

# Fachübergreifende Integration vom Umweltdaten

In: Integration von Umweltdaten, 2. Workshop 1994, 2.-4. Februar 1994, Schloss Dagstuhl.

Thomas Bandholtz  
Sema Group GmbH  
Siegburger Straße 215  
50679 Köln

## Inhalt

1 Ziele und Problemfeld.....	1
1.1 Ziele fachübergreifender Integration.....	1
1.2 Konzeptwandel.....	2
1.3 Ebenen der Integration.....	4
1.4 "private" und "public".....	5
1.5 Dynamik des Informationsbedarfs.....	6
2 Voraussetzungen für Integrationslösungen .....	7
2.1 Systemmodell eines integrativen Informationssystems.....	7
2.2 Informationsmanagement.....	9
2.3 Gestaltung der Schnittstellen.....	10
2.4 Innovation: gilt Murphy's Gesetz?.....	11
2.5 Integration ist Kommunikation.....	12

Der folgende Beitrag ist weniger ein unmittelbarer Projektbericht als vielmehr der Versuch einer Reflexion über die Integrationsproblematik auch gerade in den Aspekten, die über das rein DV-technische hinausweisen. Zu oft hat man in den letzten Jahren den Eindruck gewonnen, daß die Einrichtung einer (möglichst schnellen) Netzverbindung zwischen Rechnern, auf denen sich einzelne Informationssysteme befinden, bereits mit einer Integration *der Informationen* verwechselt wird. Drei Jahre konkrete Projekterfahrung in einem anspruchsvollen Projekt auf Länderebene in zwei Bundesländern<sup>1</sup> haben gezeigt, daß die Integration von Umweltdaten ein erheblich komplexeres Gebiet belegt. Es ist dabei von untergeordneter Bedeutung, wie die Daten *transportiert* werden. Ob Glasfaserkabel oder PC-Diskette per Fahrrad - es ändern sich lediglich die Geschwindigkeit und der Anteil menschlicher Arbeit. Die eigentliche Integrationsleistung hat mehr mit Kommunikation, Modellierung und Semantik zu tun als mit physisch greifbarer Technologie.

## 1 Ziele und Problemfeld

In der Regel ist die Ausgangslage so, daß es eine Reihe von fachlich kompetenten Informationssystemen (FIS) oder wenigstens Datensammlungen gibt, die administrativ und räumlich verteilt vorliegen, entstehen oder entstehen sollen.

Es gibt eine Instanz (z.B. Umweltministerium eines Landes), welche die Ergebnisse der Einzelsysteme insgesamt überblicken will, oder es gibt eine konkrete Aufgabenstellung, bei deren Bewältigung verschiedene Fachgebiete zusammenwirken müssen.

### 1.1 Ziele fachübergreifender Integration

Die Ziele beginnen mit einfachen Vorstellungen und entwickeln sich zwangsläufig zu einer anspruchsvollen Idee, die über den unmittelbaren Anlaß weit hinausweist.

---

<sup>1</sup> Der Autor entwickelte das Konzept für VISION-Umwelt in Niedersachsen und ist derzeit Projektleiter der Sema Group für den Einsatz von VISION-Umwelt in Umweltministerium Sachsen-Anhalt im Rahmen einer Länderkooperation.

### *"Zentraler" bzw. "verteilter" Zugriff*

Von ein und derselben Stelle aus soll der Zugriff auf mehrere FIS möglich sein. Dies allein ließe sich relativ einfach durch entsprechende Netzverbindungen und geeignete Terminal-Emulationen realisieren. An ein und demselben Arbeitsplatz können dann online die verschiedenen Anwendungen neben- oder nacheinander gefahren werden. Dieser Zugang kann an einer einzigen Stelle (zentral) oder auch überall (verteilt) ermöglicht werden.

Diese Lösung bleibt unbefriedigend, weil die einzelnen FIS noch immer unverbunden nebeneinander bestehen bleiben: jedes System behält seine eigene Oberfläche (Dialog und Visualisierung), somit wird die Bedienung durch die ständige Umstellung des Anwenders bis an die Grenze des Zumutbaren erschwert. Eine Navigation quer durch die Fachsysteme ist nicht möglich.

### *Gemeinsame Oberfläche*

Wer einmal in dieser Weise mit verschiedenen Fachsystemen gearbeitet hat, wünscht sich schnell ein durchgängiges Interaktionsmuster (Dialogführung, Begrifflichkeit, Tasten- und Mausfunktionen usw.).

Es reicht dabei nicht aus, daß die verschiedenen FIS mittels Fenstertechnik in einen gemeinsamen "Rahmen", z.B. unter Motif, gebracht werden, wenn die Interaktion in den einzelnen Fenstern unterschiedlich bleibt.

Da sich die verschiedenen Oberflächen nicht auf einen "kleinsten gemeinsamen Nenner" reduzieren lassen, muß die gemeinsame Oberfläche eine *neue* Oberfläche sein, die sich von denen der einzelnen FIS mehr oder weniger löst.

### *Explizite semantische Bezüge*

Bis hierhin haben wir eine durchgängige Navigation allein hinsichtlich der Interaktionsmuster erreicht, die semantischen Zusammenhänge der einzelnen FIS sind jedoch nach wie vor unverbunden. Dies trifft vor allem die Methoden und die Bezugssysteme. Mehrere Aussagen zu einem Umweltindikator aus unterschiedlichen fachlichen Sichten stehen vage nebeneinander. Handelt es sich z.B. tatsächlich um denselben Schadstoff, können Meßergebnisse hinsichtlich ihrer Erhebungsmethoden miteinander verglichen werden?

Bisher existieren nur wenige übergreifend gültige Begriffsdefinitionen (z.B. Abfallartenkatalog, Biotop- und Nutzungstypen). Meßmethoden und Probenamen, statistische Verfahren und Leitwertbildungen weichen stark voneinander ab, und ihre Ergebnisse sind nicht gesichert ineinander umrechenbar.

Die Integration muß nachweisen, in welchem Maße die Daten der Einzelsysteme miteinander in Bezug gesetzt werden können. Dies setzt zunächst eine intensive fachliche Zusammenarbeit der Beteiligten voraus.

