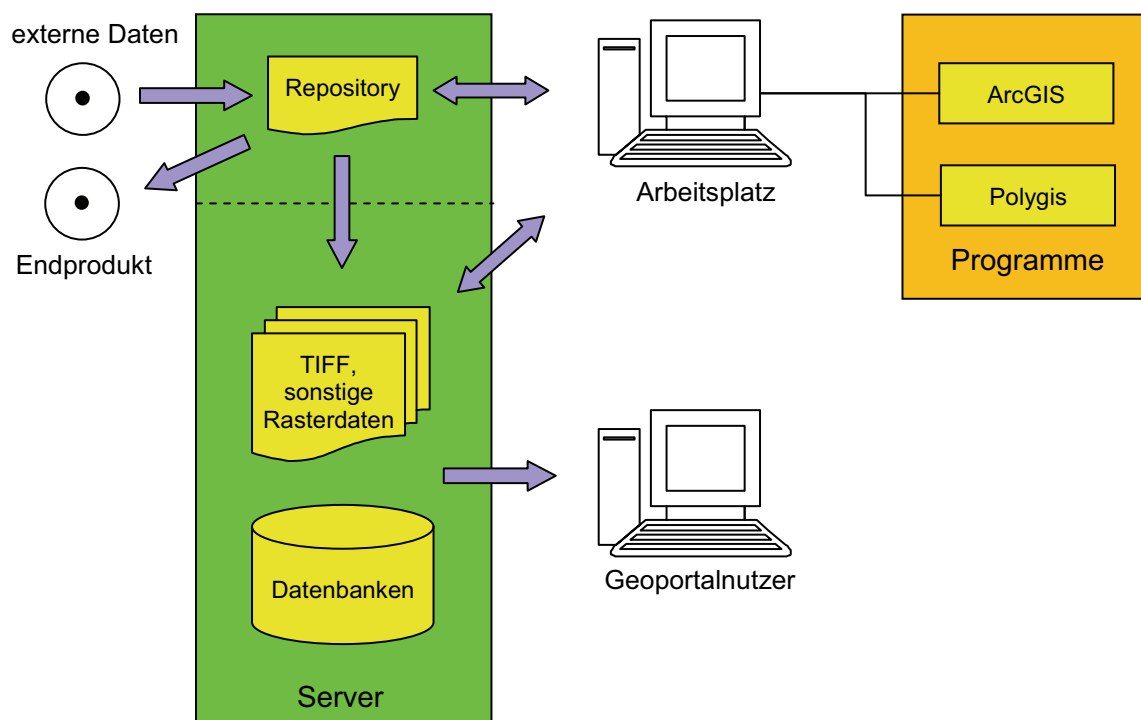


Abbildung 10: veränderter Datenfluss

Im Folgenden wird auf einige Elemente des Speichermodells genauer eingegangen.

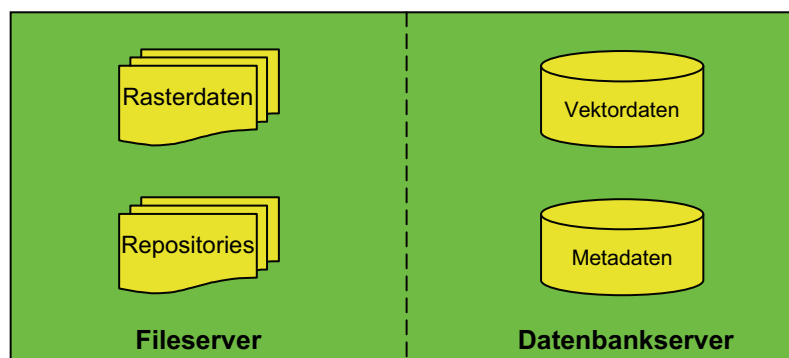


Abbildung 11: Aufteilung des Servers

Die Rasterdaten werden weiterhin im Dateisystem gespeichert. Vor allem die Bilder, bedingt durch ihre Datenstruktur, lassen sich mit einem Dateisystem leichter verwalten. Auch für das Geoportal sind die Bilder als Imagefiles besser geeignet. Auf Grund des benötigten Speicherbedarfs für die Repositories werden diese ebenfalls auf dem Fileserver abgelegt. Auf den Arbeitsplätzen wird dann lediglich eine Arbeitskopie des Repositories heruntergeladen. Die bisher verwendete Verzeichnisstruktur (siehe Abb. 2) kann für das neue Dateisystem fast vollständig übernommen werden.

