

ECO - ein Datenbanksystem zur langfristigen, integrativen Verwaltung von Datenbeständen des Forstlichen Umweltmonitorings

Andreas Schulze

Abt. Umweltkontrolle
Niedersächsische Forstliche Versuchsanstalt
Grätzelstr. 2
37079 Göttingen
aschu@nfv.gwdg.de

Abstract: Ein Datenbanksystem zur universellen Verwaltung forstökologischer Erhebungen wird hinsichtlich der konzeptionellen Grundlagen, der Realisierung und vorliegender praktischer Erfahrungen im Überblick beschrieben.

1 Rahmenbedingungen und konzeptionelle Leitlinien

Monitoring- und Inventur-Programme stellen spezifische Anforderungen an ein adäquates Datenverarbeitungskonzept. Neben der inhaltlichen, räumlichen und zeitlichen Vielfältigkeit der Erhebungsverfahren muss insbesondere der Langfristigkeit der Projekte und einer meist sektoralen Projektdurchführung Rechnung getragen werden.

Die Langfristigkeit zieht neben Personalwechsellern vor allem Methodendynamik durch technische Entwicklungen und wissenschaftliche Erkenntnisse nach sich. Sich damit verändernde, spezifische Randinformationen zu Erhebungen und Messungen (Meta-Informationen) gilt es in Hinblick auf eine zutreffende Interpretation und Auswertung der Daten - insbesondere bei der Betrachtung von Zeitreihen und/oder der Verwendung von Daten durch Dritte - entsprechend zuzuordnen und zu sichern (Integrationsebene 1: Methodik). Daneben vollziehen sich im zeitlichen Verlauf auch immer Veränderungen bei der Datenverwaltungstechnik selbst, was bei langfristigen Projekten auch projektintern zu multiplen, technisch isolierten Datenbeständen führen kann, die es unter Auswertungsaspekten nachträglich zusammenzuführen gilt (Integrationsebene 2: Zeitbezug).

Eine sektorale Projektorganisation zieht parallele, nicht direkt vergleichbare Begrifflichkeiten und Verfahren sowie technisch isolierte Datenbestände nach sich. Im Rahmen von übergreifenden Auswertungen ist es aber häufig erforderlich oder wünschenswert, die Daten unterschiedlicher fachlicher Teilgebiete parallel zu verwenden und die nicht harmonisiert erhobenen Daten miteinander zu verschneiden. Dann werden die Datenzusammenstellungen traditionell durch technische und inhaltliche Heterogenität sehr erschwert. Daher sollten sektoral erhobene Daten, die vorhersehbar miteinander verknüpft

werden müssen, möglichst technisch integriert und fachlich harmonisiert werden (Integrationsebene 3: Thematik). Dies minimiert nebenbei die Gefahr, dass einzelne Datenbestände, insbesondere aus kurzfristigeren Projekten, in Vergessenheit geraten und technisch "verschwinden". Die Notwendigkeit einer Methodendokumentation wird auch bei der sektoren-übergreifenden Integration von Daten augenfällig: abweichende Entstehungsbedingungen oder Bedeutungen der Daten müssen unmittelbar recherchierbar sein, um Fehlverwendungen auszuschließen.

Die Vielfalt denkbarer Auswertungen und Verknüpfungen im Umweltmonitoring legt unter Berücksichtigung der Kosten und Nicht-Wiederholbarkeit der Messungen nahe, die Datenbasis an den Ausgangsdaten auszurichten. Aus Urdaten können im Bedarfsfall unterschiedlichste Zielgrößen auch nach alternativen Verfahren abgeleitet werden. Routinemäßige Datenaggregationen und -verrechnungen einzelner Fachthemen müssen von Endbenutzeranwendungen unterstützt werden.

In Bezug auf die informatorische Umsetzung erfordern die erwähnten Rahmenbedingungen im Umweltmonitoring eine anspruchsvolle Datenmodellierung, die die genannten Integrationsebenen bedient und damit die traditionellen Teildatenmodelle auf einem neuen Abstraktionsniveau zusammenführt. Daneben ist die Entwicklung spezifischer Anwendungsoberflächen für die Nutzer erforderlich. Die Datenverarbeitung ist damit als ein eigenständiges Arbeitsgebiet neben den fachlichen Themen vorzusehen. Auch für die Fachwissenschaftler ergeben sich grundlegende Konsequenzen, in dem Informationsbereiche, die traditionell analog ("im Kopf", in der Feldkladde) geführt werden, vor der Übernahme in das System recherchiert und formalisiert werden müssen. Die Forderung nach technischer Integration sektoral erhobener Daten setzt darüber hinaus ggf. zunächst deren fachliche Harmonisierung, Angleichung oder Umformung voraus.

