

I.

KTB Oberpfalz VB –

Datenverarbeitung

J. Wächter
M. Friese-Haug

KTB-Report	89-2	I1-I20	3 Abb.	Hannover 1989
------------	------	--------	--------	---------------

Tiefbohrung KTB-Oberpfalz VB

I. Datenverarbeitung

Wächter, J. & M. Friese-Haug *)

<u>Inhaltsverzeichnis</u>	Seite
I.1 Zielsetzung und Organisationsstruktur der KTB-Datenverarbeitung	I 2
I.1.1 Einführung	I 2
I.1.2 Zielsetzung und Schwerpunkte des DV-Konzeptes	I 2
I.1.3 Organisationsstruktur der KTB-Datenverarbeitung	I 4
I.2. KTBBase (KTB database) - der Kern eines wissenschaftlich/technischen Informationssystems: Hardware-Konfiguration und Struktur der Datenbank	I 6
I.2.1 Einleitung	I 6
I.2.2 Hardware-Konfiguration	I 7
I.2.3 Aufgaben und Funktionen einer Datenbank	I 9
I.2.4 Entwicklungsphasen der KTBBase	I 10
I.2.5 Grundfunktionen der KTBBase	I 14
I.2.6 Konzeptionelles Schema der KTBBase	I 15
I.2.7 KTBBase als Kern eines wissenschaftlich/technischen Informationssystems	I 18
I.2.8 Perspektiven	I 19
I.2.9 Literatur	I 20

*) Anschrift der Verfasser: KTB Feldlabor
8486 Windischeschenbach

I.1 Zielsetzung und Organisationsstruktur der KTB-Datenverarbeitung

WÄCHTER, J., LAUTERJUNG, J. & GIESE, P.

I.1.1 Einführung

Im Jahr 1987 wird der Aufbau einer zentralen Datenverarbeitung an der Bohrlokation bei Windischeschenbach beschlossen, um die zu erwartenden großen Datenmengen zu sichern und zu archivieren. Das erste Ziel des DV-Konzeptes, das in Zusammenarbeit von Projektleitung und Schwerpunkt gelöst wird, ist die Planung und Beschaffung einer leistungsfähigen Rechneranlage.

Nach Erreichen dieses "Meilensteins" gilt es, die weitergehenden Ziele des DV-Konzeptes zu formulieren, die neben der Datensicherung eine interdisziplinäre Nutzung der Datenbasis beinhalten. In dieses Konzept müssen die Anforderungen aller beteiligten Interessens-Gruppen, der Projektleitung, des Schwerpunkts und der Arbeitsgemeinschaften integriert werden. Die Zielsetzung, das DV-Konzept und die Organisationsstruktur der Datenverarbeitung innerhalb des KTB sollen nachfolgend vorgestellt werden.

I.1.2 Zielsetzung und Schwerpunkte des DV-Konzeptes

Das Ziel des DV-Konzeptes ist der Aufbau eines Informationssystems, das sowohl geowissenschaftlichen als auch bohrtechnischen Ansprüchen genügt und die Basis für eine weitergehende, zusammenfassende Auswertung aller Daten bildet. Das wissenschaftlich/technische Informationssystem unterstützt den Benutzer aktiv darin, die vorhandene Information in ihrer gesamten Breite zu erfassen und zu nutzen.

Das DV-Konzept beinhaltet folgende Schwerpunkte:

1) Datensicherung und -archivierung

Eine Hauptaufgabe der DV-Gruppe muß die Datensicherung und -archivierung sein. Zu diesem Zweck wird eine zentrale Datenbank eingerichtet, die eine effiziente Speicherung der zu erwartenden großen Datenmengen gestattet. Die zu archivierenden Daten umfassen zum einen die an der Lokation anfallenden Daten (Feldlabor- und Bohrlochmeßdaten, bohrtechnische Daten), zum anderen auch "externe" Daten, welche an Universitäten und Forschungsinstituten an Material aus der Bohrung oder im regionalen Umfeld der Bohrung gewonnen wurden. Eine weitere wichtige Aufgabe in diesem Zusam-

Dr. J. Wächter, KTB Feldlabor, 8486 Windischeschenbach.

Dr. J. Lauterjung, Institut für Geowissenschaften und Lithosphärenforschung, Senckenbergstr. 3, 6900 Gießen.

Prof. Dr. P. Giese, Institut für Geophysikalische Wissenschaften, Rheinbadendallee 49, 1000 Berlin.

menhang ist eine zentrale Koordinierung des Datenflusses sowohl an der Bohrlokation und im Feldlabor, als auch zwischen Datenbank und den externen Benutzern.

2) Statistische Auswertung und graphische Darstellung

Für Dokumentationszwecke und eine gezielte Informationsfilterung und -aufbereitung muß das gesamte Datenmaterial statistisch und graphisch aufbereitet werden können. Wichtig ist eine Korrelation und Gegenüberstellung von Daten aus den verschiedenen Arbeitsbereichen: bohrtechnische Daten, geologisch-petrographische und tektonisch-strukturelle Informationen, geophysikalische und geochemische Daten, außerdem Bohrlochmeßdaten. Besondere Bedeutung hat unserer Meinung nach die graphische Aufbereitung der strukturellen und petrographischen Informationen. Als Beispiel für die Korrelation von Daten aus den unterschiedlichen Arbeitsbereichen seien hier genannt:

- Bohrkernorientierung,
- Erkennung lithologischer Wechsel durch die Korrelation geologisch-petrographischer, chemischer und bohrtechnischer Daten insbesondere bei kernlosen Bohrstrecken,
- Überprüfung seismischer Modelle durch tektonisch-strukturelle Daten,
- Eichung von Meßergebnissen des Wireline-Loggings durch geochemische und geophysikalische Daten.

3) Flexible Reporterstellung

Die zentrale Verfügbarkeit der Daten gestattet ein schnelles und flexibles Informationsretrieval. Datenextrakte können unter vielfältigen Gesichtspunkten in Reports zusammengestellt werden.

4) Schnittstelle für weitergehende Auswertungen

Das zentrale Informationssystem im Feldlabor soll als einheitliche Datenbasis für weitergehende Auswertungen und Modellbetrachtungen dienen. Es hat sich bereits eine Arbeitsgemeinschaft "Modellbetrachtungen" konstituiert, die auf eine Interpretation der Daten z. B. durch 3D-Modellrechnungen abzielt. Weiterhin ist in Zukunft die Bearbeitung des Datenmaterials durch Expertensysteme geplant.