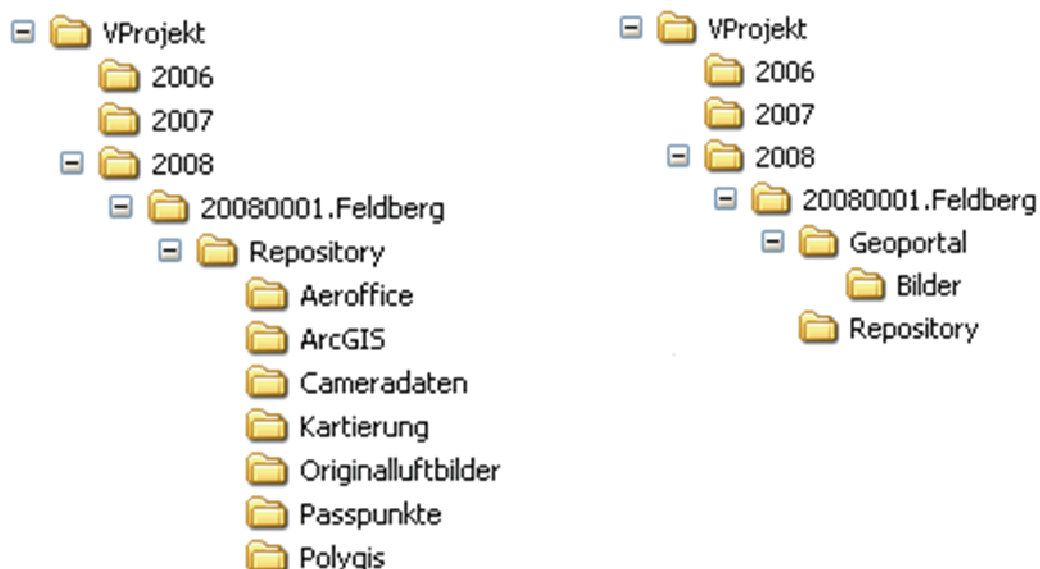


Abbildung 12: Verzeichnisstruktur links: Arbeitsplatz und rechts: Server

Auf dem Datenbankserver werden alle Vektordaten gespeichert, die für zukünftige Projekte und für das Geoportal bereitgestellt werden sollen. Die Daten werden in zwei Datenbanken gespeichert. Jede Datenbank ist einem Themenschwerpunkt zugeordnet: Kartierung und Passpunkte. In der Kartierungsdatenbank werden zum Beispiel Höhenmodelle und Umringe gespeichert. Für jede Datei, die in diese Datenbank überführt wird, wird jeweils eine neue Tabelle angelegt. Je nach Datenformat und Bearbeitungsprogramm variiert die Spaltenanzahl und die Bezeichnung der einzelnen Spalten. Somit kann keine feste Tabellenstruktur erstellt werden. Die Passpunkt-datenbank enthält nur eine Tabelle, in der alle Passpunkte gespeichert werden. Diese Datenbank existiert bereits im Unternehmen und soll ins neue Speichersystem übernommen werden. Zusätzlich befindet sich auf dem Datenbankserver die Metadatenbank ebenfalls nur mit einer Tabelle, in der nähere Informationen zu den Geodaten gespeichert werden.

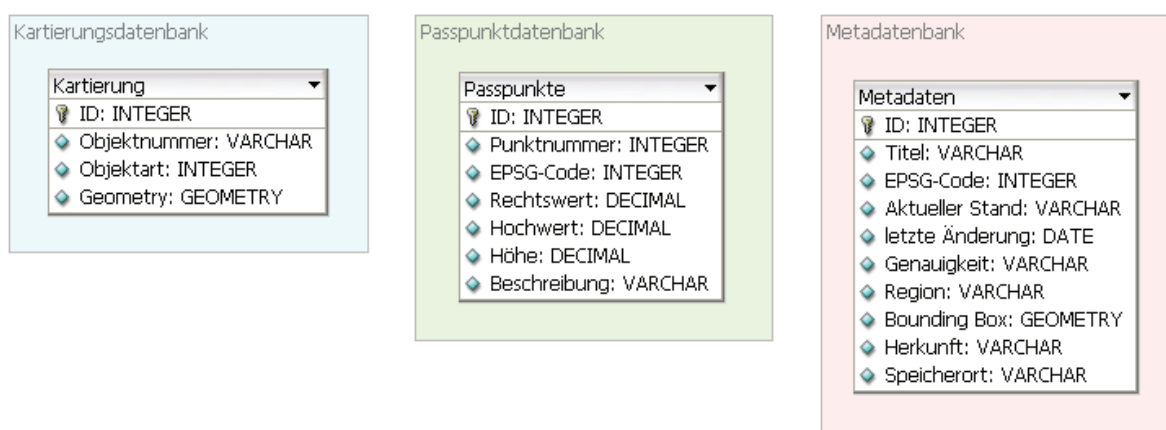


Abbildung 13: Datenbankentwurf

In der oberen Abbildung sind die jeweiligen Datenbanken dargestellt. Für die Kartierungsdatenbank ist eine Beispieltabelle angegeben. Dieses Beispiel enthält jeweils die Mindestanzahl an Attributen, die jede Tabelle in dieser Datenbank hat:

- ID
- Objektnummer
- Objektart
- Geometry.

Jedes Objekt enthält neben der ID-Nummer noch eine Objektnummer, die nicht mit der ID-Nummer übereinstimmen muss. Das Element Objektart gibt Auskunft darüber, ob es sich beim dargestellten Objekt zum Beispiel um eine Brücke oder einen Gullydeckel handelt. In der Spalte Geometry werden der Geometrietyp (z.B. Punkt, Linie) und die Koordinaten eines Objektes gespeichert. Die Aufzählung der vorgenannten Attribute ist nicht abschließend, sondern in Abhängigkeit von den gespeicherten Informationen zu den Objekten ist die Auflistung der Attribute beliebig erweiterbar. Die Tabellennamen setzen sich jeweils aus dem Projektnamen und dem Thema der Kartierung zusammen (z.B. 20080001.Feldberg_Hoehenmodell).

Die Passpunkttable enthält Informationen über die Punktnummer, den EPSG-Code, die Koordinaten mit Rechts- und Hochwert, Höhe sowie eine kurze Beschreibung des Passpunktes. Der EPSG-Code gibt Auskunft darüber, in welchem Koordinatenreferenzsystem die Koordinaten vorliegen. Liegen die Daten nicht im benötigten Koordinatensystem vor, müssen sie vor der Weiterverwendung in das gewünschte System transformiert werden.

Die Metadatenbank enthält eine einzige Tabelle und dient der Beschreibung der einzelnen Datensätze. Die Attribute der Metadatenbank orientieren sich am Minimalsatz des

Metadatenstandards der ISO 19115 (siehe Kapitel 3.3). Folgende Attribute sind in der Tabelle enthalten:

- Titel: Bezeichnung des Datensatzes (z.B. Höhenmodell Feldberg)
- EPSG-Code: Angabe zum Koordinatenreferenzsystem
- Aktueller Stand: Angabe zum Bearbeitungsstand (z.B. fertig, noch in Bearbeitung)
- letzte Änderung: Datum der letzten Änderung
- Genauigkeit: Angaben zur Genauigkeit der Darstellung der Elemente
- Region: welches Gebiet wird abgedeckt
- Bounding Box: Eckkoordinaten des abgebildeten Gebietes
- Herkunft: woher stammen die Daten (externe Anbieter, im Unternehmen erstellt)
- Speicherort: wo sind die Daten gespeichert (Pfad zur Tabelle oder Datei bzw. Angabe des externen Datenträgers).

