

Datenpflege in Geoinformationssystemen – eine Herausforderung für Geodäten

Martin Scheu

Zusammenfassung

Mit dem schnellen Anwachsen des Bestands an digitalen Geodaten rückt das Thema Datenpflege zunehmend in den Vordergrund. Ohne eine entsprechende Laufendhaltung verlieren Geodaten schnell an Wert, da sie nicht mehr als Basis für Entscheidungen großer Tragweite dienen können. Der Beitrag zeigt auf, an welchen Stellen der Prozess der Datenpflege unter Zuhilfenahme geodätischer Kompetenzen mitgestaltet und optimiert werden kann.

Summary

With regards to the increasing amount of geo-spatial data maintenance becomes a topic that interests GIS community. Lack of topicality causes worse decision support by geo-spatial data. The paper points out that geodetic skills are essential for maintaining GIS-databases. Of central interest is the transfer of incremental update and upgrade data from the custodian to the user.

1 Einleitung

Geoinformationssysteme (GIS) haben einen hohen Verbreitungsgrad in Wirtschaft und Verwaltung erreicht. Mit dem Erreichen dieses Status werden GIS aber auch wie jede andere Anwendung der Informationsverarbeitung zunehmend einer Betrachtung der Relation zwischen dem Nutzen und den laufenden Betriebskosten unterworfen.

Der Nutzen von GIS-Technologie ist heute unbestritten. Er basiert auf den Möglichkeiten der Datenanalyse, der vielfältigen Präsentation und der signifikanten Beschleunigung von Arbeitsprozessen.

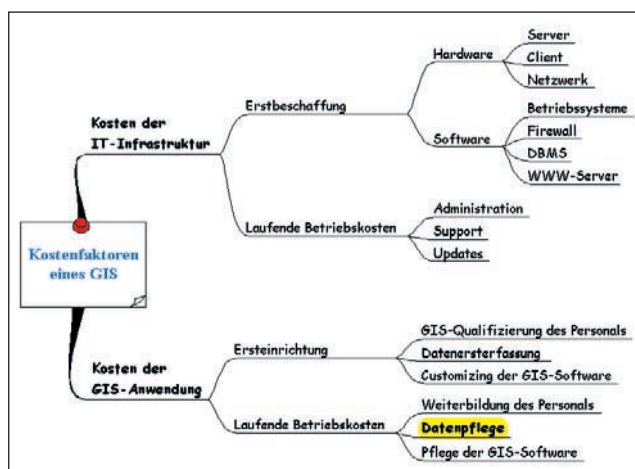


Abb. 1: Kostenfaktoren für die Einrichtung und den Betrieb eines GIS

Diesem geldwerten Nutzen stehen Kosten für die Erst-einrichtung und den laufenden Betrieb eines GIS gegen-über, deren Faktoren in Abb. 1 dargestellt sind. Dabei wird zwischen den einmaligen und laufenden Kosten für die benutzte IT-Infrastruktur und den Kosten der GIS-Anwendung differenziert.

2 Kostenfaktoren eines GIS

In der Abb. 1 werden die Kosten in zwei Faktoren aufge-spalten, deren Anteile sich in Abhängigkeit von der Art der GIS-Anwendung sehr stark unterscheiden.

2.1 Kosten der IT-Infrastruktur

Für die Berechnung der laufenden Kosten einer IT-Infra-struktur existieren verschiedene Modellansätze. Informa-tionstechnologie wird dabei nach spezifischen Kriterien bewertet. Ergebnis sind beispielsweise branchenspezifi-sche Aussagen über die zu erwartenden Kosten pro End-anwender und Jahr.

Zur IT-Infrastruktur kann auch die Datenhaltungs-komponente des GIS gezählt werden. Moderne GIS be-dienen sich hierzu Datenbank-Managementsystemen (DBMS), die neben den GIS-Daten auch die Daten ander-er Anwendungen verwalten. Durch die Möglichkeit der Mehrfachnutzung sind DBMS als ein Teil der IT-Infra-struktur zu betrachten.

2.2 Kosten der GIS-Anwendung

Für das Design der GIS-Architektur, die Systemauswahl, die Modellierung der Daten und die GIS-Einführung lie-gen Erfahrungen vor, die sich in Standardwerken wieder-spiegeln (Behr 1998, Bernhardt 1994).