### *Fachübergreifende Recherchen und Anwendungen*

Zentraler Zugriff, gemeinsame Oberfläche und semantische Sicherheit sind die Voraussetzungen für eine interessenbezogene Navigation quer durch die FIS. Die Frage geht dabei meist aus von einem Wirkungsfaktor (z.B. Schadstoff), der in verschiedenen Medien untersucht wird, oder aber von einem konkreten Raum- und Zeitbezug. Letztlich geht es immer um Wirkungsgefüge im Raum über die verschiedenen Umweltmedien hinweg.

Hierbei zeigt sich, daß die beste Integration der sektoralen Daten nicht integrierter sein kann als die Blickwinkel, unter denen diese erhoben wurden. Die Informatik muß hierbei sehen, daß es sich nur sehr begrenzt um ein daten- bzw. DV-technisches Problem handelt.

## **1.2 Konzeptwandel**

Seit den achtziger Jahren zeichnet sich in der Informatik ein grundlegender Konzeptwandel hinsichtlich der Strukturierung komplexer Systeme ab. Der Übergang von der "zentralistischen" zur "offenen" Sichtweise vollzieht sich mit einer gewissen Zwangsläufigkeit. Denn einerseits stoßen einzelne Systeme, die von verschiedenen Zentren aus aufgebaut wurden, zunehmend aneinander, ohne daß eines von ihnen sich als neues Metazentrum etablieren kann. Und andererseits zeigt in den Systemen selbst, daß der zentralistische Ansatz mit zunehmender Komplexität immer schwerer durchzuhalten ist.

Man gesteht also, mehr oder minder schweren Herzens, den Sub- und Teilsystemen eine gewisse Autonomie zu. Die Integration verlagert damit ihren Schwerpunkt von der Vorab-Modellierung in der Systemplanung zur permanenten, die Fortentwicklung der Systeme begleitenden Aufgabe.

Sie wird damit selbst zu einem Teilsystem in der Landschaft der offenen, verteilten Systeme, sie ist im Grunde jetzt ein offenes System wie alle anderen, nur eben mit der spezifischen Aufgabe, bestehende Teilsichten miteinander in Verbindung zu bringen.

Die jeweils zu betrachtenden Systeme werden dadurch kleiner, denn ich sehe nicht mehr ständig das (oft imaginäre) Gesamtsystem, sondern betrachte Teilsysteme und die Verbindungen zwischen ihnen. DV-technisch hat dies seine Umsetzung bereits weitgehend gefunden. Begriffe wie "Downsizing" und "Client-Server"-Modell bilden diesen Wandel ab. Der Siegeszug der vernetzten Workstation ist so zu erklären.

Die DV-technische Abbildung einer "offenen" Systemlandschaft eilt jedoch dem entsprechenden Wandel der Informationssysteme voraus: Die Interaktion verteilter, offener *Information* kann nicht daraus bestehen, daß man vernetzte Workstations beschafft und auf jeder den Zugang zum bisherigen Informationssystem installiert.

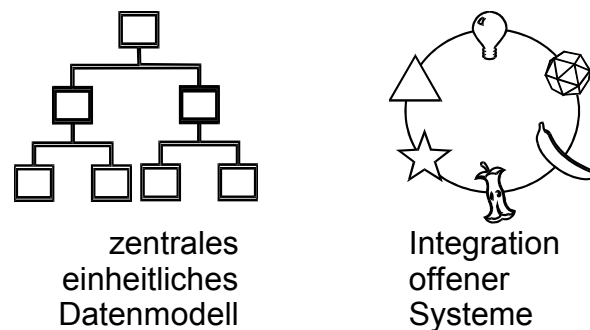


Abb. 1 Konzeptwandel in der Auffassung komplexer Informationssysteme

Die Fortentwicklung von der offenen Systemtechnik zum Einsatz offener Informationssysteme bedingt vor allem eine Neuorientierung der Datenmodelle. Die Idee des *unternehmens- (oder behörden-) weiten Datenmodells* muß sich der Autonomie der Teilsysteme öffnen, ohne diesen blindlings zu verfallen. Dabei gilt die einfache Regel: je mehr das Datenmodell auf die lokale Autonomie eingeht, desto größer wird der Aufwand für die Kommunikation zwischen den Teilsystemen. Hierbei lassen sich drei Ansätze unterscheiden: das ganzheitliche, der methodische und der Schnittstellenkonzept.

#### *Einheitliches Datenmodell*

Die Idealvorstellung eines einzigen, umfassenden Datenmodells, in das sich alle Einzeldaten reibungslos einfügen wie Teile eines aufgelösten Puzzlespiels, ist auch in einer offenen Systemwelt präsent. Besonders in den neuen Bundesländern sind derartige Konzepte wieder aufgeflackert, da man die Chancen für einen "Neubeginn" zu erkennen meinte, der Informationssysteme gewissermaßen "auf der grünen Wiese" errichten könnte.

Derartige Bemühungen vernachlässigen offenbar die Autonomie, welche auch in neuen Fachsystemen faktisch existiert und existieren muß. Jedes Gebiet lokalen Expertentums hat seine eigene Dynamik, seine eigenen Denkstrukturen und häufig auch eine eigene Mentalität. Entscheidend ist letztlich, daß die Arbeit des "Generalisten" derjenigen des "lokalen Experten" ständig hinterherhinken muß.

Der ganzheitliche Ansatz hat trotzdem seine Funktion: Integration kann nur als ständige Iteration zwischen einer ganzheitlichen Vision und der Wahrnehmung einer heterogenen Menge von lokalen Realitäten wirksam werden. Angestrebt werden Vollständigkeit, Eindeutigkeit, Durchgängigkeit und Redundanzfreiheit. Zu vermeiden sind Überlappungen in den Zuständigkeiten. In den Behörden ist das ganzheitlich-hierarchische Modell ja bereits in den Verwaltungsstrukturen verankert, die einen wesentlichen Einfluß auf die Modellierung auch der Informationssysteme ausüben.