4 Umsetzung

4.1 Software

Das Konzept zur Speicherung von Geodaten wird mit Hilfe von Open Source Produkten umgesetzt. „Der Ausdruck Open Source (engl.) steht für quelloffen, einerseits in dem Sinne, dass der Quelltext eines Programms frei erhältlich ist, andererseits für 'offene Quelle', also dass ein Werk frei zur Verfügung steht. Software gilt als Open Source, wenn sie bestimmte Kriterien erfüllt, die in ihrer Open-Source-Lizenz geregelt sind.“⁵

Dabei beschränkt sich der Begriff aber nicht ausschließlich auf Software, sondern kann auch auf Wissen und Information allgemein ausgedehnt werden.

4.1.1 Server

Für die Wahl eines geeigneten Fileservers waren folgende Kriterien ausschlaggebend:

- freie / kostenlose Lizenz
- einfache Bedienung
- einfache Nutzerverwaltung.

Diese Kriterien erfüllen nachfolgende Server: Quick 'n Easy FTP Server [Quick], GuildFTPd [GuildFTPd] und FileZilla Server.

Aller drei Server zeichnen sich durch eine einfache Bedienung aus. Rechte von Gruppen und einzelnen Nutzern lassen sich einfach regeln. Vor allem der Quick 'n Easy FTP Server lässt sich besonders einfach bedienen, da sich die Oberfläche stark an Windows orientiert. Ein weiterer Vorteil dieses Servers liegt darin, dass alle Einstellungen in einer XML-Datei abgelegt werden und somit eine Installation unnötig ist. Der Server kann daher von jedem Ort der Festplatte aus gestartet werden. GuildFTPd Server zeichnet sich durch einen großen Funktionsumfang aus. Rechte können sowohl für Gruppen als auch für User definiert werden. Zu den Rechten zählen unter anderem Bandbreitenbeschränkungen, virtuelle Pfade, Timeouts und Anzahl der Verbindungen. Vorhanden sind außerdem Logging-Funktionen, grafische Anzeigen von Verbindungen und Übertragungsraten, Update-Check, Event-Manager usw.. Das Programm kann zusätzlich über Plug-Ins erweitert werden (z.B. Statistik). Trotz der einfachen Bedienung des Quick 'n Easy FTP Servers und des großen Funktionsumfangs des GuildFTPd Servers erhielt der FileZilla Server den Vorzug, da neben dem Server auch ein Client vorhanden ist, der FileZilla FTP Client [FileZilla]. Der Funktionsumfang des FileZilla Servers ist für die Bedürfnisse des

⁵ Quelle: [Open Source]

Unternehmens ausreichend, die Verwendung eines anderen Servers bringt keinen größeren Vorteil für die Firma. Des Weiteren sind der Server und der Client bereits im Unternehmen vorhanden und die Mitarbeiter sind mit der Bedienung vertraut.

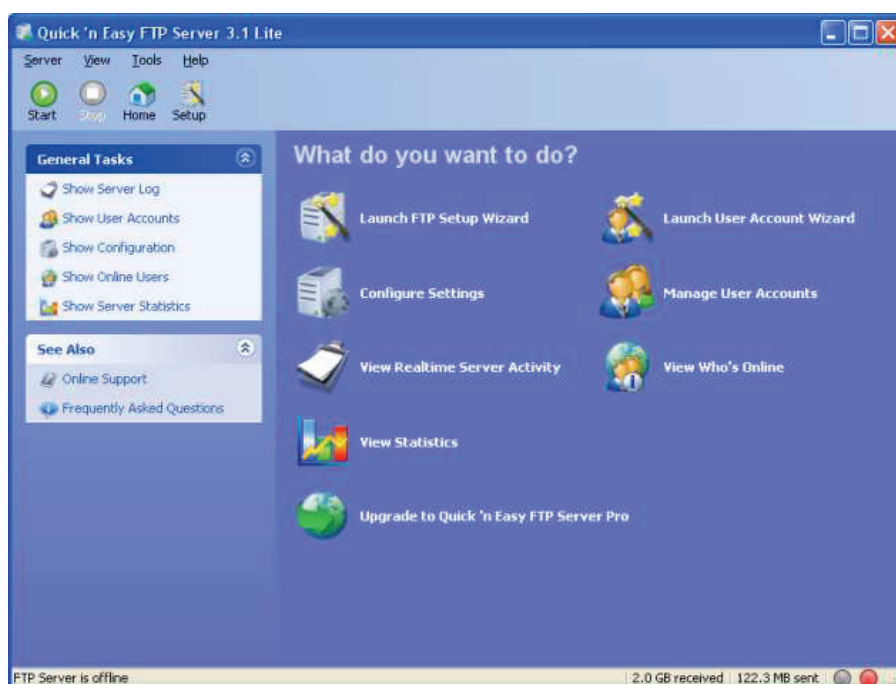


Abbildung 14: Quick 'n Easy FTP Server [Quick]