Die Basis des DV-Konzeptes bildet der Aufbau einer zentralen Datenbank, in welche bereits vorhandene Datenbestände und die laufend anfallenden Daten effizient archiviert werden. Die technische Entwicklung vor allem der optischen Speicher gestattet die Aufnahme von speicherintensiven Informationen in die Datenbank (z. B. Referenz-Logs der Bohrlochgeophysik, Bild- und Dokumentarchive). Besonders die in zahlreichen externen Projekten an Universitäten oder Forschungsinstituten gewonnenen Daten, verlangen eine flexible Anpassung und Erweiterung der Datenbankstruktur, um auch diese Ergebnisse für weitergehende Auswertungen nutzbar zu machen.

Neben der Datenarchivierung und -sicherung gewinnt die statistische Auswertung und graphische Präsentation mit zunehmender Breite der Datenbasis an Bedeutung. Nur eine integrierte Bearbeitung aller Daten verspricht eine zuverlässige Interpretation der Einzelergebnisse. Auch in diesem Bereich ist eine flexible Anpassung des Konzeptes notwendig, um den wachsenden Anforderungen der Datenbankbenutzer gerecht zu werden.

Der Schlüssel zum Erfolg des DV-Konzeptes liegt also in einer flexiblen Anpassung an steigende Anforderungen. Wesentlicher Punkt bei der zielgerichteten Realisierung des DV-Konzeptes ist die Koordinierung der Entwicklungs- und Beschaffungsmaßnahmen. Zum einen müssen Parallelentwicklungen vermieden werden, zum anderen gewährleisten gemeinsame Standards und Schnittstellen einen reibungslosen Datenfluß.

I.1.3 Organisationsstruktur der KTB-Datenverarbeitung

Um diese Aufgaben erfolgreich zu lösen, gliedert sich die Organisationsstruktur der KTB-Datenverarbeitung in drei Ebenen mit jeweils unterschiedlichen Aufgabenbereichen:

- die Lenkungsgruppe,
- die DV-Arbeitsgruppe,
- die DV-Gruppe Hannover und die DV-Gruppe an der Bohrlokation.

Die Lenkungsgruppe ist die oberste Entscheidungsinstanz innerhalb der KTB-Datenverarbeitung. Sie setzt sich aus führenden Vertretern der Projektleitung (Rischmüller, Kehrer) und des Schwerpunkts (Emmermann, Harjes) zusammen (Abb. 11.1).

Die DV-Arbeitsgruppe untersteht der Lenkungsgruppe und erarbeitet für die Lenkungsgruppe Vorschläge zur Lösung aktueller DV-Problemstellungen. Die Arbeitsgruppe besteht aus Mitarbeitern der Projektleitung (Draxler, Engeser, Hoffers, Reimann) und des Schwerpunkts (Friese, Lauterjung, Wächter). Ihre Aufgabe ist die Weiterentwicklung des DV-Konzeptes, die Koordinierung der DV-Aktivitäten zwischen Projektleitung, Schwerpunkt und Arbeitsgemeinschaften und die Koordinierung der Beschaffung von Hard- und Software.

Die zentrale Datenverarbeitung an der Lokation wird von Mitarbeitern des Schwerpunkts (Friese, Wächter) geleitet. Zu ihren Aufgaben gehört die Datensicherung, das System- und Datenbankmanagement und außerdem die Abwicklung des Datentransfers von und zu den externen Projekten. Daneben sind die Software-Entwicklung, der Aufbau des Informationssystems und die Benutzerbetreuung als wesentliche Aufgaben zu nennen.

Die DV-Gruppe Hannover (Hoffers, Reimann, Operator: Sulzbacher) führt die Datensicherung der Projektleitung und das Systemmanagement an der KTB-Rechneranlage in Hannover durch. Wichtige Aufgaben sind daneben die Durchführung von Beschaffungsmaßnahmen, der Abschluß von Verträgen mit Dritten und die Antragstellung. Der