#### *Einheitliche Methoden in der Datenorganisation*

Ein etwas bescheidenerer Ansatz liegt in der Vereinbarung konsistenter Methoden für die Datenmodellierung in allen Fachbereichen. Dies kann sich beziehen auf Vorgaben wie: die Verwendung relationaler Datenbankentwürfe, konsequenter Einsatz bestimmter abstrakter Datentypen z.B. für die Ablage von Zeitreihen in Meßnetzen, durchgängige Kodierungsverfahren von Schlüsselsystemen speziell bei hierarchischen Strukturen, einheitlicher Umgang mit Zeitstempeln usw. Wesentlich ist der Einsatz eines Data-Dictionaries, welches die Anwendung solcher Regeln in den lokalen Datenstrukturen beschreibt.

Dieser Ansatz akzeptiert jede lokale fachliche Dynamik und zielt lediglich auf eine Vereinfachung der Kommunikation durch gewisse strukturelle Spielregeln. Die konsequente Einhaltung dieser Regeln

erlaubt es Anwendungen wie z.B. VISION-Umwelt, mit bisher unbekanntem Daten aus verschiedensten Fachgebieten sofort korrekt zu arbeiten.

Derartige Entwicklungen befinden sich jedoch noch im Stadium von Prototypen. Unabhängig davon besteht die Einschränkung, daß sich bestehende "gewachsene" Fachsysteme nur mit großem Aufwand entsprechend restrukturieren lassen. Dieses Modell eignet sich primär für den Einsatz in fachübergreifenden Informationssystemen mit eigener Datenhaltung und für von Grund auf neu zu gestaltende Fachsysteme.

#### *Definierte Schnittstellen*

Dieser Ansatz akzeptiert vollständig jede Autonomie in den Fachsystemen selbst und beschränkt sich auf Regelungen, die den punktuellen Austausch von Daten untereinander ermöglichen. Die aktive Integration erfordert dann stets genau die Transformationen in den Datenstrukturen und oft auch in den Bezugssystemen, die bei der Datenmodellierung vernachlässigt wurde.

Eine gültige *Definition* der Schnittstellen sorgt immerhin dafür, daß die Eingangsdaten formal konstant bleiben, sodaß die Transformation stets derselbe Prozeß sein kann.

Solche Schnittstellen können im Einzelfall bilateral vereinbart werden. Wesentlich effektiver ist es insgesamt, wenn die Schnittstellen weiter vereinheitlicht werden, indem - in ähnlicher Weise wie oben für die Datenhaltung beschrieben - einheitliche Methoden für die Modellierung von Schnittstellen verwendet werden.

Hier gibt es auch international gültige Standards wie z.B. EDIFACT (ISO 9735), auf die sich lokale Systeme abstützen können. Man kann jedoch nicht erwarten, in solchen Normen gebrauchsfertige Lösungen vorzufinden.

### **1.3 Ebenen der Integration**

Die Integration, und besonders die fachübergreifende Integration von Umweltdaten erweist sich damit als äußerst vielschichtiger Prozeß, der weit über die Datentechnik und Datenmodellierung hinausweist.

<i>kulturell</i>	Anwendung Zusammenarbeit Kompetenz
<i>operativ</i>	Datenschutz Vorgangsmanagement Implementierung
<i>semantisch</i>	Modellierung Aggregationsebenen Begriffsdefinitionen
<i>syntaktisch</i>	Austauschformate Datenmodelle Zeichensätze
<i>physikalisch</i>	Netzprotokolle Netzverbindungen Binärformate

Abb. 2 Ebenen der fachübergreifenden Integration

#### *physikalisch*

Binärformate, Netzverbindungen und -protokolle, Blockgrößen, Baudraten usw.

Diese Ebene ist heute grundsätzlich gelöst. Die Problematik im Einzelfall kann auf einen Vorrat erprobter Verfahren und Standards und auf Produkte zurückgreifen. Hier müssen sich die Beteiligten einigen, Lösungen sind unter Umständen zu finanzieren und zu installieren, müssen jedoch nicht neu entwickelt werden.

#### *syntaktisch*

Zeichensätze (Umlaute!), Datentypen, Datensatzstrukturen, Austauschformate usw.

Hier verlassen wir zunehmend das Feld gesicherter Standardisierung. Es ist erstaunlich, welche Probleme bereits die Behandlung von Umlauten und Sonderzeichen in verschiedenen Zeichensätzen heute noch immer aufwerfen kann. Die syntaktische Datenstruktur ist bis auf wenige verbreitete elementare Datentypen (Ganz- oder Gleitkommazahl, Zeichenkette usw.) weitgehend individuell und muß bei der Entwicklung und Anwendung von Austauschformaten beschrieben werden.

### semantisch

Begriffsdefinitionen, Erhebungs- und Bewertungsverfahren, Aggregationsebenen, Raumbezüge.

Hier entsteht ein immenser Handlungsbedarf, auf den bereits unter "Ziele" hingewiesen wurde. Die Ansätze zu Bildung von weithin gültigen Begriffskatalogen und Thesauri stecken heute noch in ihren Kinderschuhen. Die Vergleichbarkeit von Ergebnissen, die mit unterschiedlichen Methoden gewonnen wurden, ist größtenteils ungeklärt. Oft sind auch die Verfahren, die bei der Bildung von Grenz- oder Richtwerten eingesetzt werden, nicht kompatibel mit denen der Datenerhebung.

### operativ

Anwendungsrealisierung, Vorgangssteuerung, Qualitätskontrolle usw.

Die Datenflüsse müssen auch im Sinne einer Vorgangssteuerung beschrieben und kontrolliert werden. Hierzu müssen Verfahren eingeführt und eingehalten werden. Soft- und Hardware ist zu beschaffen bzw. zu entwickeln. Schließlich muß die Integration einer permanenten Qualitätskontrolle unterworfen werden.

### kulturell

Datenschutz, Zuständigkeit, Motivation, Kompetenz, Vereinbarungen, Zusammenarbeit usw.

Datenintegration bedingt auch die Integration von sozialen und kulturellen Strukturen. Die Handelnden und Verantwortlichen sind letztlich Menschen, die bestimmte Ziele verfolgen. Sie haben eine definierte Rolle im Geschäftsverteilungsplan und in der administrativen Hierarchie. Integration setzt hier ein kooperatives Verhältnis voraus.