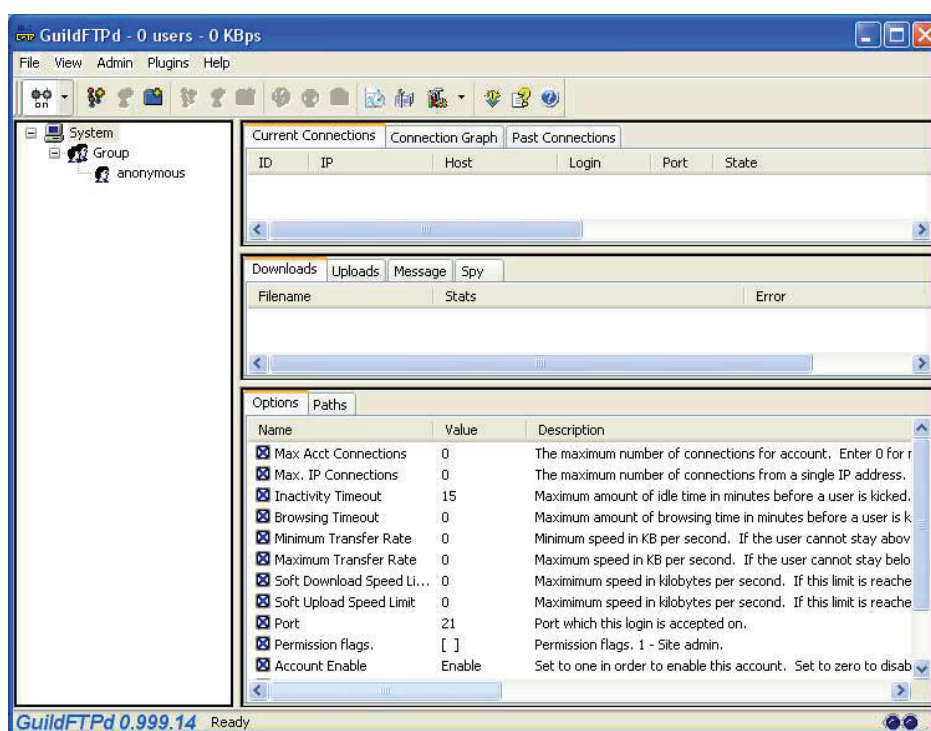


Abbildung 15: GuildFTPd Server [GildFTPd]

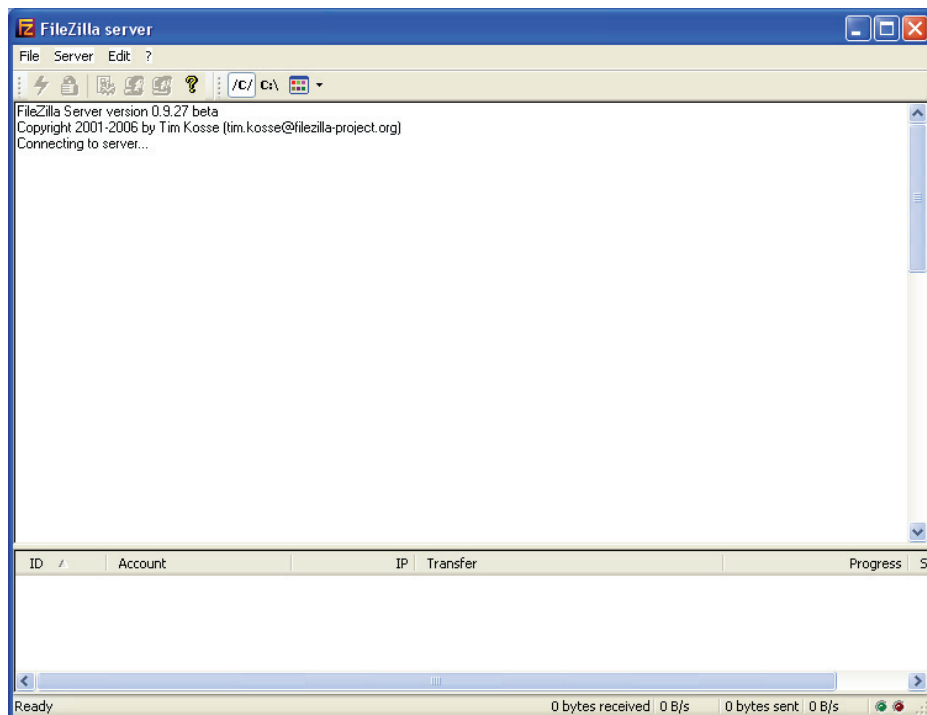


Abbildung 16: FileZilla Server [FileZilla]

4.1.2 Datenbank

Für die Datenbank fiel die Entscheidung auf das Produkt PostgreSQL [Postgre] mit der Erweiterung PostGIS [PostGIS]. Die Erweiterung dient der geographischen Darstellung von Objekten. Folgende Kriterien waren für die Entscheidung ausschlaggebend:

- freie/kostenlose Lizenz
- Unterstützung geometrischer Datentypen und Funktionen
- einfache Anbindung an ein Web Map Service (Geoportal)
- Verbindungen zu ArcGIS und Polygis möglich
- einfacher Datenimport und -export
- Erfahrungen mit diesem Produkt im Unternehmen.

Der Import von Geodaten kann auf unterschiedliche Arten erfolgen. Die Daten können direkt als SQL-Befehle über die kommandozeilenorientierte SQL-Schnittstelle psql oder die grafische Benutzeroberfläche pgAdmin III [pgAdmin] verfasst werden. Die SQL-Befehle können auch als eine Ansammlung innerhalb einer Datei gespeichert und an die SQL-Schnittstelle übergeben werden. Als eine weitere Möglichkeit für den Datenimport und -export stellt PostGIS die Funktionen Shapeloader (shp2pgsql) und Shapedumper (pgsql2shp) bereit. Für den Shapeloader ist auch ein Graphical User-Interface (GUI) vorhanden. Mit diesem GUI (siehe Abb. 17) ist der Datenimport besonders einfach, da

lediglich der Tabellenname, Datenbankname, Host und Login angegeben werden müssen und die Datei automatisch in die angegebene Datenbank eingefügt wird.

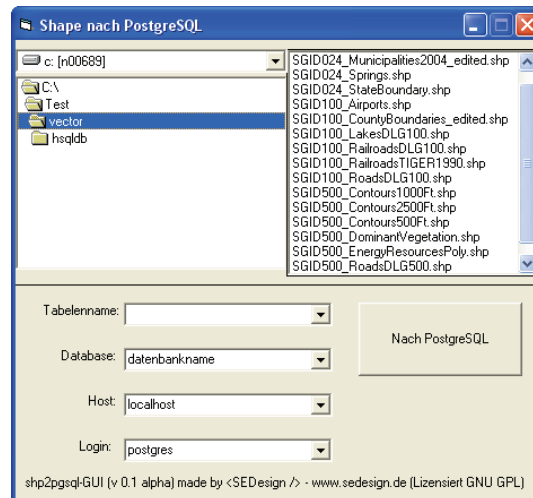


Abbildung 17: shp2pgsql-GUI [shp2pgsql]

4.1.3 Versionskontrollsystem

Im Unternehmen gibt es keine Erfahrungen im Umgang mit Versionskontrollsystemen. Das System sollte daher auch für Neueinsteiger einfach zu bedienen und überschaubar sein. Mehrere Mitarbeiter greifen gleichzeitig auf dieselben Daten zu. Demzufolge muss das System den parallelen Zugriff auf die Daten gewährleisten, ohne dass dabei Inhalte verloren gehen. Des Weiteren muss das System von unterschiedlichen Arbeitsplätzen aus den Zugriff über ein Netzwerk auf die Repositories ermöglichen. Aufgrund der Anforderungen an das Versionskontrollsystem sind für die Wahl eines geeigneten Systems folgende Kriterien ausschlaggebend:

- freie / kostenlose Lizenz
- einfache Bedienung
- Unterstützung des Kopieren-Ändern-Zusammenführen Modells
- Netzwerkzugriff auf Repositories
- Verfügbarkeit von GUIs.