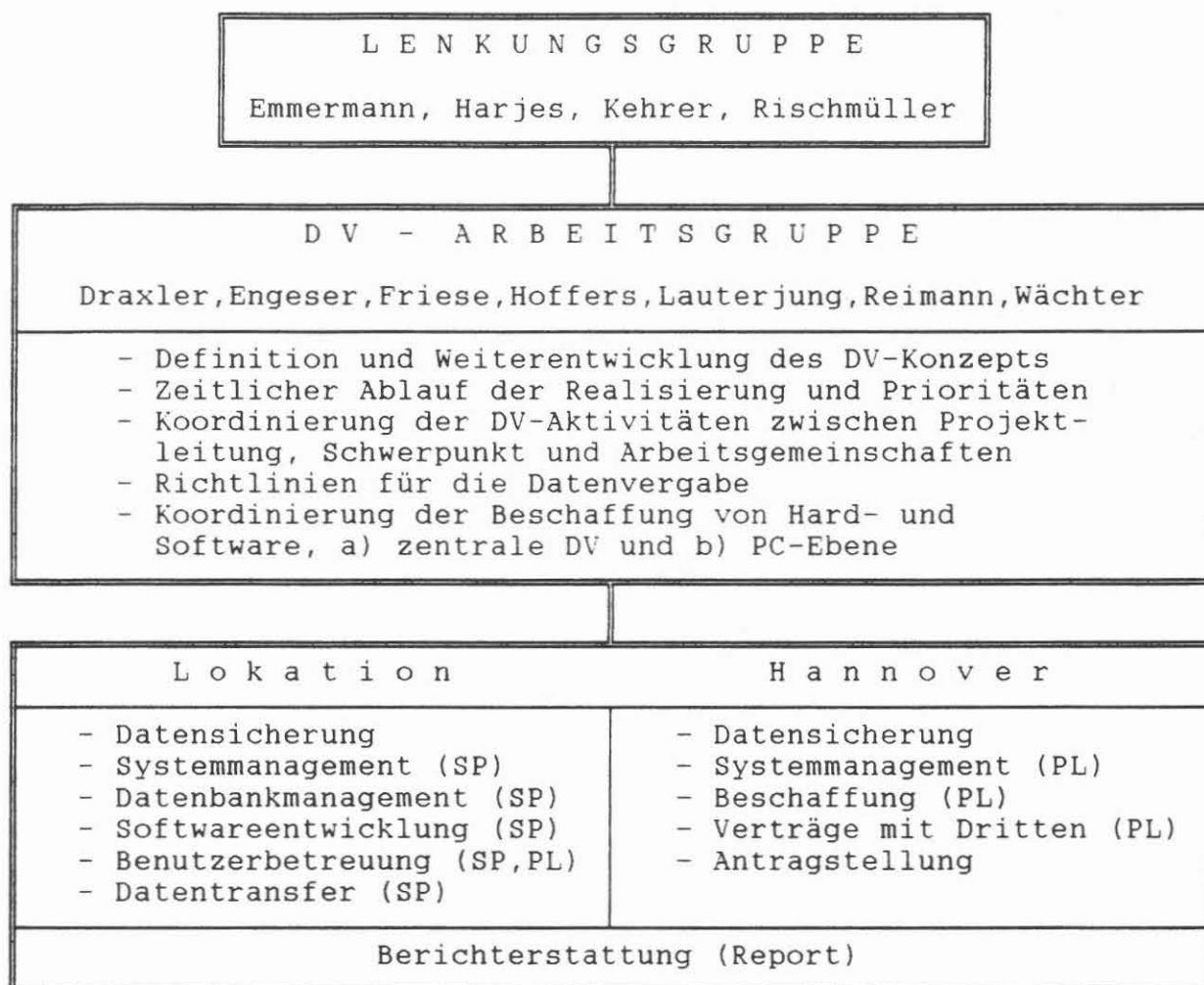


Abb. 11.1: Organisationsstruktur der KTB-Datenverarbeitung

Operator unterstützt zeitweise die DV-Gruppe an der Lokation in den Aufgaben Software-Entwicklung und Benutzerbetreuung.

Diese Organisationsstruktur bietet eine wichtige Voraussetzung dafür, daß die Anforderungen aller beteiligten Interessengruppen eingebracht und flexibel in das DV-Konzept, in Beschaffungs- oder Entwicklungsarbeiten umgesetzt werden können.

I.2. **KTBase (KTB database) - der Kern eines wissenschaftlich/
technischen Informationssystems: Hardware-Konfiguration
und Struktur der Datenbank**

Wächter, J. & M. Friese-Haug

I.2.1 Einleitung

Das KTB ist ein interdisziplinäres geowissenschaftliches Grundlagenforschungsprojekt. Die Basis zur Umsetzung der geowissenschaftlichen Zielvorstellungen bildet eine interdisziplinäre, methodisch weit gefächerte, umfassende Datengewinnung:

- Im Feldlabor erfolgt eine kontinuierliche Untersuchung von Kernmaterial, Bohrklein, Bohrspülung und Gesteinsfluiden, die in den Arbeitsgruppen Geologie/Tektonik, Petrologie/Petrographie, Geochemie und Geophysik mit einem breiten Spektrum von Methoden durchgeführt wird.

- Durch die Mudlogging-Unit werden kontinuierlich die Parameter des Spülungsumlaufes und des Spülungsschemismus gemessen.

- Ein umfangreiches, geophysikalisches Bohrlochmeßprogramm (Wireline-Logging, Experimente im Bohrloch) bildet einen wesentlichen Teil der Untertage-Datengewinnung.

- Geowissenschaftliche Umfelduntersuchungen liefern Daten für eine räumliche oder regionale Interpretation der o. g. Informationen.

- Parallellaufende und nachfolgende Forschungsprojekte sorgen durch den Rückfluß von Ergebnissen für eine ständige Erweiterung der Datenbasis.

Zur Archivierung, Dokumentation, Verwaltung und Nutzung dieser umfassenden, dynamisch wachsenden Datenbestände wird derzeit ein zentrales Rechenzentrum im Feldlabor eingerichtet. Basierend auf einer zeitgemäßen Hardware-Ausstattung wird ein zentrales Datenbanksystem aufgebaut. Die KTBase (KTB database) bildet den Kern eines wissenschaftlich/technischen Informationssystems, das eine integrierte, interaktive Nutzung aller Datenbestände gewährleisten soll.

In diesem Bericht soll der Stand der Entwicklung dargestellt werden. Der erste Teil (Kap. 2) stellt die zur Verfügung stehende Hardware des KTB-Rechenzentrums vor und gibt einen Überblick über die Struktur der Rechnervernetzung innerhalb des Feldlabors.

Der zweite Teil (Kap. 3-7) widmet sich der KTB-Datenbank KTBase. Das Kapitel 3 enthält eine kurze Einführung in eine Standard-

Dr. J. Wächter, Dipl.Min. M. Friese-Haug, KTB Feldlabor,
8486 Windischeschenbach.

Datenbankstruktur und erläutert die besonderen Leistungsmerkmale von Datenbanken. Darauf aufbauend werden im Kapitel 4 die zur Entwicklung der KtBase geplanten Schritte vorgestellt.

Derzeit steht die erste Phase der Datenbankentwicklung mit dem Entwurf des Datenbank-Schemas vor ihrem Abschluß. Vorausgegangen ist eine Informationsanalyse, in der sowohl Struktur und Definition der produzierten Daten als auch Anforderungen an die Leistungsmerkmale und die Funktionen der Datenbank ermittelt und dokumentiert wurden. In den Kapiteln 5 und 6 sollen die wichtigsten Ergebnisse, die Grundfunktionen und die Struktur der KtBase, vorgestellt werden. Das Kapitel 7 erläutert die Aufgaben der KtBase im Rahmen des projektierten wissenschaftlich/technischen Informationssystems.

Abschließend sollen im Kapitel 8 die Perspektiven für den Aufbau des wissenschaftlich/technischen Informationssystems diskutiert werden, wie sie sich aus der momentanen Situation heraus darstellen.

I.2.2 Hardware-Konfiguration

Die Computersysteme VAX 6210, MicroVAX II und GPX 3200 der Firma DIGITAL EQUIPMENT GMBH bilden die zentralen Komponenten des Rechnernetzes (Abb. I2.2) im Feldlabor. Jeder der drei Knoten hat unterschiedliche Funktionen im Netzwerk. So wird die VAX 6210 ausschließlich für Datenbankverwaltung und aufgesetzte Datenbankanwendungen eingesetzt. Die MicroVAX übernimmt Aufgaben der Bürokommunikation und Textverarbeitung, sowie zu einem späteren Zeitpunkt die Verbindung zu den Telekommunikationsdiensten der Post, Datex-P und Telefax. Die GPX 3200 ist neben ihrer Funktion als Plotterserver, zentrales System (Boot Member) für die Arbeitsplatzstationen VAXstation 2000 (Satellite Members).

Die zentralen Massenspeicher können wahlweise über zwei Rechner, die VAX6210 und die Micro VAX angesprochen werden. Dieses Verfahren gewährleistet den Datenzugriff bei Ausfall eines Systems.

Die Rechner sind auf der Basis von Ethernet zu einem lokalen VAXcluster (LAVC) verbunden. Die Vernetzung ermöglicht

- eine zentrale Verwaltung von Programmen und Daten bei gleichzeitiger dezentraler Nutzung,
- die gemeinsame Nutzung spezieller Peripheriegeräte wie Schnelldrucker, Laserdrucker, Elektrostplotter, etc. und
- eine zentrales Systemmanagement.