## 1.4 "private" und "public"



Abb. 3 "private" und "public" im Fachinformationssystem

Für den einzelnen Integrationspartner stellt sich der Zugriff auf "seine" Daten durch andere wie eine Veröffentlichung ("Datenbekanntgabe") dar. Ist Umfang und Verfahren dieses Zugangs zu vage, entsteht zwangsläufig eine rechtliche oder fachliche Verunsicherung, die jeden Integrationsversuch stark belastet. Derartige Bedenken betreffen meist folgende Aspekte:

#### Validität kontra Aktualität

Die Bedenken der Datenquellen werden um so intensiver, je weiter der Zugriff sich den Rohdaten nähert. Die unmittelbar erhobenen Daten enthalten Meßfehler und "gestörte" Werte, die zunächst zu bereinigen sind. In der Tat können Mißinterpretationen von "Ausreißern" zu verhängnisvollen Fehlentscheidungen oder auch Unterlassungen führen. Rein formale Methoden können Unwahrscheinlichkeiten aufdecken, zur verantwortlichen Beurteilung gehört allerdings lokales Expertenwissen. Solche Validierungen erfordern Reihenbetrachtungen, Kontextanalysen und ggfs. Nacherhebungen - und damit Zeit.

Dem entgegen steht das Interesse an *aktueller* Information auf Seiten der Integrationspartner. Wenn man bedenkt, daß z.B. Meßintervalle im Sekundenbereich liegen können, fachliche Bedenken der Meßnetzbetreiber aber noch bei einer 3-stündigen oder gar nur täglichen Bekanntgabe bestehen, und

als eigentlich gesichert nur der Jahresbericht angesehen wird, so versteht man auch den Unmut mancher Integrationspartner, die gegenüber einer solch geduldigen Verantwortlichkeit kein Verständnis mehr aufbringen können.

Der Übergang von analogen (Papier) zu digitalen Medien im Berichtswesen übt hier einen starkern Druck aus, indem er den technischen Publikationszyklus enorm beschleunigt.

#### *Interpretierbarkeit*

Selbst valide Daten können zu Fehlinterpretationen führen. Ungleiche Erhebungsmethoden oder irritierende Umgebungseinflüsse wie jahreszeitliche Schwankungen oder etwa die Stilllegung einer Fabrik sind den Daten selbst nicht ohne weiteres anzusehen. Daher enthalten die regelmäßig publizierten Berichte der Fachsysteme stets Kommentare und zusammenfassende Bewertungen.

Es wird nun befürchtet, daß die Daten aus diesem Zusammenhang gelöst werden, wenn sie in einen fachlich oder räumlich übergreifenden Kontext überführt sind. Aus DV-technischer Sicht kommt erschwerend hinzu, daß die Kommentierungen kaum formal strukturiert sind und sich damit nur wie Textdokumente behandeln lassen.

#### *Schutz vor Mißbrauch*

Ob nun ein einleitender Betrieb nach BIMSChG befürchtet, die Konkurrenz könne über seine Abwässer Rückschlüsse auf Verfahrensgeheimnisse ziehen, ob man die Erfahrung gemacht hat, daß nach einer Weitergabe von Daten zu wissenschaftlichen Zwecken dies Daten gewinnbringend weiterverkauft wurden, ob nach einer unbedachten Bekanntgabe geschützten Artenvorkommens Orchideen und Falkeneier gezielten Plünderungen zum Opfer fielen - jede Weitergabe von Informationen birgt juristische, kommerzielle und ethische Risiken.

Es handelt hier jedoch um die Integration innerhalb einer Organisationseinheit, also z.B. Umweltministerium eines Landes, und nicht um eine Bekanntgabe an die Öffentlichkeit. Entsprechende Strukturen des Datenschutzes können daher verantwortlich entwickelt und eingehalten werden.

## 1.5 Dynamik des Informationsbedarfs

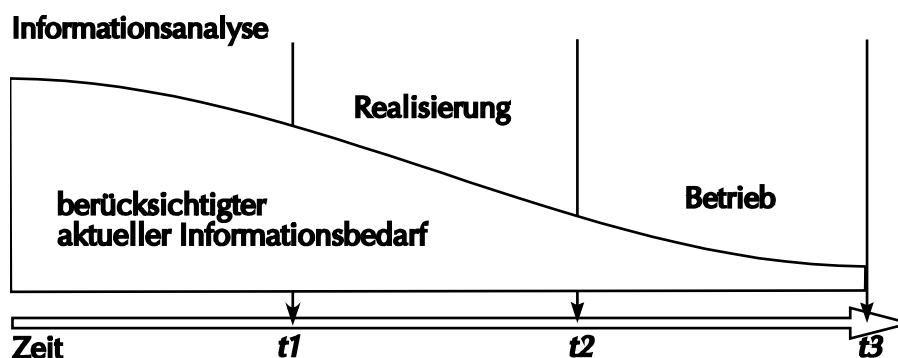


Abb. 4 Dynamischer Informationsbedarf und Phasenmodell

Hinzu kommt ein ungeheuer schneller Wandel der Relevanz, der den einzelnen Themen innerhalb der Umweltinformatik insgesamt zugewiesen wird. Auch dringt die Datenerhebung selbst in immer neue Gebiete vor.

Dagegen sieht das klassische Phasenmodell in der Softwareentwicklung ein Nacheinander von Informationsanalyse, Modellierung, Realisierung und Betrieb vor.

Dies führt dazu, daß das DV-System einen Informationsbedarf zum Zeitpunkt  $t_1$  festschreibt, mit diesem zum Zeitpunkt  $t_2$  in Betrieb geht und schließlich zum Zeitpunkt  $t_3$  seine volle Produktivität entfalten könnte - hätte sich der Informationsbedarf nicht ständig weiterentwickelt.

Wenn solche klassischen Integrationsansätze scheitern, dann tun sie dies in erster Linie, weil sie eine frühe Momentaufnahme des Informationsbedarfs zugrunde legen und sich damit zum Zeitpunkt ihrer Fertigstellung auf Informationsstrukturen beziehen, die schon gar nicht mehr aktuell sind. Die stets begehrten "neuesten" Daten und Zusammenhänge lassen sich mit dem eigentlich ebenso neuen

System schon nicht mehr darstellen und auswerten, und damit ist das System bereits veraltet, bevor es richtig in Betrieb gegangen ist.