Das Versionskontrollsystem Subversion (SVN) erfüllt diese Kriterien. Im Open Source Bereich gehört Subversion zu den bekanntesten Versionskontrollsystemen. SVN hebt sich vor allem durch die verfügbaren GUIs von anderen Versionskontrollsystem ab. In den letzten Jahren wurde eine Vielzahl von Oberflächen auf Betriebssystem- und Programmenebene entwickelt. Auf zwei GUIs soll an dieser Stelle kurz eingegangen werden.

TortoiseSVN

Bei TortoiseSVN [TortoiseSVN] handelt es sich um eine Erweiterung des Windows-Explorers. Das Programm stellt die gängigsten Subversion-Funktionen im Explorer-Kontextmenü dar (siehe Abb. 18). Der Zustand einzelner Dateien und Ordner einer Arbeitskopie wird mit speziellen Icons abgebildet.

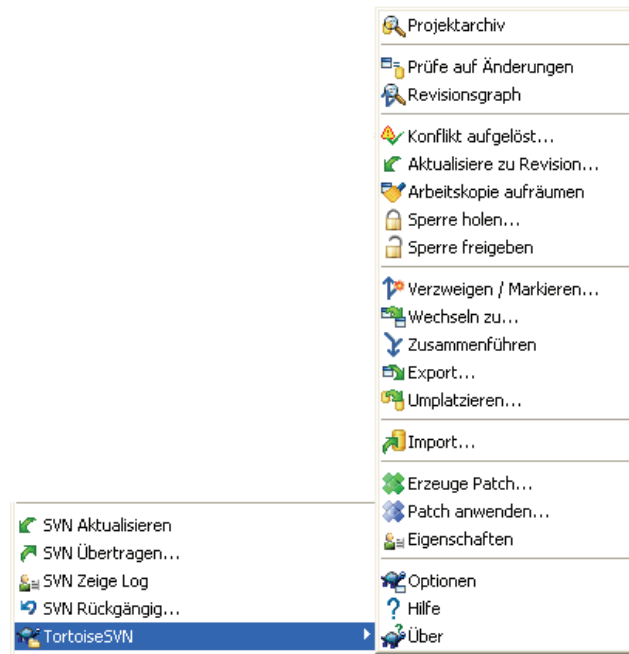


Abbildung 18: TortoiseSVN im Explorer-Kontextmenü

RapidSVN

Im Gegensatz zu TortoiseSVN ist dieser GUI-Client betriebssystemunabhängig. RapidSVN [RapidSVN] wurde mittels der wxWindows-Bibliothek entwickelt. Die Besonderheit bei RapidSVN liegt darin, dass eine C++-API für Subversion implementiert wurde. Subversion selbst bietet nur eine C-API.

Neben den bereits genannten GUIs gibt es unter anderem noch:

- SubCommander (ähnlich RapidSVN) [SubCom]
- Subclipse (für Eclipse) [Subclipse]
- AnkhSVN (für Visual Studio IDE) [AnkhSVN]
- WebSVN (Serverfrontend) [WebSVN].

Im Unternehmen kommt TortoiseSVN aufgrund der einfachen Bedienung und der Integration in das Explorer-Kontextmenü zum Einsatz.

4.2 Nutzung von Subversion

FileZilla Server und PostgreSQL/PostGIS sind bereits installiert, daher soll im Folgenden näher auf den Umgang mit Subversion eingegangen werden. Zum Testen wird Subversion auf einem Arbeitsplatz installiert. Subversion wird in der Version 1.5 [Subversion] verwendet. Nach der Installation wird ein Repository angelegt.

Für den Zugriff auf das Repository stehen drei Möglichkeiten zur Verfügung:

svn: Dabei handelt es sich um ein Subversion eigenes Protokoll. Für den Zugriff ist ein eigener svnserve-Server nötig. Für einen verschlüsselten SSH-Tunnel Zugriff gibt es die Möglichkeit *svn+ssh*.

http: Der Zugriff erfolgt hierbei über das WebDAV-Protokoll und einem entsprechend konfigurierten Webserver. Beim *https* wird der Zugriff zusätzlich durch eine SSL-Verschlüsselung gesichert.

file: Direkter Zugriff auf das Repository über das lokale Dateisystem.

Da eine Nutzung des Repository über das Internet nicht vorgesehen ist, wird der Zugriff über *svn* realisiert. Um den *svnserve*-Server zu starten, wird folgender Befehl auf der Kommandozeile ausgeführt:

```
svnserve -d -r D:\VProjekt
```

-d → Daemonmodus

-r → Basis des bereitzustellenden Verzeichnisses

Im Basisverzeichnis werden alle Repositories gespeichert, die der Server bereitstellen soll. Zum eigentlichen Anlegen des Repository wird das Tool *svnadmin* verwendet. Der Befehl sieht wie folgt aus:

```
svnadmin create D:\VProjekt\2008\20080001.Feldberg\Repository
```

Für jedes Repository wird generell eine Standard-Konfiguration erzeugt, die sich in der Datei *svnserve.conf* befindet. In dieser Datei werden die Zugriffsrechte verwaltet. Nach dem Anlegen des Repository sind die Arbeiten auf der Server-Seite abgeschlossen.