Die Prozeßrechner im Feldlabor, im allgemeinen Industriestandard-PC XT und AT von IBM, können ausgerüstet mit Ethernetkarte und PCSA-Software in gleicher Weise an der Netzwerkkommunikation teilnehmen. Die Funktionalität von PCSA umfaßt unter anderem Terminalemulation, Einrichtung von virtuellen Platten und Dateiübertragung von und zu VAXSystemen.

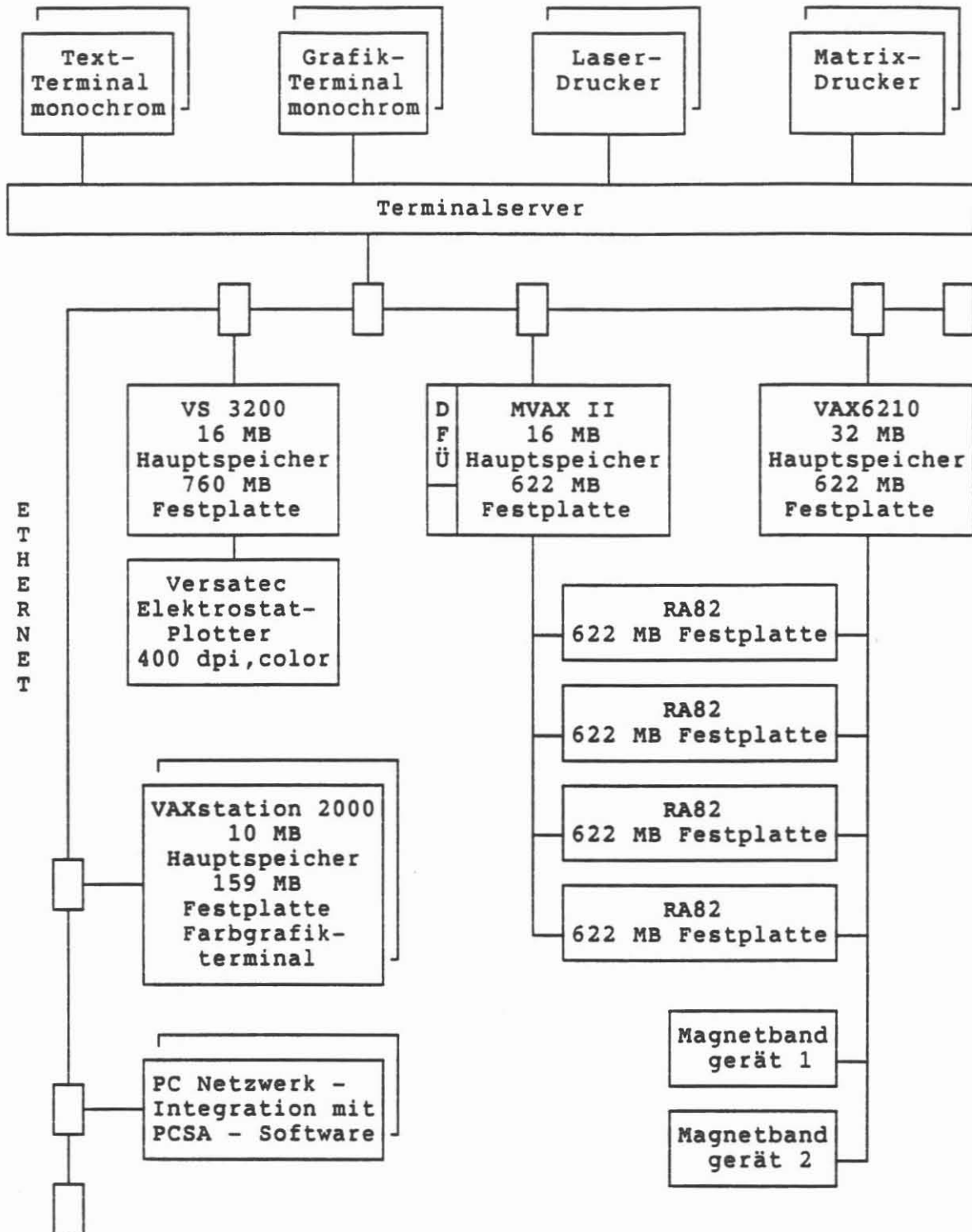


Abb. I2.2: Schematische Darstellung der Hardware-Konfiguration.

I.2.3 Aufgaben und Funktionen einer Datenbank

Mit zunehmender Leistungsfähigkeit von Computersystemen hat sich die Art der Nutzung von numerischen Berechnungen auf den Schwerpunkt Verwaltung von Informationen verlagert. Datenbanksysteme (DBS) sind Hilfsmittel zur rechnergestützten Verwaltung großer Datenbestände. Ein DBS besteht aus zwei Komponenten, der Datenbank (DB), in der zusammengehörige Informationen einer Miniwelt abgelegt sind, und der Verwaltungsssoftware (DBMS), mit deren Hilfe die Datenbasis verwaltet, aktualisiert und Anwendungen in geeigneter Form zur Verfügung gestellt werden.

Wesentliches Ziel beim Einsatz eines Datenbanksystems ist die Unabhängigkeit der Datenbasis sowohl von ihrer physischen Darstellung als auch von den externen Sichten der Datenbank. Datenunabhängigkeit beinhaltet eine flexible Anpaßbarkeit des Datenbanksystems:

- neue Anwendungen sollen keine Änderung der laufenden Anwendungen, der Datenbasis oder der Speicherstruktur erforderlich machen,
- Veränderungen in der Informationsstruktur haben auf laufende Anwendungen keine Auswirkungen.
- Veränderungen der Speicherstruktur haben außer den erwünschten Verbesserungen der Leistungsfähigkeit keine Rückwirkungen auf das Datenbank-Konzept oder die Anwendungsprogramme.

Der ANSI/SPARC Architekturvorschlag für Datenbanken hat das Ziel diese Datenunabhängigkeit in konkreten Datenbanksystemen durchzusetzen. Die vollständige Definition eines Datenbanksystems gliedert sich danach in drei Ebenen, die jeweils durch ein eigenes Schema beschrieben werden.

- externes Schema: Das Schema enthält die Festlegung von individuellen Sichten (views) der Datenbasis. Den Benutzern bzw. den Anwendungsprogrammen sind nur begrenzte Ausschnitte des konzeptionellen Schemas der Datenbank zugänglich. Der Benutzer hat den Eindruck, als ob die Datenbank nur die für ihn relevanten Informationen enthält.
- konzeptionelles Schema: Das Schema enthält den Gesamtvorrat an globalen Definitionen für die betreffende Miniwelt, deren relevante Informationen in der Datenbank erfaßt werden sollen. Die Informationsstrukturen werden mit Hilfe eines Datenmodells formuliert. Das konzeptionelle Schema bildet die Basis für die Definition des externen und internen Schemas.
- internes Schema: Das Schema bestimmt die physische Struktur der Datenbank und legt fest wie und wo die Daten gespeichert werden. Daneben können Systemparameter (z. B. Puffergrößen) und die Art und Weise des Zugriffs auf Datensätze kontrolliert werden.