Auch über die zu erwartenden Kosten für eine GIS-Einführung liegen Erkenntnisse aus einer Reihe von Großprojekten, beispielsweise der Versorgungswirtschaft, vor. Die Aufwendungen umfassen die Beschaffung, An-passung und Einführung der GIS-Software sowie die Datenersterfassung.

Die laufenden Betriebskosten einer GIS-Anwendung werden insbesondere dann zum Gegenstand des Inter-esses, wenn Anwender in Verwaltung und Industrie der internen Leistungsrechnung unterworfen sind. Wie für andere IT-Anwendungen, beispielsweise Bürosoftware,

werden für die Nutzung eines GIS »Service Level Agreements« (SLA) vereinbart, die die Kosten für Support und Softwarepflege kalkulatorisch berücksichtigen.

Damit rückt mit der »Datenpflege« ein Kostenfaktor in den Fokus, der in der Vergangenheit oftmals nicht adäquat berücksichtigt wurde. Die Folge sind heute die auch als »Datenfriedhöfe« bezeichneten Bestände von veralteten GIS-Daten, die keine Grundlage für Entscheidungen großer Tragweite sein können.

3 Teilaspekte der Datenpflege

Der Prozess der Datenpflege schließt sich unmittelbar an die Datenersterfassung an. Generell steigen die Anforderungen an diesen Prozess immer mit dem Wiederbeschaffungswert der Daten selbst. Mit der Komplexität und Vielschichtigkeit der Geodaten steigt naturgemäß der wirtschaftliche Aufwand für deren Laufendhaltung.

Der Prozess der Datenpflege ist klar von dem der Datenmigration zu unterscheiden. Migrationsverfahren werden immer dann angestoßen, wenn Geodaten in ihrer Gesamtheit verändert werden. Anlass hierfür kann beispielsweise ein Wechsel des Raumbezugssystems oder eine komplettes Redesign der Anwendung sein. In der Bundesrepublik bereiten derzeit eine Reihe katasterführender Stellen die Migration der Daten des Liegenschaftskatasters in das ALKIS®-Datenmodell (AdV 2002) vor. Eine Datenmigration ist immer ein projektspezifisch individueller Vorgang, der bestimmten Regeln folgend geplant und durchgeführt werden kann (Scheu und Rose 2002).

Der Vorgang der Datenpflege dagegen ist ein sich regelmäßig wiederholender Arbeitsprozess, bei dem nur ein geringer Anteil der Daten gelöscht, ergänzt oder modifiziert wird. Jeder GIS-Betreiber stellt so dauerhaft die Zuverlässigkeit und Aktualität der Daten nach Abschluss der Ersterfassung sicher. Die hohen Erstinvestitionen in den Datenbestand werden damit nachhaltig vor Verlust geschützt.

3.1 Geo-Basisdaten und Geo-Fachdaten

Für eine systematische Betrachtung der Datenpflege ist es sinnvoll, zwischen Geo-Basisdaten und Geo-Fachdaten zu unterscheiden (Scheu 2000).

Geo-Basisdaten erwirbt der Betreiber eines GIS von der öffentlichen Verwaltung oder von privaten Anbietern und stellt diese in sein GIS ein. Beispiele hierfür sind digitale Rasterkarten, ATKIS-Daten, ALK/ALB-Daten oder GDF-Navigationsdaten.

Geo-Fachdaten dienen der Abbildung von raumbezogenen Objekten der Realwelt, die für die Arbeitsprozesse des GIS-Betreibers essentiell sind, aber kein Bestandteil der Geo-Basisdaten sind. Solche Objekte sind beispielsweise die Filialen einer Handelskette, die Haltepunkte einer Buslinie oder die Betriebsmittel von Leitungsnetzen.

3.2 Fortführungen erster und zweiter Art

Sowohl für die Erfassung der Fortführungsinformationen vor Ort als auch für den Integrationsprozess im GIS ist es sinnvoll, zwischen zwei Arten von Fortführungen zu differenzieren (Scheu et al. 2000).

3.2.1 Fortführungen erster Art

Fortführungen *erster Art* entstehen auf der Grundlage von konkreten Veränderungen der Realwelt, die in einer Veränderung von GIS-Objekten münden.