Im nächsten Schritt wird eine Arbeitskopie des Repository auf den lokalen Arbeitsplatz geladen. Danach müssen noch Ordner angelegt werden. Dies geschieht auf der Client-Seite mit TortoiseSVN. Neue Ordner werden über den Menüpunkt *TortoiseSVN* → *Projektarchiv* erstellt. Der weitere Arbeitsablauf sieht wie folgt aus:

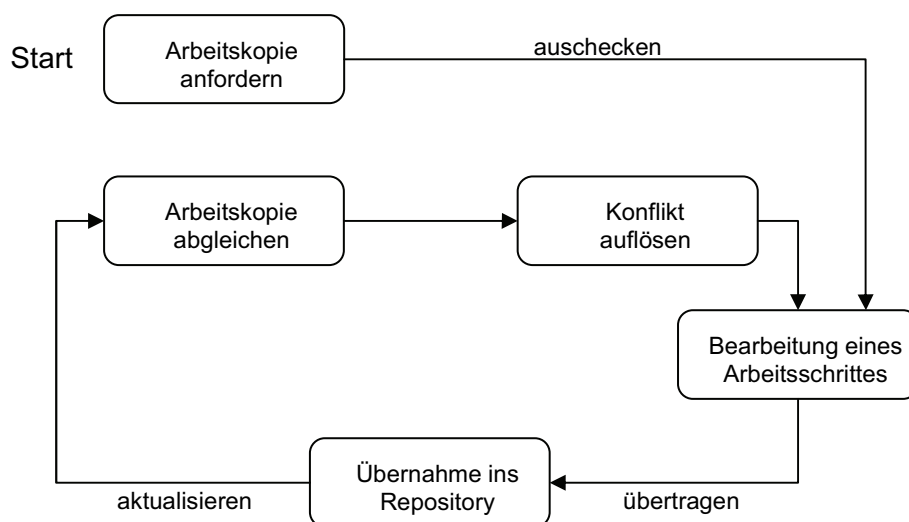


Abbildung 19: Arbeitszyklus mit Subversion

In der nachfolgenden Abbildung ist ein Auszug aus der Arbeitskopie dargestellt.

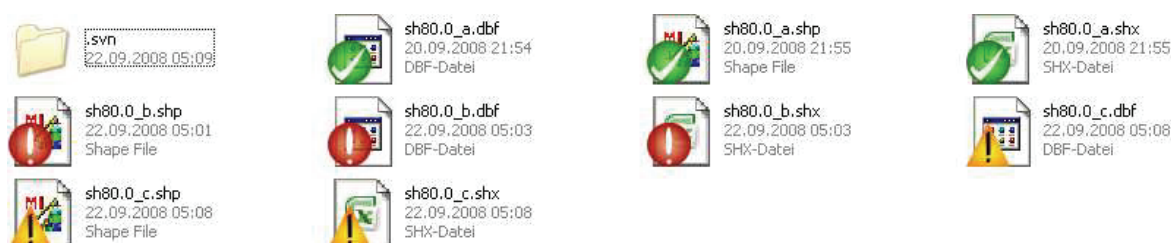


Abbildung 20: Auszug aus der Arbeitskopie

Die grünen Symbole zeigen an, dass die Datei in der Arbeitskopie mit der Originaldatei im Repository übereinstimmt und somit keine Veränderungen an der Datei vorgenommen wurde. Eine Datei mit einem roten Symbol wurde hingegen verändert, aber die Veränderung wurde noch nicht ins Repository übernommen. Gelbe Symbole weisen auf einen Konflikt bei der Übertragung der Arbeitskopie ins Repository hin. Das bedeutet, im Repository existiert bereits eine geänderte Version der Datei. Um den Konflikt zu lösen, muss manuell eine Kopie der Datei vom Repository mit der eigenen Datei zusammengefügt werden. Die von TortoiseSVN bereitgestellte Merge-Funktion ist für die meisten Geodaten aufgrund der Datenformate nicht anwendbar.

5 Backup

Für die Datensicherung im Unternehmen kommen verschiedene Backup-Strategien, Datensicherungstypen und Backup-Medien in Betracht. An dieser Stelle sollen lediglich einige dieser Strategien und Medien näher erläutert und verglichen werden. Welches Backup für das Unternehmen am besten geeignet ist, darauf wird an dieser Stelle nicht weiter eingegangen.

5.1 Backup-Strategien

Um sich vor Datenverlust zu schützen, ist eine regelmäßige Sicherung der Daten wichtig. Dafür ist eine sinnvolle Backup-Strategie notwendig. Es wird unterschieden zwischen vollständigem, differenziellem und inkrementellem Backup.

Typen	vollständig	differenziell	inkrementell
Beschreibung	Es werden alle ausgewählten Daten gesichert und als gesichert markiert.	Es werden nur die Daten gespeichert, die sich seit der letzten vollständigen Datensicherung geändert haben bzw. neu hinzugekommen sind.	Es werden nur die Daten gespeichert, die sich seit der letzten Sicherung geändert haben bzw. neu hinzugekommen sind.
Vorteil	- schnelle Datenwiederherstellung	- geringerer Speicherbedarf auf Backup-Medium, - geringerer Aufwand bei der Datenwiederherstellung im Gegensatz zur inkrementellen Sicherung	- schnelle Datensicherung, - wenig Speicherplatz auf Backup-Medium benötigt
Nachteil	- täglich vollständige Datensicherung und Archivierung → hoher Datensicherungsaufwand und Speicherbedarf	Im Gegensatz zur inkrementellen Sicherung größere zu speichernde Datenmenge und längere Datensicherungszeit.	Hoher Zeitaufwand bei der Datenwiederherstellung, da die letzte Vollsicherung und mehrere inkrementelle Sicherungen hintereinander in der richtigen Reihenfolge eingespielt werden müssen.

Tabelle 4: Backup-Strategien⁶

⁶ Quelle: [Backup]

In der Regel werden mehrere Strategien zu einer Backup-Strategie kombiniert, um eine hohe Datensicherheit zu garantieren.

Für ein vollständiges Backup existieren verschiedene Möglichkeiten, wie Komplettsicherung, Festplatten-Image und Mirroring.

Komplettsicherung

Bei der Komplettsicherung wird der gesamte Datenbestand einer Festplatte gesichert. Für einzelne Rechner oder kleine Netzwerke mit zentralem Fileserver ist dies die sicherste Variante. Der Nachteil dieser Strategie ist zum einen der enorme Zeitaufwand und zum anderen werden unzählige Backup-Medien mit großen Kapazitäten benötigt, da moderne Festplatten 500 GB und mehr an Daten erfassen können.