Aus der zentralen Verwaltung des Datenbestandes einer Miniwelt

mit Hilfe eines DBMS und einer 3-Ebenen-Architektur der Datenbank ergeben sich wesentliche Vorteile gegenüber einer dezentralen Datenspeicherung und -verwaltung:

- Durch ein globales konzeptionelles Schema der Datenbank lassen sich unerwünschte Redundanzen in den Datenbeständen erkennen und beseitigen. Bei der Entwicklung des konzeptionellen Schemas werden mögliche negative Auswirkungen auf die Konsistenz der Datenbasis aufgedeckt, wie sie bei Update-Operationen wie Löschen, Einfügen und Verändern von Datensätzen auftreten können.

- Zugriffe auf Daten werden durch das DBMS zentral gesteuert. Aus diesem Grunde werden Inkonsistenzen vermieden, die durch unkontrollierte Updates hervorgerufen werden können. Die zentrale Steuerung garantiert außerdem eine reibungslose Multiuser-Nutzung der Datenbank und die Realisierung einheitlicher Sicherungskonzepte und Zugriffskontrollen.

- Für die einzelnen Benutzer der Datenbank kann im externen Schema die logische Sicht der Daten beliebig formuliert werden. Das Datenbanksystem kann dem Benutzer (oder dem Anwendungsprogramm) Daten in der geeigneten Form zur Verfügung stellen.

Neben den o. g. Vorteilen soll an dieser Stelle eine weitere wichtige Eigenschaft von Datenbanksystemen hingewiesen werden, die eine Wiederherstellung des Ausgangszustandes nach System-Ausfällen oder Hardware-Schäden ermöglichen (Recovery). Während des Betriebes werden vom DBMS Veränderungen der Datenbasis protokolliert. Mit Hilfe dieser redundant gespeicherten Information kann das DBMS den Datenbankzustand bis unmittelbar vor den Zeitpunkt des Ausfalls rekonstruieren.

(weitergehende Datenbankliteratur: VOSSEN 1987, LOCKEMANN & SCHMIDT 1987)

I.2.4 Entwicklungsphasen der KTBBase

Die Aufgabe der KTBBase liegt zum einen in der effektiven Archivierung der Datenbestände, zum anderen in einem flexiblen Informationsretrieval. Dieses Ziel ist unter dem Gesichtspunkt einer sich permanent erweiternden Datenbasis nur durch eine klare, detailliert dokumentierte Struktur der Datenbank zu erreichen, die flexibel erweitert werden kann.

Wie in Kap.3 gezeigt wurde, gliedert sich die Beschreibung einer Datenbank in drei Schemata, das externe, das konzeptionelle und das interne Schema, von denen das konzeptionelle Schema die Informationsstrukturen enthält, welche die interessanten Sachverhalte der zu beschreibenden Miniwelt, in diesem Fall der Miniwelt KTB, darstellen. Das externe und das interne Schema bauen auf dem konzeptionellen Schema auf.

Der erste und wichtigste Schritt des Datenbankentwurfs ist deswegen die Entwicklung des konzeptionellen Schemas. Zur Darstellung

des Schemas sind prinzipiell zwei Wege möglich, die Beschreibung des Schemas mit Hilfe des Implementierungsentwurfs, d. h. in der Datendefinitionssprache des benutzten DBMS oder eine Schema-Beschreibung mit Hilfe eines semantischen Datenmodells.

Der erste Weg mag beim Aufbau einer einfachen Anwendung mit einer statischen Struktur zulässig sein. Die Datenstrukturen und Datenbeziehungen innerhalb des KTB sind jedoch komplex und müssen, wegen des Datenrückflusses aus den parallellaufenden Forschungsprojekten, flexibel erweitert werden können. Aus diesem Grund erfolgt die Darstellung der Informationsstrukturen mit Hilfe eines semantischen Datenmodells, in diesem Fall des Entity-Relationship-Modells (ERM). Das semantische Datenmodell bildet die Grundlage für eine übersichtliche graphische Darstellung der Datenstrukturen sowohl auf der Arbeitsgruppen-Ebene als auch auf der Ebene des Gesamtschemas der Datenbank. Ein weiterer Vorteil dieser Darstellung liegt in der Unabhängigkeit des Datenbankschemas von der Datendefinitionssprache des verwendeten DBMS, wodurch eine Portierbarkeit der Datenbank auf andere Systeme bzw. DBMS erreicht wird.

Die Basis zur Entwicklung des Datenbankschemas bildet eine Informationsanalyse, in der die relevanten Sachverhalte der untersuchten Miniwelt ermittelt werden. Zur Durchführung der Analyse werden in der Literatur keine standardisierten Vorgehensweisen vorgestellt. Eine fachübergreifende wissenschaftlich/technische Nutzung setzt jedoch eine allgemeinverständliche Datendefinition und Beschreibung der Methoden voraus.

Ausgehend von diesen Anforderungen wurden zur Ermittlung der benötigten Information Interviews mit den Datenproduzenten und späteren Nutzern der Datenbank durchgeführt, die unter folgenden Gesichtspunkten standen:

- Definition der Daten. Die Definition umfaßt nicht nur eine kurze, präzise Bezeichnung der Parameter eines Datensatzes und deren Dimension, sondern auch eine klare, allgemeinverständliche Beschreibung.

- Definition der Methode. Die Interpretation der Daten setzt neben der Bedeutung der Einzelparameter auch Kenntnisse über die Methode voraus, mit welcher die Daten gewonnen wurden.

- Lagerung der Daten. Dieser Punkt ergibt sich aus dem verzögerten Aufbau einer zentralen Datenverarbeitung an der Lokation. Bislang wurden die Daten auf unterschiedlichen Datenträgern von den Arbeitsgruppen selbst verwaltet. Die Strukturierung der Datensätze erfolgte nach den speziellen Anforderungen der Arbeitsgruppen. Zum Laden der Datenbank sind deshalb z. T. komplexe Konvertierungen der Datensätze notwendig.

Neben der Definition der Daten und der Methode, besteht ein weiterer Schwerpunkt der Informationsanalyse in der abstrakten Formulierung von Grundfunktionen, welche die Datenbank unterstützen muß (siehe auch Kap. 5). Die Grundfunktionen bilden die Basis für

den Entwurf der übergeordneten, verknüpfenden Informationsstrukturen des Datenbankschemas.