Die notwendige Toleranz gegenüber einem sich verändernden Informationsbedarf und -angebot stellt dem integrierenden System eine scheinbar groteske Aufgabe: Es muß bereit sein, auch solche Daten und Sichten zu integrieren, die ihm noch gar nicht bekannt sind.

## 2 Voraussetzungen für Integrationslösungen

Bisher wurde in groben Zügen das Problemfeld umrissen, in welchem die Integration von Umweltdaten sich bewegt. Aus diesen Überlegungen heraus sollen Lösungsbeiträge der Fachinformatik entwickelt werden. Der enge Raum dieses Beitrags zwingt zu einer Konzentration auf wenige Schwerpunkte:

- ☞ ein Systemmodell für nicht nur "offene" sondern *aktiv integrative* Informationssysteme,
- ☞ die Implementierung eines übergreifenden Informationsmanagement-Systems,
- ☞ die Gestaltung der Schnittstellen zwischen den Teilnehmern.

### 2.1 Systemmodell eines integrativen Informationssystems

Im Folgenden soll in groben Zügen ein spezialisiertes Informationssystem umrissen werden, daß sich *ausschließlich* der Integration von Umweltdaten widmet - es verfügt also über keinerlei "eigene" Daten, sondern ist allein damit beschäftigt, die Daten anderer Informationssysteme in einen "neuen" Zusammenhang zu stellen.

In der Praxis mögen heute die Fälle häufiger sein, in denen ein eigenständiges Informationssystem (z.B. Biotope) fallbezogen Informationen aus einem anderen derartigen Informationssystemen (z.B. Altlasten) verwenden möchte. In jedem einzelnen Fall scheinen hierfür pragmatische Regelungen auszureichen, die keinen Eingriff in die Struktur der beteiligten Systeme erfordern. Die Konzentration auf pragmatische Einzellösungen führt jedoch dazu, daß die Integrationsproblematik mit einer bunten Vielfalt individueller Ad-Hoc-Lösungen überzogen wird, die bald nicht mehr zu überschauen und zu pflegen sind.

Daneben entstehen Anforderungen an rein integrative Systeme etwa im Bereich der Führungs- oder Managementinformationssysteme oder auch beim Einsatz von GIS, wo die (kartographische) Methode der Auswertung im Vordergrund steht und nicht mehr ein einzelnes Thema. Diese Ansätze haben aus unserer Sicht den Reiz, daß für sie die Integration das einzige Mittel des Datenzugriffs überhaupt darstellt. Sie sind gewissermaßen "Spezialisten" im Integrations-"geschäft", von denen modellhafte und professionelle Lösungen zu erwarten sind.

Ausgangspunkt eines solchen Systems ist die Idee, nicht die Modellierung eines definierten Informationsbestands anzubieten, sondern eine Art "implementierte Methode", mit der sich die jeweiligen Ausprägungen der aktuellen Informationsbestände ohne Änderungen des eigenen Datenmodells komfortabel und sicher erfassen lassen.

Die Abbildung 5 zeigt eine grobe Struktur aus Anwendungen, integrierten Daten und verteilten Quelldaten.

Beginnen wir mit der mittleren Schicht (dynamisches Datenmodell): Hier sehen wir links ein aktives Data Dictionary (bei VISION das "Objektarten-Verzeichnis", OAV), und zum anderen einen *dynamischen* Satz konkreter Objektdaten, deren aktuelle Struktur eindeutig aus dem aktuellen, deskriptiven Inhalt des OAV ableitbar ist.

Die Anwendungen greifen zunächst nicht auf die aktuellen Objektdaten selbst zu, sondern ausschließlich auf das OAV, also auf die aktuelle Datenbeschreibung. Diese beinhaltet deskriptive und operative Metadaten einschließlich der Zuordnung von Selektions- und Darstellungsmethoden. Die Anwendungen müssen in der Lage sein, sich mit diesen Informationen so zu parametrisieren, daß sie auf die momentanen konkreten Daten vollständig und methodisch adäquat eingehen können.

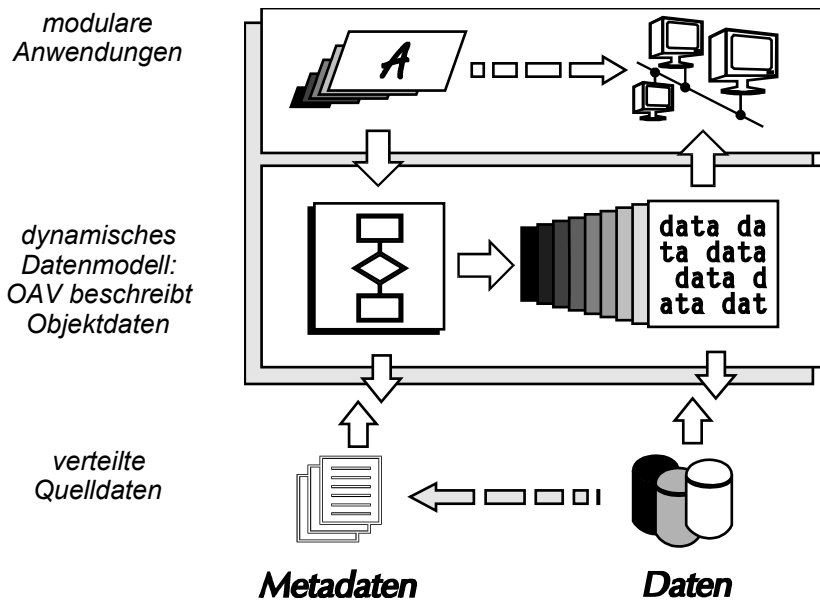


Abb. 5 Systemmodell eines integrativen IS

In der Schicht unten, als Daten-Fundament, stehen die verteilten Quell- oder Rohdaten, die es zu integrieren gilt. Auch hier wird zunächst ausschließlich eine Strukturbeschreibung im OAV abgelegt, woraus eine zwingende Tabellenstruktur für die Aufnahme der Daten selbst abgeleitet werden kann.

Damit das System funktioniert, müssen die Daten nach gewissen normativen Regeln beschrieben werden, sie müssen gewissermaßen in "genormte Container" verpackt werden, mit denen die Anwendungen umgehen können. Das System benötigt vereinheitlichte abstrakte Datentypen, mit denen Selektions- und Darstellungsmethoden implementiert werden können.