Festplatten-Image

Festplatten-Image bietet sich für Einzelplatzsysteme an. In wöchentlichen Abständen wird ein Festplatten-Image auf einem externen Datenträger oder auf einer anderen Festplatte angefertigt. Die Struktur einer oder mehrerer Partitionen bzw. der gesamten Festplatte wird 1:1 (inkl. Betriebssystem und Anwendungsprogramme) auf dem Backup-Medium abgebildet. Darin liegt auch der Vorteil gegenüber einer Komplettsicherung. Sollte es zum Beispiel einmal zum Datenverlust oder Hardwaredefekt kommen, wird das Festplatten-Image einfach auf den Rechner kopiert und es kann sofort mit dem Stand des Images weiterarbeiten. Bei normalen Backups hingegen müssen zunächst das Betriebssystem und die Anwendungssoftware neu installiert werden. Der Nachteil dieser Strategie besteht darin, dass nur der letzte Arbeitsstand des letzten Festplatten-Images vorhanden ist. Zwischenzeitliche Änderungen sind somit verloren.

Mirroring

Eine weitere Strategie für Einzelplatzrechner ist die Verwendung eines RAID1-Systems. Für diese Backup-Strategie werden zwei identische Festplatten benötigt. Die Daten der ersten Festplatte werden dabei durch permanente Spiegelung auf die Sicherungsplatte übertragen. Der Vorteil hierbei liegt in der schnellen Verfügbarkeit der gesicherten Daten, da die defekte Festplatte einfach durch eine Mirror-Festplatte ersetzt werden kann. Der Nachteil ist die fehlende Redundanz. Daten, die auf der Hauptfestplatte gelöscht werden, werden auch simultan auf der Mirror-Festplatte gelöscht. Somit ist diese Lösung eher zum Schutz gegen Hardwaredefekte an den Festplatten als zur redundanten Datensicherung geeignet.

Unabhängig von der gewählten Strategie sollten folgende Punkte [Backup] immer beachtet werden:

Regelmäßigkeit: Die Daten sollten immer in regelmäßigen Abständen gesichert werden. Diese Abstände orientieren sich an den jeweiligen Anwendungen.

Aktualität: Die Aktualität orientiert sich an der Anzahl der Datenänderungen. Je öfter wichtige Daten verändert werden, desto häufiger sollten diese gesichert werden.

Ständige Prüfung auf Vollständigkeit: Die Datensicherungen sollten in regelmäßigen Abständen auf Vollständigkeit und Fehler überprüft werden.

Automatische Datensicherung: Manuelle Datensicherung kann durch menschliche Fehler beeinflusst sein. Fehler sind aber auch bei der automatischen Sicherung nicht vollständig auszuschließen.

Verwendung von Standards: Die Verwendung von Standards macht die Datenwiederherstellung einfacher.

Zeitfenster: Der Zeitpunkt des Sicherungsvorgangs sollte wohl überlegt sein, da es unter Umständen zu Problemen in der Produktionsumgebung kommen kann (z.B. Beeinträchtigung des Datentransfers, Zugriffsmöglichkeiten).

5.2 Backup-Medien

In der nachfolgenden Tabelle sind drei Backup-Medien für die Datensicherung aufgelistet.

Backup-Medien	Streamer-Band	CD /DVD	Festplatte
Speicher- kapazität	- einige hundert MB bis über 1 TB	- CD: unkomprimiert ca. 700 MB - DVD: unkomprimiert 4,7 - 8,5 GB	- einige hundert MB bis 1 TB
Vorteil	- preiswert: 0,40 € pro GB - klein, leicht → transportabel	- preiswert: 0,20 € - 0,80 € pro GB - leicht zu transportieren - mit jedem CD-/DVD-Laufwerk nutzbar	- schneller Zugriff, hohe Geschwindigkeit - hohe Lebensdauer und Zuverlässigkeit - preiswert: 0,30 € -0,40 € pro GB
Nachteil	- niedrige Zugriffs-geschwindigkeit - empfindlich gegen magnetische Störfelder	- langsamer Zugriff - relativ geringe Lebensdauer - bei der Sicherung größerer Datenmengen aufgrund der Vielzahl der einzusetzenden Backup-Medien ungeeignet	- nur in externer Variante transportfähig, dabei aber sehr empfindlich gegen Erschütterungen, Vibrationen und Stöße

Tabelle 5: Backup-Medien⁷

Aufgrund der hohen Speicherkapazitäten kommen für die Datensicherung im Unternehmen Streamer-Bänder und Festplatten in Betracht.

⁷ Quelle: [Backup]

6 Zusammenfassung und Ausblick

Ausgehend von den im Unternehmen anfallenden Daten und der Untersuchung verschiedener Speichermöglichkeiten wurde das vorliegende Konzept für die Speicherung von Geodaten entwickelt.

Durch den Einsatz eines Versionskontrollsystems wird gewährleistet, dass unabhängig vom Bearbeitungsstand und der Anzahl der am Projekt tätigen Mitarbeiter an verschiedenen Arbeitsplätzen der parallele Zugriff auf die entsprechenden Daten möglich ist. Dafür wird für jedes Projekt ein Repository auf dem Server angelegt und eine Arbeitskopie auf den Arbeitsplätzen gespeichert.

Für die Nutzung der Daten im Geoportal und für nachfolgende Projekte werden die Rasterdaten im Dateisystem und die Vektordaten im Datenbanksystem auf dem Server gespeichert. Voraussetzung dafür ist die Aufteilung des Geodatenservers in Fileserver und Datenbankserver. Um eine Übersicht über den gesamten Datenbestand zu haben, werden neben den Vektordaten auch Metadaten im Datenbanksystem gespeichert. Mit Hilfe der Metadaten wird die Suche nach bestimmten Daten erleichtert.

Für die Umsetzung des Speichersystems kommen folgende Open Source Produkte zum Einsatz:

- FileZilla Server und FileZilla Client für das Dateisystem
- PostgreSQL / PostGIS und pgAdmin III für das Datenbanksystem
- Subversion und TortoiseSVN für das Versionskontrollsystem.