Erfasst wird, welches Objekt nicht mehr existiert bzw. welches Objekt hinzugekommen ist. Für die Erfassung der geometrischen Informationen neuer GIS-Objekte ist dabei ein bedarfsgerechtes Verfahren zu wählen. Ergebnis einer Neuaufnahme ist eine Objektbeschreibung, die neben dem neuen Objektidentifikator (OID) alle geometrischen und nichtgeometrischen Attribute enthält.

3.2.2 Fortführungen zweiter Art

Fortführungen *zweiter Art* betreffen grundsätzlich nur die Attribute eines bereits erfassten GIS-Objekts. Die OID bleibt stabil, da der Lebenszyklus des Objekts noch nicht beendet ist.

Attribute können sich aufgrund von örtlichen Beobachtungen ändern. Beispielsweise kann sich das Attribut »Kronendurchmesser« für ein GIS-Objekt »Baum« nach einer Begehung des Bestands ändern.

Fortführungen *zweiter Art* können aber auch erforderlich werden, wenn die Integrität der GIS-Daten nach einer Fortführung *erster* oder *zweiter Art* wiederhergestellt werden muss. Ein Beispiel hierfür ist die Wiederherstellung der geometrischen Nachbarschaft zwischen Geobasisdaten und Geo-Fachdaten in einem Netzinformationssystem (NIS) nach erfolgter Veränderung der Koordinateninformation für bestehende Katasterobjekte (Scheu 2000).

3.3 Der Prozess der Datenpflege

Betrachten wir die Pflege von Geo-Fachdaten, so ist der GIS-Betreiber für den gesamten Prozess der Datenpflege selbst verantwortlich. Abb. 2 zeigt vereinfacht den Prozessablauf der Datenpflege von der Informationsgewinnung bis zum fortgeführten GIS-Datenbestand.

Der mit dem einzelnen Fortführungsfall verbundene Aufwand ist von verschiedenen Faktoren abhängig:

- (1) Zeitliche und räumliche Verteilung der relevanten Veränderungen in der Realwelt
- (2) Wahl der Erfassungsmethode für Veränderungen in der Realwelt
- (3) Automationsgrad bei der Integration der Änderungsinformation in das GIS

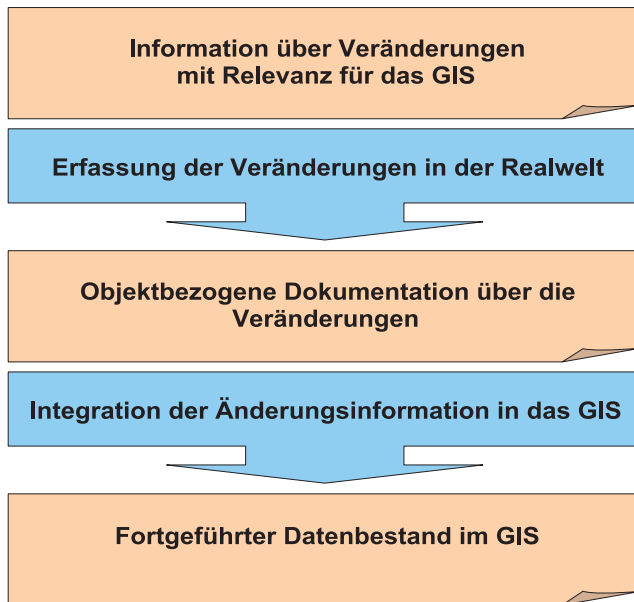


Abb. 2: Vereinfachter Prozessablauf der Datenpflege

Die Faktoren unter (1) werden durch die Art der GIS-Anwendung selbst vorgegeben und bilden einen festen Rahmen für die Arbeitsschritte (2) und (3).

Für die Konzeption und Realisierung der Arbeitsprozesse (2) und (3) sind geodätische Kompetenzen unentbehrlich. Die Vorgänge der Auswahl der Erfassungsmethode und der Integration der Änderungsinformationen in den Bestand an GIS-Daten sollen näher betrachtet werden.