Wesentlich ist dabei, daß nicht nur "flache" Tabellen verarbeiten können, sondern komplexe Datenstrukturen bis hin zu attributierten Mehrfach-Relationen wie z.B. bei der Modellierung von Meßnetzen, wo es um die Dimensionen Ort, Zeit, gemessener Parameter, statistische Funktion und Meßintervall geht.

Der Verweis auf die "Views" des relationalen Modells, mit denen sich einzelne vorselektierte Sichten auf die komplexe Modell in Form flacher, "virtueller" Tabellen ablegen lassen, kann diese Beschränkung nicht überwinden. Wieviele solcher Views müssen durch Variierung von 5 Parametern wohl vorab erzeugt werden, um alle Kombinationen zu berücksichtigen?

Auf die *Navigation und Visualisierung* in einem solchen System wird an anderer Stelle ausführlich eingegangen.<sup>2</sup> Der wesentliche Unterschied zu konventionellen Informationssystemen liegt in der Struktur des Dialogaufbaus: Die dynamischen Informationsbestände erzwingen eine Abkehr von der klassischen 1:1-Zuordnung zwischen "Datensatz" und Bildschirmmaske.

Wir verwenden nicht eine fixe Gesamt-Maske für jede Objektart, sondern wenige konstante Masken für die Selektion des Themas und den folgenden Dialog zur Selektion und Ausgabe der Objekteigenschaften über alle Objektarten hinweg.

Die Auswahl einer Eigenschaft aktiviert die Selektionsmethode, die der betreffenden Eigenschaftsklasse zugeordnet ist. Der entscheidende Vorteil: Die Zahl der eingesetzten Dialoge *wächst nicht* mit der Zahl der integrierten Objektarten (=fachlichen Themen). Dies entlastet zum einen den Anwender, indem er sich stets in einem überschaubaren Feld wiederkehrender Interaktionsmuster bewegt. Es entlastet die Entwicklung, da neue Dialoge nur bei grundsätzlich neuen Eigenschaftsklassen geschaffen werden müssen. Und es entlastet das System, weil die Anzahl der zu verwaltenden Dialoge auch auf lange Sicht nicht ins Uferlose wachsen kann.

<sup>2</sup> vgl. in: Denzer u.a. (Hrsg.), 4. Workshop Visualisierung von Umweltdaten, 1994

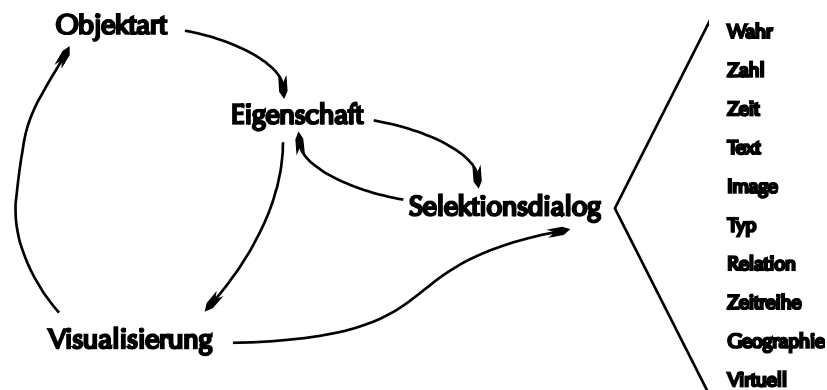


Abb. 6 Grundstruktur der Navigation

Vor allem aber kann das System nur mit solchen Mitteln tatsächlich neuartige Datenbestände integrieren, die zum Zeitpunkt seiner Erstellung noch nicht bekannt waren. Voraussetzung ist, daß die neuen Datenbestände mit den bekannten abstrakten Datentypen beschrieben, kommuniziert und verarbeitet werden können.

## 2.2 Informationsmanagement

Wir können uns daher nicht länger alleine mit dem integrierenden System und einigen Import-Schnittstellen beschäftigen, sondern müssen das Gesamtsystem mit all seinen Teilnehmern und ihrem Informationsaustausch untereinander sehen. Jede Beschränkung auf die Betrachtung bilateraler Einzelfälle, die auch einzeln gelöst werden könnten, scheint nur auf den ersten Blick einfacher zu sein. Sie führt zu unflexiblen und begrenzten Lösungen, die bei jeder weiteren Anforderung erweitert oder neu entwickelt werden müssen.

Es ist effektiver, die Informationsflüsse insgesamt zu untersuchen und so zu klassifizieren, daß wiederverwendbare Typen von Schnittstellen entstehen. Ebenso wird es bei zunehmendem Betrieb sinnvoll, den Datenaustausch über einen spezialisierten Server zu kontrollieren.

Das vorgeschlagene Modell sieht hierfür ein zentrales Vermittlungssystem vor, etwa in der Art einer spezialisierten "mailbox", welches die Schnittstellendefinitionen verwaltet, eingehende Daten zwischenlagert und automatisch verteilt und den planmäßigen Ablauf dieser Vorgänge überwacht.

Die Abbildung 7 zeigt hier einige typische "Rollen" von Teilnehmern im Gesamtsystem. Dort sehen wir zwei Informationsdienste (A,B), welche Daten über Eingangs-Schnittstellen an das aktive Informationsmanagement-System (AIMS) übergeben. Es handelt sich hierbei z.B. um einzelne Meßnetze der Umweltüberwachung.

Daneben stellt C ein "federführendes" System dar, das aufgabenbezogene Informationen anderer Systeme sammelt und zusammenstellt, um die Ergebnisse wiederum an das AIMS zu übergeben.

Das System D bildet ein Beispiel für ein funktionales Server-System: Es bezieht Daten vom IMS, wendet Methoden darauf an und speist Ergebnisse ins IMS zurück. Dabei kann es sich um eine Modellrechnung handeln, um statistische Auswertungen, oder aber auch um einen Archiv-Server.

Schließlich gibt es den Typus des rein integrativen Systems, hier als "Führungsinformationssystem" VISION, welches vom AIMS Daten über mehrere Schnittstellen bezieht, um ein fachübergreifendes Monitoring zu betreiben. Aus der Sicht eines Systems für die Öffentlichkeitsarbeit könnte dieses System auch die Rolle eines Funktionsservers einnehmen.