Durch den Implementierungsentwurf findet die Übersetzung des konzeptionellen Schemas in das Datenmodell des verwendeten DBMS statt. Die KTBasis wird unter Verwendung der Datenbanksoftware RDB/VMS (Produkt der Firma DIGITAL Equipment GmbH) aufgebaut, die auf dem relationalen Datenmodell basiert. Die Darstellung der Daten erfolgt intern in Form von Tabellen (Relationen), deren Spalten als Attribute bezeichnet werden.

Vor der Definition der Relationen und ihrer Attribute wird eine Normalisierung der Datenstrukturen durchgeführt. Basierend auf der Normalformen-Lehre (z. B. DATE 1986) werden die Informationsstrukturen unter dem Gesichtspunkt betrachtet, mögliche Inkonsistenzen der Datenbasis aufzudecken, die bei der Veränderung, dem Löschen oder dem Einfügen neuer Datensätze z. B. durch redundante Informationen entstehen können (Update-Anomalien).

Abschließend werden semantische und referentielle Integritätsbedingungen festgelegt, die bei der Eingabe neuer Datensätze aktiv durch das DBMS überwacht werden. So kann z. B. festgelegt werden, daß die Summe von Analysenwerten einen Toleranzwert nicht über- oder unterschreitet oder daß Analysen sich auf eine real vorhandene Probe beziehen.

Nach der Entwicklung des konzeptionellen Schemas und der Ableitung des Implementierungsentwurfs besteht der nächste Schritt zum Aufbau der KTBasis im Entwurf des internen Schemas. Die Möglichkeiten zum Gestalten des internen Schemas sind vom verwendeten DBMS abhängig. RDB/VMS gestattet z. B. eine flexible Aufteilung der Datenbasis in mehrere Dateien (Multifile Datenbank), die Festlegung der physischen Speicherung und die Definition von Indizes, die einen erheblich beschleunigten Zugriff auf einzelne Datensätze ermöglichen. Um diese Möglichkeiten auszuschöpfen, sind detaillierte Kenntnisse über die Nutzung der Datenbank notwendig, die bei weitem über die o. g. Grundfunktionen der Datenbank hinausgehen. Die benötigten Informationen werden durch eine Transaktionsanalyse gewonnen, die auf genauen Angaben der einzelner Benutzergruppen über die beabsichtigte Nutzung der Datenbank basiert.

An dieser Stelle sei erwähnt, daß für den Aufbau der KTBasis das externe Schema zunächst keine Bedeutung hat. Bei einem wissenschaftlich/technischen Informationssystem hat der Benutzer Zugriff auf die gesamte Datenbasis, sodaß die Definition von Views entfällt. Das externe Schema entspricht also in diesem Fall dem konzeptionellen Schema. Die Definition von Views wird erst mit der Entwicklung von Anwendungsprogrammen wichtig, die auf die Datenbasis zugreifen, und kann zum entsprechenden Zeitpunkt erfolgen.

Nach dem Abschluß der oben genannten Schritte und der vollständigen Definition des Datenbankschemas kann die Testphase beginnen. Ihr Ziel ist erreicht, wenn die Datenbank sowohl bei Updates als

auch bei Abfragen (Retrieval) eine gutes Reaktionsverhalten zeigt. Der erfolgreiche Abschluß der Testphase setzt voraus, daß die sorgfältig geplanten Bedingungen der Testphase die Anforderungen des Routinebetriebs widerspiegeln. Die Testdaten müssen also in ihrem Umfang den Realdaten entsprechen. Zeigt die Datenbank bei bestimmten Transaktionen nicht das gewünschte Verhalten, sind Modifikationen des internen Schemas eventuell sogar Änderungen des konzeptionellen Schemas erforderlich, sodaß die oben dargestellten Entwicklungsschritte der Datenbank unter Umständen mehrfach durchlaufen werden.

Nach Abschluß der Testphase und der Entwicklung eines Datensicherheitskonzeptes, das auf den noch zu erarbeitenden Richtlinien der DV-Arbeitsgruppe basiert, kann die KTBasis für die allgemeine Nutzung freigegeben werden. Auch während des Routinebetriebes muß jedoch das Betriebsverhalten der Datenbank weiter überwacht werden. Gegebenenfalls sind Anpassungen des internen Schemas an Veränderungen des Nutzungsprofils vorzunehmen.

Die Schritte des Datenbankentwurfs und der -implementation sollen abschließend stichpunktartig zusammengefaßt werden:

I Konzeptionelles Schema der Datenbank

a) im semantisches Datenmodell (ERM)

- 1) Ermittlung der Datenstrukturen in den Fachbereichen
 - Definition der Attribute
 - Definition der Datenstrukturen
 - derzeitige Speicherung und Lagerung der Daten
- 2) Definition der Entities
- 3) graphische Darstellung der Datenstrukturen
- 4) Schema der Datenbank
- 5) Dokumentation

b) Implementierungsentwurf in RDB/VMS

- 1) Normalisierung und Beseitigen von Update-Anomalien
- 2) Definition der Attribute und Attribut-Domains
- 3) Definition der Relationen
- 4) Definition von semantischen und referentiellen Integritätsbedingungen
- 5) Dokumentation

II Entwurf des internen Schemas

- a) Transaktionsanalyse
- b) Dokumentation

III Testphase

- a) Laden der Testdaten
- b) Tuning

IV Entwicklung eines Datensicherheitskonzeptes

V Offizielle Freigabe der Datenbank und abschließende Dokumentation

Die Schritte zur Entwicklung einer leistungsfähigen Datenbank sind in Bezug auf ihren Zeitaufwand keineswegs gleichwertig. Zeitkritisch sind im wesentlichen zwei Punkte, die Entwicklung des konzeptionellen Schemas und die Testphase in ihrer Wechselwirkung mit dem physischen Design. Derzeit ist mit der Erstellung einer ersten Dokumentation die Endphase des konzeptionellen Entwurfs erreicht.

Aufbau und Struktur der KTBBase werden in differenzierter Form dokumentiert. Die zu erwartenden Erweiterungen der Datenbasis sind mit einer wachsenden Zahl von Relationen und Attributen verbunden und machen eine permanente Aktualisierung der Dokumentation erforderlich. Darüber hinaus muß die Dokumentation eine effektive Nutzung der Datenbasis unterstützen. Diese Merkmale machen eine besondere Art von Dokumentation erforderlich, die als Online-Dokumentation realisiert werden soll.

I.2.5 Grundfunktionen der KTBBase

Das konzeptionelle Schema einer Datenbank (vgl. Kap. 3) enthält den Gesamtvorrat an globalen Definitionen der abzubildenden Miniwelt und beinhaltet neben den relevanten Informationsstrukturen (Entities) auch die Definition der Beziehungen zwischen ihnen (Entity-Relations). Das Schema beschreibt also nicht nur den Datenvorrat, sondern spiegelt auch den Informationsfluß und die speziellen Informationsbedürfnisse der Miniwelt wieder.