3.4 Geodätische Kompetenz in der Datenpflege

Der Nachweis für die Bedeutung Geodätischer Kompetenz für die Laufendhaltung von Geo-Basisdaten wird seit dem 19. Jahrhundert erbracht. Trotz sich schnell wandelnder technologischer und politischer Randbedingungen konnten die Aufgaben der Katasterführung und der Abbildung der Erdoberfläche bedarfsgerecht erfüllt werden.

3.4.1 Auswahl adäquater Erfassungsmethoden

Für die örtliche Erfassung der geometrischen Attribute ist die geodätische Methodenkompetenz gefordert, zumal die Palette praxisreifer Erfassungsverfahren sich stark erweitert hat. Der Geodät kennt die gesamte Breite der möglichen Erfassungsmethoden: Verfahren der terrestrischen Einzelpunktbestimmung, Oberflächenscanner, Photogrammetrie, Fernerkundung oder satellitengestützte Positionierungsverfahren.

Die geodätische Erfassungskompetenz erlaubt die Auswahl eines adäquaten Aufnahmeverfahrens, das den beim GIS-Betreiber vorgefundenen Voraussetzungen entspricht. Für jedes Fortführungsverfahren sollten diese Randbedingungen im Zuge einer Vorabuntersuchung er-

mittelt werden. Hierbei sind beispielsweise folgende Fragen zu beantworten:

- Welche Merkmale hat der Bestand an Geo-Fachdaten, der fortgeführt werden soll?
- Welches Personal soll eingesetzt werden?
- Welche Investitionen können getätigt werden?
- Welche Datenqualität der Geo-Fachdaten ist erforderlich, um alle darauf basierenden Arbeitsprozesse beim GIS-Betreiber optimal zu unterstützen?
- Welche Geo-Basisdaten werden bei den örtlichen Arbeiten benötigt?
- Welche Maßnahmen der Qualitätssicherung sollen ergriffen werden?

Die Auswertung aller Fragen mündet in Zielsetzungen, die gemeinsam mit dem GIS-Betreiber gewichtet werden, bevor die adäquate Erfassungsmethode gewählt werden kann. Die Technologie ist dann gemeinsam mit dem GIS-Betreiber zu beschaffen, zu parametrisieren und in der Organisation einzuführen.

3.4.2 Integration von Fortführungsdaten in das GIS

Bei der Integration der Fortführungsdaten in das GIS ist zwischen Fortführungen *erster* und *zweiter* Art zu unterscheiden. Bei Fortführungen *erster* Art werden GIS-Objekte gelöscht (DELETE) oder hinzugefügt (INSERT), bei Fortführungen *zweiter* Art existierende Objekte angesprochen, um deren Attribute zu verändern (UPDATE).

In beiden Fällen kann die Datenintegrität des GIS gefährdet werden. Geodätische Verarbeitungskompetenz ist insbesondere dann gefragt, wenn hierbei Herausforderungen mit geometrischem Bezug entstehen.

Ein derzeit aktuelles Beispiel ist hier die Übernahme von in der Realwelt unveränderten Gebäudeobjekten, deren geometrische Lage neu bestimmt wurde. Diese Fortführungen *zweiter* Art führen zu Veränderungen der Koordinatenwerte als Attribute bestehender GIS-Objekte. Dadurch kann das Prinzip der Nachbarschaft nachhaltig verletzt werden. Geodätische Kompetenz ist hier gefragt, um die Interpolationsverfahren der Geodatenhomogenisierung (Gielsdorf und Gründig 1997, Scheu und Philipp 2003) mit einem möglichst hohen Automationsgrad in den Fortführungsprozess zu integrieren.

4 Zur Übertragung für Fortführungsinformationen

Jeder GIS-Betreiber ist heute bestrebt, seine Geo-Fachdaten einer Mehrfachnutzung innerhalb und außerhalb der eigenen Organisation zuzuführen. Diese Art der Verwertung kann zur Refinanzierung der Betriebskosten beitragen.