Wesentlich ist, daß jeder Teilnehmer seine Eingangsdaten, die verarbeitenden Prozesse und seine Ausgangsdaten formal und inhaltlich definiert. Das AIMS erhält nun eine eigenständige Aufgabe. Weder erhebt es Daten, noch "veredelt" es sie, noch integriert es sie, vielmehr verwaltet es den Verkehr der Teilsysteme untereinander. Damit qualifiziert es sich zugleich für die Rolle des Gateways zu den Fernverkehrsnetzen.

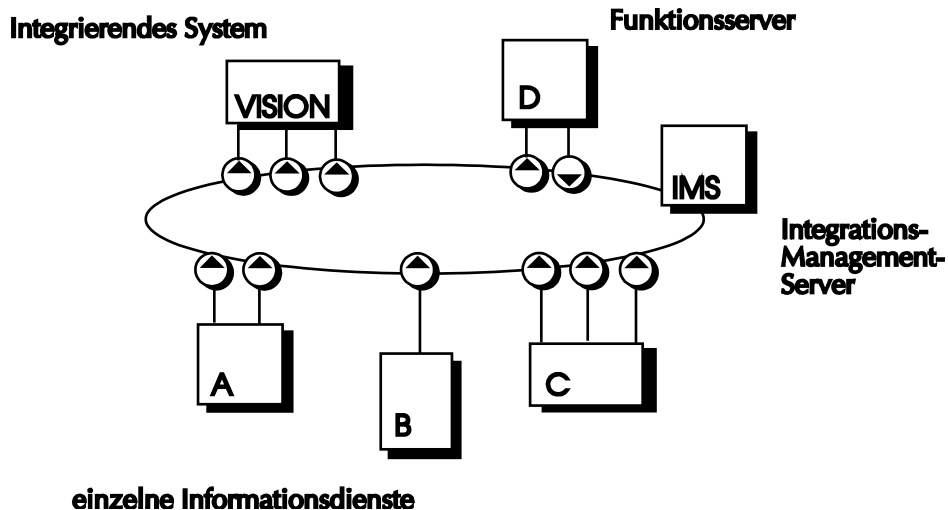


Abb. 7 Rollen im Informationsmanagement (Beispiele)

Für diese Aufgaben benötigt das AIMS Datenstrukturen, die die Strukturen der Kommunikation zwischen den Systemen abbilden können. Hier muß das "Wer, Was, Wann, Wie und Wem" des Datenaustauschs verzeichnet sein, um jederzeit nachvollzogen oder auch nach definierten Regeln initiiert bzw. abgeufen, oder auch angemahnt zu werden.

## 2.3 Gestaltung der Schnittstellen

Es sollte deutlich geworden sein, daß hier nicht der universelle und spontane Zugriff aller Teilnehmer untereinander organisiert werden soll. Die Abkehr vom Ideal des umfassenden Gesamtdatenmodells wie auch die Berücksichtigung einer "privaten" Zone in den Teilsystemen entsprechen viel eher dem, was bisher klassisch unter "Berichtswesen" verstanden wird. Der Austausch definierter und vom Urheber freigegebener Berichte hat sich prinzipiell (auch ohne DV) bewährt und soll optimiert und weiterentwickelt werden. Gut organisierte Fachsysteme dokumentieren ihre Berichtsstrukturen (auch als "Datenbekanntgabe" bezeichnet) bereits seit vielen Jahren in analoger Form. Hier geht es primär um die Einführung DV-technischer Verfahren, die vielleicht aufdecken, daß das bisherige Berichtswesen konsequenter modularisiert und formalisiert werden muß. In anderen Fällen muß zunächst erreicht werden, daß das Wissen um die Datenbekanntgabe, das bisher allein in den Köpfen einiger weniger Spezialisten existierte, explizit festgehalten wird.

Eine solche Schnittstellenbeschreibung sollte Angaben zu den folgenden Themen enthalten:

### *Urheber*

Erhebende Behörde und Ansprechpartner.

### *Inhalt*

Was wird mit den Daten erfaßt, welche Methoden wurden angewendet, usw. Dies ist oft nicht so formalisiert zu beschreiben, wie es ein konventioneller Datensatz erfordern würde. Hier ist ein Freitextdokument (wenigstens ergänzend) erforderlich. Wichtig ist der Verweis auf Bezugssysteme (z.B. Abfallartenkatalog, Soffdatenbanken usw.).

### *Austauschformat*

In jedem Fall soll ein Austauschformat definiert, dokumentiert, nachvollziehbar und wiederholbar sein.

Austauschformate sollten stärker nach formalen Gesichtspunkten als nach inhaltlichen bestimmt werden. Also nicht ein Format für die lufthygienische, ein zweites für die Gewässergüteüberwachung und so fort, sondern ein einziges Format für Zeitreihendaten in räumlich verteilten Meßnetzen (mit dem man "nebenbei" auch die komplette Landkreis- oder Gemeindestatistik erfassen kann).

Austauschformate sollten sich auf ISO- bzw. DIN Standards (wie z.B. EDIFACT, ISO9735) abstützen. Man kann jedoch nicht erwarten, hier stets gebrauchsfertige Lösungen vorzufinden. Der Einsatz von lokal vereinbarten Formaten, die sich an Standards anlehnen, ist oft realistischer als das Abwarten einer umfassenden Normung.

## Zugriffsrechte

Zugriffsbeschränkungen müssen für alle potentiellen Teilnehmer definiert und vom Informationsmanagement überwacht werden. Hierzu müssen systemweit Rechte ("Rollen") und Benutzerklassen vereinbart werden.

## Vorgangsbeschreibung

Es ist zu unterscheiden zwischen periodischen Berichten, ereignisgesteuerten "Episodenverläufen", Informationsdiensten auf Abruf usw., die jeweils unterschiedliche "Auslöser" aufweisen. Das Informationsmanagement und die Empfänger müssen sich auf die entsprechenden (aktiven oder passiven) "Bezugsbedingungen" einstellen können.