2 Umsetzung

Die Entwicklung des Datenbanksystems ECO an der Niedersächsischen Forstlichen Versuchsanstalt richtete sich konzeptionell an den genannten Aspekten aus. Als strategisches Ziel wurde die sukzessive Integration der Datenbestände unterschiedlicher Erhebungsniveaus und diverser Projekte des Forstlichen Umweltmonitorings sowie Fallstudien bei einer Nutzungsdauer von mindestens zehn Jahren festgelegt.

Folglich bildete die Modellierung der Datenstruktur einen anfänglichen Arbeitsschwerpunkt, wobei es im Wesentlichen darum ging, die konkreten Begriffe und Vorgänge der realen Projektwelten so weit wie nötig zu abstrahieren und die Entitäten damit universell verwendbar und zukunftssicher zu gestalten. Die entwickelte Datenstruktur nimmt bis auf wenige Ausnahmen keinerlei inhaltliche Auslegung vorweg, d.h. es lassen sich alle im weiteren Rahmen vorstellbaren Inhalte abbilden. Erreicht wurde dies durch die Modellierung Benutzerdefinierbarer Klassen von Datenobjekten, die beliebige Inhalte verwalten können, ohne datenbank-strukturelle Modifikationen zu bedingen [HS97]. Neben der Schaffung der Integrationsfähigkeit selbst wird hierdurch der administrative Pflegeaufwand enorm vermindert. Das Datenmodell wird weiterhin dadurch charakteri-

siert, dass es hierarchische Verweisstrukturen zwischen Teilinformationen aufbaut, die zu Redundanzfreiheit führen (eindeutige Datenlage) und die Abarbeitung aller Informationsteilbereiche in einer logischen Reihenfolge erzwingen (lückenlose Datenlage). Methodische Informationen werden damit in ihrer Bedeutung auf das Niveau der Messdaten gehoben [KS02]. Das Datenmodell wurde unabhängig vom später zum Einsatz kommenden RDBMS bzw. der Programmumgebung entworfen. Die Komplexität der Relationen und die abgeschätzten Datenmengen definierten erst die Anforderungen an die für die Implementierung notwendigen Software-Werkzeuge. Als serverseitiges backend wurde Oracle gewählt, die clients wurden mit Delphi entwickelt.

Die Benutzung der Datenbank setzt grundsätzlich keine Datenbankstruktur- oder SQL-Kenntnisse voraus. Das Design der Anwendungsprogramme und deren Interaktion orientiert sich neben den fachlichen Aspekten direkt an der Arbeitsorganisation spezifischer Teilprojekte und den unterschiedlichen Teilaufgaben, die verschiedene Nutzergruppen darin auszuführen haben [SH97]. Dies fördert insgesamt die Akzeptanz und führt damit zu einer vollständigen Datenlage. Strategisch vollzog und vollzieht sich die Anwendungsentwicklung nicht im klassischen Stil auf Basis von Pflichtenheften, sondern eher iterativ in permanenter, enger Abstimmung zwischen Fachanwendern und Programmierern.

Als software-technische Mittelschicht zwischen abstrakter Datenbankstruktur und spezifischen Applikationen wurden "Virtuelle Tabellen" eingeführt, die einen rationellen und sicheren Vollzugriff auf beliebig definierbare Teilstrukturen und -inhalte erlauben. "Virtuelle Tabellen" sind als PL/SQL-Prozeduren serverseitig implementiert und kapseln die Datenstruktur aus Sicht der Anwendungen vollständig. Der jeweilige SQL-Quellcode kann programmorientiert automatisiert generiert werden [HS98].

3 praktische Erfahrungen

Im nunmehr zehnten Betriebsjahr werden die Bereiche Deposition, Luftqualität, Meteorologie, Bodenhydrologie, Kronenzustand und Standortkartierung routinemäßig verwaltet (ca. 8 - 12 Mio. Datensätze/a). Dies bedeutet die - durchaus erst nachträgliche - Integration von zuvor unter räumlichen, zeitlichen oder inhaltlichen Aspekten in jeweils eigenen Umgebungen verwalteten Datenbeständen, deren Zugriff und Verarbeitbarkeit dadurch enorm verbessert wurde [ES05]. Beispiele sind die integrierte Verwaltung von im Feld erhobenen Mengenwerten und den zugehörigen chemischen Laborergebnissen, die zu Stoffflüssen zu verrechnen sind oder Messungen identischer Parameter mit sehr unterschiedlichen Erhebungsmethodiken, die auf unterschiedlichen Datenpfaden erfasst werden (manuell vs. Datenlogger) und routinemäßig zur Plausibilitätskontrolle vergleichend gegenüberzustellen sind. Die integrierte Datenbasis gewährleistet per se eine hohe formale Datenkonsistenz und ermöglicht neben der einheitlichen Datenverwaltung die Anwendung zentraler, automatisierter Verfahren zur Qualitätssicherung (Plausibilitätsprüfungen, Vollständigkeitsprüfungen) und Datenweiterverarbeitung (Aggregation, Verrechnung), wodurch die Datenqualität und -verlässlichkeit gegenüber individuellen, meist recht fehlerträchtigen Verfahren deutlich verbessert wird.