Um den Informationsfluß innerhalb der Miniwelt KTB und die Informationsbedürfnisse der Datenbanknutzer im Schema der KTBBase abzubilden, wurden die zukünftigen Aufgaben der Datenbank in Form von Grundfunktionen beschrieben und sowohl in das Schlüsselkonzept als auch in die übergeordneten, verknüpfenden Informationsstrukturen umgesetzt.

Die Grundfunktionen lassen sich aus den allgemeinen Zielsetzungen der KTB-Datenverarbeitung ableiten, wie sie im DV-Konzept dargestellt sind, und beinhalten die Archivierung von Daten, die Dokumentation von Proben und Daten und die Nutzbarmachung von Dokumentation und Daten.

Daneben müssen Leistungen des Feldlabors für externe Projekte unterstützt werden, z. B. die Proben- und Datenvergabe, außerdem die Proben- und Datenauswahl. Auch die Überwachung des Daten- und Probenrückflusses muß von der KTBBase unterstützt werden. Dabei ist zu berücksichtigen, daß der Datenrückfluß auch in Form von Publikationen, Berichten und Reports erfolgen kann.

Die Aufgaben der Datenbank lassen sich durch folgende Grundfunktionen beschreiben:

I ARCHIVIERUNG

- a) von Informationen über Proben und Daten
 - Dokumentation der Probenart, der Rahmenbedingungen der Probenahme und der Lagerung
 - Dokumentation durchgeführter und laufender Untersuchungen
 - Dokumentation der Struktur und der Definition von Daten
- b) von Daten
 - Speicherung von internen und externen Daten und abstrakten Untersuchungsergebnissen
- c) Dokumentation und Durchführung zentraler Korrekturen

II RETRIEVAL

- a) von Informationen über Proben und Daten
 - Informationen über die Lagerung des Probenmaterials
 - Informationen über geologische Rahmenbedingungen
 - Informationen über durchgeführte und laufende Untersuchungen
 - Informationen über die Definition der gespeicherten Daten
 - Informationen über die Struktur der Daten
 - Informationen zum Verknüpfen von Daten
 - Informationen über Korrekturen
- b) von Daten
 - Retrieval von internen und externen Daten oder abstrakten Untersuchungsergebnissen

III Definition eines BEZUGSSYSTEMS

- für das Retrieval von Daten
- für das Retrieval von Proben
- für die Darstellung von Ergebnissen
- für die Interpretation von Ergebnissen

I.2.6 Konzeptionelles Schema der KTBBase

Unter Berücksichtigung der oben skizzierten Grundfunktionen zeigt das konzeptionelle Schema der Datenbank eine Gliederung in fünf übergeordnete, logische Informationsstrukturen (Abb. I2.3), die einen, hier nicht weiter dargestellten, komplexen internen Aufbau besitzen. Die Informationsstrukturen und ihre wichtigsten Funktionen sollen nachfolgend kurz vorgestellt werden:

1) Zentral-Archiv

An dieser zentralen Stelle der Datenbank werden Informationen archiviert, die innerhalb des Projektes eine besondere Bedeutung

KTBase

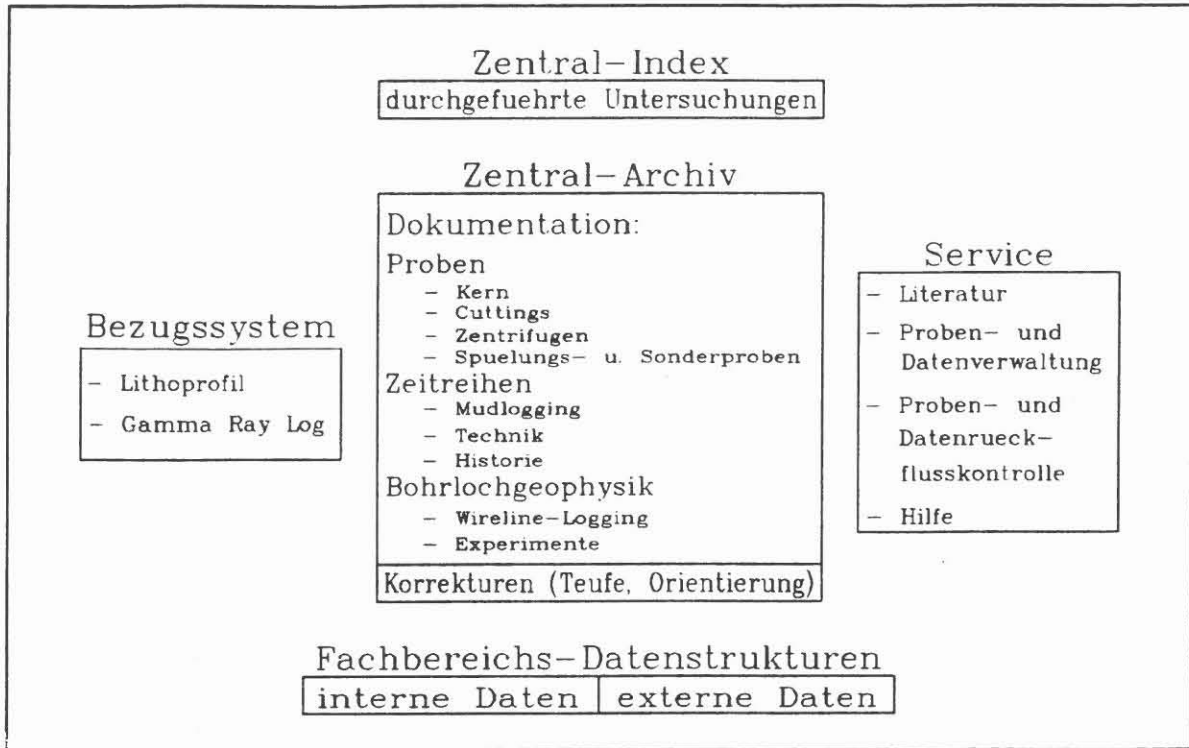


Abb. I2.3: Konzeptionelles Schema der KTBase.

besitzen. Dazu gehören zum einen die Probendokumentation und die Dokumentation der geophysikalischen Messungen im Bohrloch, zum anderen die Online-Daten, die den technischen Ablauf der Bohrung und das Mudlogging chronologisch dokumentieren. Wichtiger Teil des Zentral-Archivs sind daneben die Datenstrukturen, in denen die Teufenkorrekturen für Kern- und Spülungsproben abgelegt sind.