Auf der anderen Seite stützen sich eine Reihe von GIS-Betreibern auf Geo-Basisdaten, die von externen

Anbietern erstellt und gepflegt werden. Dies senkt den Pflegeaufwand erheblich, macht aber gleichzeitig die fortlaufende Übernahme von digitalen Fortführungsinformationen notwendig. Beide an einem solchen Austauschverfahren beteiligten Institutionen stehen zumeist in einem Vertragsverhältnis, deren Grundlage eine eindeutige Leistungsbeschreibung ist. Diese enthält Aussagen über den Umfang und die Qualität der zu liefernden Fortführungsdaten, die Häufigkeit der Übernahme, das Datenformat und letztendlich die entstehenden Unkosten für die Geodaten und deren Nutzung.

4.1 Anforderungsprofil

Das Übertragungsverfahren sollte bestimmten Anforderungen genügen, um den Aufwand für die Entwicklung und den dauerhaften Betrieb zu minimieren: Hauptziel muss die Maximierung des Automationsgrads bei der Erzeugung der Fortführungsdaten im GIS-Quellsystem und der Integration der Daten im GIS-Zielsystem sein. Der Schlüssel zum Erreichen dieses Ziels ist die Verwendung von Standards.

Generell ist die Verarbeitung von Geodaten zum Gegenstand von Normungsaktivitäten auf nationaler und internationaler Ebene geworden (FIG 2002). Mit der Serie ISO 19100 kommt das ISO/TC211 (<http://www.isotc211.org>) dem Bedürfnis nach Festlegung von Standards für die Verarbeitung von Geodaten nach.

Standards der GI-Industrie, Entwicklungen der IT-Industrie und national festgelegte Austauschformate ergänzen die Vorgaben der ISO, an denen sich jedes Austauschverfahren messen lassen muss. Mit der Erfüllung dieser Vorgaben ist eine universelle Verwendbarkeit und Langlebigkeit des Austauschverfahrens sichergestellt.

Aufbauend auf Standards sollte das gewählte Übertragungsverfahren Fortführungen *erster* und *zweiter Art* in einem einzigen Verfahren abbilden.

Derzeit entstehen vielfach noch Engpässe in den verwendeten IT-Infrastrukturen, die eine Minimierung des Datenvolumens erforderlich machen. Neben der Datenkomprimierung hat die Auswahl des Übertragungsverfahrens für die Fortführungsinformationen selbst entscheidenden Einfluss auf die zu transferierende Datenmenge. Mit der Datenmenge steigen in aller Regel auch die Laufzeiten beim Aus- und Einspielen der Daten in den beteiligten GIS.

4.2 Varianten für die Übertragung von Fortführungsinformationen

In Abhängigkeit von der Art der Geo-Basisdaten, den Möglichkeiten der beteiligten Systeme und der Intensität der Datenpflege können Fortführungsinformationen auf der Basis von drei verschiedenen Strategien transferiert werden (Jacoby 1996):

- Beim *Komplettaustausch (bulk file update)* wird der gesamte Bestand an Geo-Basisdaten zyklisch übernommen. Hier ist aus der Sicht des Endnutzers kein Unterschied zur Erstübernahme sichtbar. Dabei gehen alle nachträglich vorgenommenen Anpassungen immer wieder verloren.
- Beim Austausch von geometrisch oder inhaltlich begrenzten *Teilbereichen (block data update)* tauscht der Endnutzer z. B. einzelne Kilometerquadrate oder Ebenen in seinem Bestand an Geo-Basisdaten aus.
- *Differentielle Verfahren (differential update)* ermöglichen Fortführungen auf der Ebene der Geoobjekte und minimieren so das Volumen der Fortführungsdaten. Die Geo-Basisdaten sind hier objektstrukturiert, ein eindeutiger Objektidentifikator (OID) erlaubt das gezielte Einfügen, Löschen und Verändern von Objekten.

Fortführungsinformationen werden derzeit auf der Basis von Formaten der GI-Industrie oder nationalen Standards ausgetauscht. Dabei kommen alle drei der oben benannten Verfahren zum Einsatz.

Sowohl hinsichtlich des Datenvolumens als auch vom erreichbaren Automationsgrad her bieten die differentiellen Verfahren die meisten Vorteile. Sicherlich werden hier aber auch die höchsten Ansprüche an die Datenstruktur und die beteiligten Systeme gestellt.

Differentielle Verfahren ermöglichen auch die Trennung von Fortführungen *erster* und *zweiter Art* auf der Ebene der Objekte (Scheu et al. 2000). Dies ist auch unbedingt erforderlich, wenn eine automationsgestützte Integration von Fortführungsdaten erreicht werden soll.