## 2.4 Innovation: gilt Murphy's Gesetz?

Mit den bisherigen Mitteln lassen sich vor allem kontinuierliche Informationsflüsse strukturieren, also z.B. periodische Berichte von Überwachungssystemen. Die Zyklen der Berichterstattung können dabei beliebig schnell werden, die Grenze liegt allein in der Leistungsfähigkeit von Hardware begründet. Auch ereignisabhängige "on line" Schaltungen im Krisenmanagement können in dieser Weise implementiert werden. Entscheidend ist, daß die Schnittstelle bereits vorab definiert wurde und daher vom Informationsmanagement und vom Empfänger "bedient" werden kann.

Im Sinne der bekannten Murphyschen Gesetze läßt sich also etwa folgende Regel formulieren: Wie viele Schnittstellen Du auch immer realisiert hast, der erste wirkliche Informationsbedarf wird sich über keine von ihnen abdecken lassen.

Wir sind also gut beraten, wenn wir uns mit der Frage beschäftigen, wie weit und wie schnell *Innovationen* hinsichtlich des Informationsangebots oder auch der Nachfrage mit den bisher bezeichneten Anforderungen vertragen. Man möchte den Aufwand reduzieren, möglichst auch über das Informationsmanagement abdecken und am besten "gleich auf der Schnittstelle" erledigen können.

Der informationstechnische "Trick", an dem sich scheinbar der Münchhausen selbst am Zopfe aus dem Sumpf ziehen möchte, ist eigentlich simpel: Neben Austauschformaten für Meßreihen, Adressen, Anlagenbeschreibungen usw. wird ein weiteres Austauschformat für Schnittstellenvereinbarungen implementiert, mit dem die Ankündigung eines neuen Informationstyps beschrieben werden kann.

Die Voraussetzung besteht dabei in kommunizierbaren "Begriffen", mit denen die Nachricht auf Bekanntes zurückgreifen kann. Dies erscheint noch realistisch, soweit es um den deklarativen Teil allein geht: ein bekanntes Meßnetz kann z.B. über die Schnittstelle mitteilen, daß es jetzt nicht mehr nur in Tagesintervallen, sondern stündlich berichtet. Eine solche Nachricht kann auf der Empfangsseite leicht verstanden werden. Schwieriger jedoch ist es, mit diesen neuen Informationen auch umgehen zu lernen. Ein empfangendes System, das bisher Tageswerte verarbeitet, muß zunächst entscheiden, ob es sich für Stundenwerte überhaupt interessiert, und wenn ja, wie es diese zukünftig in seine Verarbeitungs- und Berichtsprozeduren einbeziehen will. Nicht gerade leichter ist es für einen Informationsdienst, eine über die Schnittstelle formulierte neue Anfrage sofort zu beantworten.

Wie flexibel oder "elegant" auch immer die Schnittstelle Innovationen berichtet, stets haben wir es mit einer Verzögerung auf der Empfängerseite zu tun, die erst lernen muß, auf diese Innovation zu reagieren. Dieser Lernprozeß läßt sich ganz einfach deshalb nicht über die Schnittstelle steuern, weil er nicht vom Absender entschieden werden kann, sondern nur vom Empfänger.

Dieser Vision einer "selbsterklärenden" Schnittstelle steht ein anderes Konzept eher zur Seite als gegenüber: das der *parametrisierbaren Informationsdienste*. Ein Meßnetz könnte neben dem Tages- oder stündlichen Bericht auch einzelne Dienst im Sinne von Abfragefunktionen anbieten, die räumlich, zeitlich oder hinsichtlich der gemessenen Komponenten parametrisierbar sind. Hier geht die Initiative nicht mehr vom Urheber aus, der ein allgemeines unterstelltes Informationsinteresse zu befriedigen sucht, sondern vom Empfänger, der eine konkrete Frage stellt wie z.B.: nach den Stundenwerten von drei bestimmten Meßkomponenten an einem bestimmten Tag, oder auch nach allen Altlastverdachtsflächen mit dem Branchenschlüssel Tankstelle in einem Landkreis.

Auch solche Funktionen müssen in allen Aspekten einer Schnittstelle vorbereitet und implementiert werden. Hier liegt die Hauptlast bei der Datenquelle. Ist eine solche Schnittstellenfunktion aber erst einmal sauber parametrisiert und hinsichtlich des Datenzugriffs und der Übergabe ausgeführt, so kann sie über lange Zeit hinweg in allen erdenklichen Anwendungssituationen genutzt werden.

## 2.5 Integration ist Kommunikation

All diese DV-technischen oder formal strukturierenden Hilfsmittel allein können noch keine Integration von Umweltinformationen bewirken. Hier kann lediglich ein kommunikativer Prozeß unterstützt werden, der von den Anwendern getragen wird.

Die Anwender finden sich dabei mit einer neuen Situation konfrontiert: jeder von ihnen bekommt es mit den Fachgebieten der anderen zu tun, und er muß zulassen, daß sich die anderen mit seinem eigenen Fachgebiet beschäftigen. Die hierbei auftretenden Verständnisschwierigkeiten werden nicht erleichtert, wenn der Zugang zu den Daten in "freier" Weise erfolgt, also ohne den Filter des jeweils zuständigen Fachmanns.

Das Schnittstellenmodell bietet in dieser Hinsicht den Vorteil, daß nur Informationen für die Integration zugänglich werden, die vorher ausdrücklich dafür freigegeben wurden. Ein "Nachfassen" in die private Sphäre des einzelnen Fachsystems ist nicht ohne Abstimmung mit den Fachleuten möglich.

Ein solches Informationssystem versucht also nicht, die interdisziplinäre Beratung durch multimediale Automatismen zu ersetzen. Die rasche Zusammenstellung aller Informationsbeiträge wird erleichtert oder auch erst ermöglicht. Deren Sichtung und Bewertung bis hin zur Entscheidungsreife bleibt weiterhin ein kommunikativer Prozeß zwischen Menschen. Die Steigerung der Verfügbarkeit verteilter Einzelinformationen schraubt die Anforderungen an formelle und informelle Zusammenarbeit eher höher.

Es erscheint daher wichtig, über *organisatorische* Strukturen nachzudenken, die gesicherte interdisziplinäre Betrachtungen auf der Basis der gesteigerten Informationsbreite erleichtern bzw. erst ermöglichen.