Das neue Speichermodell ist für das Unternehmen mit einem Mehraufwand für die Mitarbeiter verbunden. Die Mitarbeiter haben neben der Bearbeitung der Projekte zusätzlich die Aufgabe, die Vektordaten in die Datenbanken und die Rasterdaten in das Dateisystem manuell zu übertragen und auf den aktuellen Stand zu halten sowie die Metadatentabelle zu bearbeiten. Der daraus resultierende Nutzen in der Folgezeit, wie zum Beispiel die schnellere Suche nach bestimmten Daten, rechtfertigt diesen Aufwand. Um den Aufwand zu reduzieren, ist die manuelle Übertragung der Daten später durch eine automatische zu ersetzen. Hierfür müssen geeignete Eingabemasken entwickelt werden.

Geoinformationssysteme gewinnen ständig an Bedeutung. Damit verbunden steigt der Umfang der Geodaten in den Unternehmen. Für die Speicherung der bei der Bearbeitung anfallenden Geodaten kann das vorliegende Konzept nicht nur vom Unternehmen GeoSense, sondern auch von anderen Unternehmen genutzt werden. Es müssen

lediglich die Verzeichnisstruktur und die Datenbanken den Gegebenheiten im Unternehmen entsprechend angepasst werden.

Literaturverzeichnis

[ArcGIS 05]	GI Geoinformatik GmbH (Hrsg.) (2005) ArcGIS 9 – das Buch für Einsteiger Herbert Wichmann Verlag - Heidelberg
[AnkhSVN]	http://www.ankhsvn.tigris.org , 12.08.2008
[Backup]	http://www.computer-datenrettung.de , 18.08.2008
[DB]	http://www.wcs.uni-paderborn.de/cs/boettcher/lehre/WS0203/dbis1/dbw99k1.html , 29.07.2008
[deegree]	http://deegree.org , 04.08.2008
[Dublin Core]	http://dublincore.org , 08.08.2008
[FileZilla]	http://www.filezilla.de , 31.07.2008
[Göbel 02]	Göbel, S. (2002): Intelligente graphisch-interaktive Benutzerführung zu Geodaten-Archiven Computer Graphik Edition, Band 8 Fraunhofer IRB Verlag
[GuildFTPd]	http://guildftpd.com , 31.12.2008
[ISO]	http://www.iso211.org , 08.08.2008
[Joos 00]	Joos, G. (2000): Zur Qualität von objektstrukturierten Geodaten Schriftenreihe Heft 66, Neubiberg Studiengang Geodäsie und Geoinformation, Universität der Bundeswehr München
[Open Source]	http://www.ppk.net/knowhow/informationstechnologie/open_source.htm , 29.07.2008
[pgAdmin]	http://www.pgadmin.org , 04.08.2008
[Polygis]	http://www.polygis.de , 04.08.2008
[Postgre]	http://www.postgresql.org , 04.08.2008
[PostGIS]	http://postgis.refractory.net , 04.08.2008
[Quick]	http://www.pablosoftwaresolutions.com/html/quick__n_easy_ftp_server.html , 12.08.2008

[RapidSVN]	http://www.rapidsvn.tigris.org , 12.08.2008
[RaVe]	http://www.geoinformation.net , 29.07.2008
[Subclipse]	http://www.subclipsevn.tigris.org , 12.08.2008
[SubCom]	http://www.subcommander.tigris.org , 12.08.2008
[SVN]	http://www.subversion.tigris.org , 12.08.2008 Dokumentation: Version Control with Subversion For Subversion 1.5 (Compiled from Revision 2983 Version 1.4.4)(2007) Collins-Sussman, B.; W. Fitzpatrick, B.; Pilato, C. M.
[TortoiseSVN]	http://www.tortoisesvn.tigris.org , 12.08.2008
[Wasserman 06]	Wassermann, T. (2006) Versionsmanagement mit Subversion - Installation, Konfiguration, Administration mitp
[WebSVN]	http://www.websvn.tigris.org , 12.08.2008

Abbildungsverzeichnis

<i>Abbildung 1: Datenfluss</i>	3
<i>Abbildung 2: Verzeichnisstruktur</i>	4
<i>Abbildung 3: Datenbanksystem</i>	9
<i>Abbildung 4: Modell Sperren-Ändern-Freigeben</i>	13
<i>Abbildung 5: Modell Kopieren-Ändern-Zusammenführen⁴</i>	14
<i>Abbildung 6: Metadaten zur Beschreibung von Geodaten nach ISO 19115 (vereinfachte Darstellung)</i>	16
<i>Abbildung 7: Metadaten zur Beschreibung von Geodaten nach Dublin Core</i>	16
<i>Abbildung 8: Use-Case Diagramm</i>	17
<i>Abbildung 9: Gesamtmodell des Speicherkonzeptes</i>	18
<i>Abbildung 10: veränderter Datenfluss</i>	19
<i>Abbildung 11: Aufteilung des Servers</i>	19
<i>Abbildung 12: Verzeichnisstruktur links: Arbeitsplatz und rechts: Server</i>	20
<i>Abbildung 13: Datenbankentwurf</i>	21
<i>Abbildung 14: Quick 'n Easy FTP Server [Quick]</i>	24
<i>Abbildung 15: GuildFTPd Server [GildFTPd]</i>	24
<i>Abbildung 16: FileZilla Server [FileZilla]</i>	25
<i>Abbildung 17: shp2pgsql-GUI [shp2pgsql]</i>	26
<i>Abbildung 18: TortoiseSVN im Explorer-Kontextmenü</i>	27
<i>Abbildung 19: Arbeitszyklus mit Subversion</i>	29
<i>Abbildung 20: Auszug aus der Arbeitskopie</i>	29

Tabellenverzeichnis

<i>Tabelle 1: zu verwaltende Bilddateien.....</i>	<i>5</i>
<i>Tabelle 2: weitere zu verwaltende Geodaten</i>	<i>6</i>
<i>Tabelle 3: Datenformate</i>	<i>7</i>
<i>Tabelle 4: Backup-Strategien.....</i>	<i>30</i>
<i>Tabelle 5: Backup-Medien</i>	<i>33</i>