Seitens der Benutzer wird neben der universellen Verfügbarkeit der Daten besonders der Umstand begrüßt, bei den Datenbank-Inhalten nicht mehr - wie bei dezentralen, eher individuellen Lösungen verbreitet - mit u.U. differierenden Versionen von Informationen konfrontiert zu werden - also sicher zu sein, mit den allein relevanten Daten zu arbeiten. Die entsprechenden Anwendungsprogramme sind inzwischen so weit ausgereift, dass zentrale Arbeitsbereiche weitestgehend ohne Involvierung von IT-Personal durch eingewiesene Mitarbeiter autark gemanagt werden können. Damit wird insgesamt der nicht zu vernachlässigende initiale Aufwand bei der Integration eines Projektbereichs in das System wohl zumindest aufgewogen. Vergegenwärtigt man sich den tatsächlichen, immer wiederkehrenden Aufwand von Datenzusammenstellungen, -konvertierungen und -kontrollen, wie er bei verteilten Systemen unvermeidlich ist, so dürfte mit der System-einführung die Gesamteffizienz bei verbesserter Datenqualität deutlich gestiegen sein.

Parallel wird die technische Effizienz durch das Vorhalten vergleichsweise weniger, zeitlich meist sehr persistenter und universell einsetzbarer Software-Module erhöht. Die strukturelle Stabilität des Datenbankkerns erleichtert darüber hinaus z.B. Spezialabfragen und die Fehlersuche wesentlich. Der Aufwand für die rein technische Administration der Datenbank ist überraschend gering. Das Anwachsen des Datenbestandes (aktuell ca. 120 Mio. Werte) vollzog sich bisher performanz-neutral bzw. waren kritische Teilfunktionen ggf. gut optimierbar.

Ein Ende der Nutzbarkeit zeichnet sich derzeit nicht ab. Eine grundlegende Modell-revision ist als Studie erarbeitet worden. Der Arbeitsschwerpunkt liegt derzeit neben dem kontinuierlichen Ausbau des Systems auf der weiteren Verbreiterung der Datenbasis und der Entwicklung von entsprechenden Applikationen. Ferner wird das System aktuell von 4 weiteren Landes-Forschungseinrichtungen im Rahmen des Forstlichen Umweltmonitorings routinemäßig betrieben bzw. bzgl. seiner Einsetzbarkeit untersucht.

Literaturverzeichnis

- [HS97] Hoppe, J.; Schulze, A.: ECO - Datenbank zur Stoffbilanzierung in Waldökosystemen. Teil 1: Datenbankstruktur und Virtuelle Tabellen. Berichte aus dem Forschungszentrum Waldökosysteme Göttingen, Reihe B 54, 1997; 124 S.
- [SH97] Schulze, A.; Hoppe, J.: Qualitätssicherung bei der Bilanzierung von Stoff- und Wasserflüssen in Waldökosystemen durch datenbankgestützte Arbeitsorganisation. In: (Geiger, W. et al., Hrsg.). Umweltinformatik '97. 11. Int. Symp. Umweltinformatik Strasbourg 1997. Umwelt-Informatik aktuell 15, 1997; S. 216-224.
- [HS98] Hoppe, J.; Schulze, A.: Virtuelle Tabellen als Werkzeug zur spezifischen Abbildung inhomogener Datenbestände in einem homogenen Datenmodell. In: (Hoppe, J. et al., Hrsg.). Vernetzte Umweltinformation. Praxis d. Umwelt-Informatik 7, 1998; S. 125-140.
- [KS02] König, N.; Schulze, A.: Qualitätssicherung im forstlichen Umweltmonitoring durch Methodendokumentation und deren datentechnische Integration am Beispiel der chemischen Analytik. UBA-Texte 66-02, 2002; S. 28-36.
- [ES05] Evers, J.; Schulze, A.: Zur Schlüsselrolle eines integrierten Datenmanagements für die verbesserte Nutzarmachung von Daten des Forstlichen Umweltmonitorings. Zeitschrift für Agrarinformatik 13, 2005; S. 9-17.