Voraussetzung hierzu ist aber eine funktionale Unabhängigkeit der verwendeten Objektidentifikatoren (OID) von den Attributen, die sich bei Fortführungen *zweiter Art* verändern können. Diese einfache Grundregel wird vielfach verletzt und führt zu unnötigem Mehraufwand, wenn beispielsweise die Veränderung einer Koordinateninformation die Vergabe eines neuen Objektidentifikators erzwingt.

5 Zusammenfassung und Ausblick

Mit dem schnell voranschreitenden Aufbau von Geodatenbeständen rückt das Thema Datenpflege weiter in den Mittelpunkt des Interesses. Geodätische Kompetenz liefert wichtige Beiträge zum reibungslosen Ablauf aller Prozesse, die die Erfassung, Verarbeitung, Übertragung und Integration von Fortführungsinformationen zum Gegenstand haben. Aus der Kenntnis möglicher Varianten heraus fällt der Geodät als Ingenieur Entscheidungen mit großer Tragweite für den dauerhaften Erfolg eines GIS-Einsatzes.

Die in diesem Beitrag vorgenommene Unterscheidung zwischen Fortführungen *erster* und *zweiter Art* wird mittelfristig an Bedeutung gewinnen. Das Zusammen-

wachsen interoperabler GIS-Systeme zu Geodaten-Infrastrukturen wird die Spiegelung von riesigen Datenbeständen obsolet werden lassen. Der Endnutzer erhält die gewünschte Kombination von Informationen durch den Direktzugriff auf vernetzte Geodatenserver. Informationen über Fortführungen *erster Art* werden dann nicht mehr ausgetauscht. Fortführungen *zweiter Art* hingegen haben nach wie vor direkten Einfluss auf die Effizienz der Geodaten-Infrastruktur und müssen im Rahmen eines systemübergreifenden Geodaten-Managements behandelt werden.

Geodätische Kompetenzen bilden so einen unentbehrlichen Beitrag für die Ersteinrichtung und Pflege von GIS-Systemen und deren Daten. Die Aufgabe des Geodäten besteht dann darin, die Aktualität und Zuverlässigkeit von Informationen, die zukünftig aus einer vernetzten Geodaten-Infrastruktur heraus gewonnen werden, über die Systemgrenzen hinweg zu wahren.

Literatur

- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik: Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok). <http://www.adv-online.de/>
- Behr, F.J.: Strategisches GIS-Management. Wichmann-Verlag, Heidelberg, 1998.

- Bernhardt, U.: Geo-Informationssysteme in Energieversorgungsunternehmen. VWEW-Verlag, Frankfurt/Main, 1994.
- FIG: International Federation of Surveyors: FIG Guide on Standardisation; How to enhance FIG's role in the process of creating and maintaining official standards. <http://www.fig.net/figtree/>
- Gielsdorf, F., Gründig, L.: Nachbarschaftstreue Anpassung auf der Basis des Membranmodells. ZfV 122, S. 208–218, 1997.
- Jacoby, S.: Incremental update and associated issues. Proceedings of Western Australian Land Information System (WALIS) Forum '96, April 23–24 1996, the Burswood Convention Centre, Australia.
- Scheu, M., Rose, A.: Data Migration of GIS-Systems – The Component View. FIG XXII International Congress Washington, D.C., USA, April 19–26 2002, <http://www.fig.net/figtree/>
- Scheu, M.: Geo-Informationssysteme mit großmaßstäbigem Anwendungsbezug. Habilitationsschrift, DGK Reihe C, Dissertationen, Heft Nr. 522, München, 2000.
- Scheu, M., Effenberg, W., Williamson, I.: Incremental update and upgrade of Spatial Data. ZfV 125, S. 115–120, 2000.
- Scheu, M., Philipp, S.: Geo-Information Management in utilities: A posteriori Integration of Digital Cadastral Maps. FIG Working Week 2003, Paris, France, April 13–17, 2003.

Anschrift des Autors

Dr.-Ing. Martin Scheu
grit GmbH Berlin
Maxstraße 3A
13347 Berlin
scheu@grit.de