Das Archiv gliedert sich in die:

- Probendokumentation (Kernmaterial, Cuttings, Spülungsproben, Sonderproben)
- Zeitreihendokumentation (Mudlogging, Technik, Historie)
- Dokumentation bohrlochgeophysikalischer Messungen (Wireline-Logging, Experimente)

Neben der Archivierung liegt die Funktionalität des Zentral-Archivs in

- der Überwachung der Datenkonsistenz,
- der Durchführung von datenbankweiten Korrekturschritten,
- dem Bezug zum Referenzlog und
- der Verknüpfung von Informationen.

2) Fachbereichs-Datenstrukturen

In den Fachbereichs-Datenstrukturen werden die preprozessierten Daten der einzelnen Feldlabor-Arbeitsgruppen (interne Daten) und der Datenrückfluß externer Projekte (externe Daten) archiviert. Die Daten werden durch Untersuchungen an Probenmaterial gewonnen.

3) Zentral-Index

Im Zentral-Index werden Information darüber gehalten, welche Untersuchungen an einzelnen Proben durchgeführt wurden. Über den Index können Informationen aus dem Zentral-Archiv, dem Bezugssystem (z. B. Lithoprofil) und den Fachbereichs-Datenstrukturen miteinander verknüpft werden.

4) Bezugssystem

Das Bezugssystem ist ein Referenzlog, das eine zentrale Bedeutung für den überwiegenden Teil der Datenbankanwendungen besitzt z. B.

- die Interpretation von Daten
- die graphische Darstellung von Daten
- die Auswahl von Proben
- die Darstellung von technischen Parametern und Mudlogging-Daten
- die geotektonische Interpretation.

Das derzeit wichtigste Bezugssystem ist das Lithoprofil. Die zusätzliche Aufnahme weiterer Logs z. B. Gamma-Log erscheint sinnvoll.

5) Service

Die Service-Funktionen basieren auf vier Gruppen von Informationsstrukturen:

a) Die LITERATUR-Verwaltung unterstützt folgende Funktionen:

- die Verwaltung von abstrakten externen oder internen Daten, die z. B. als Abstract und/oder als graphische Darstellung vorliegen können,
- die Verwaltung von Fachliteratur der Arbeitsgruppen.

b) PROBEN- und DATENVERWALTUNG. Die Datenstruktur unterstützt die ordnungsgemäße Abwicklung des Proben- und Datenversands.

c) DATEN-RÜCKFLUSS-KONTROLLE. Diese Datenstruktur bildet die Schnittstelle zwischen den externen Datenproduzenten und der Datenbank. Überwacht werden z. B. die Einhaltung von Terminen und der mögliche Probenrückfluß.

d) Die HILFE-Funktion basiert auf differenzierten Informationen über die Definition von Attributen, Entities bzw. Relationen, die Struktur der Datenbank und mögliche Verknüpfungen von Informationen. Für den Datenbankadministrator bedeutet diese Funktion eine wesentliche Arbeitserleichterung bei einer Erweiterung oder Veränderung der Datenbasis. Für den Benutzer bietet sie eine wichtige Hilfe beim Navigieren und freien Arbeiten mit der Datenbank.

I.2.7 KTBBase als Kern eines wissenschaftlich/technischen Informationssystems

Die Zielsetzung des DV-Konzeptes beinhaltet zwei Schwerpunkte, erstens die Dokumentation, Archivierung und Sicherung der gewonnenen Daten und zweitens ihre wissenschaftlich/technische Nutzbarmachung. Durch den Aufbau eines wissenschaftlich/technischen Informationssystems sollen diese Ziele erreicht werden.

In kommerziellen Datenbankanwendungen sind für den dominierenden Teil der Benutzer nur selektierte Informationen der Datenbasis zugänglich. Im Gegensatz dazu steht dem autorisierten Benutzer eines wissenschaftlichen Informationssystems die Gesamtheit der gespeicherten Information zur Verfügung. Darüber hinaus muß ein wissenschaftliches Informationssystem den Benutzer aktiv darin unterstützen, die gesamte Breite der zur Verfügung stehenden Informationen zu erfassen und für seine kreative Arbeit nutzbar zu machen.

In Kap. 2 wurden die wesentlichen Funktionen eines Datenbanksystems vorgestellt. Für das Retrieval von Daten stehen auf der Basis von Abfragesprachen wie RDO oder SQL vielfältige Möglichkeiten zur Verfügung, Informationen zu verknüpfen und zu selektieren. Unter dem Gesichtspunkt der wissenschaftlichen Arbeit mit diesen Informationen können diese Funktionen jedoch nur als elementar bezeichnet werden. So würde die Abfrage bestimmter Ergebnisse der geologischen Aufnahme oder einer geophysikalischen Untersuchungsmethode eine Liste alphanumerischer bzw. numerischer Werte ergeben. Einer Interpretation der Daten kommt man also auf diesem Wege kaum näher.

Die Aufbereitung der Daten für Interpretationsprozesse ist vielmehr die Aufgabe einer Software-Shell, die sich aus der Perspektive des Informationsflusses gesehen zwischen die Datenbank und den Benutzer schaltet. Die von einem Benutzer angeforderte Information wird aus der Datenbank extrahiert und über eine Schnittstelle an die Datenbank-Shell übergeben. Dort findet eine Umsetzung der Daten z. B. in eine graphische Darstellung statt, die dem Benutzer als Ergebnis seiner Informationsanforderung präsentiert wird.

Die beiden Schwerpunkte des DV-Konzeptes werden also innerhalb des wissenschaftlich/technischen Informationssystems durch zwei unterschiedliche Komponenten realisiert:

- Die Aufgaben Dokumentation, Archivierung und Sicherung werden durch die KTBBase übernommen, die als Kern des Informationssystems die Daten- bzw. Wissensbasis verwaltet.

- Die wissenschaftlich/technische Nutzbarmachung dieser Daten ist die Aufgabe der Datenbank-Shell, die den Systemkern KTBBase umgibt. Hier sollen dem Benutzer unter einer einheitlichen Oberfläche interaktive Ein- und Ausgabefunktionen, Spreadsheet-Funktionen zur weiteren Aufbereitung von Daten, außerdem Statistik- und Grafikanwendungen zur Verfügung gestellt werden. Auch die

aktive Unterstützung des Benutzers in der Erschließung der verfügbaren Informationen ist eine Funktion der Datenbank-Shell.

I.2.8 Perspektiven

Zu einem Zeitpunkt, an dem die Vorbohrung abgeschlossen ist, stellt sich die Situation der DV an der Bohrlokation folgendermaßen dar:

- Die technische Installation und Inbetriebnahme des KTB-Rechenzentrums ist noch in vollem Gange und wird voraussichtlich im Juni abgeschlossen werden können.
- Der Aufbau der Datenbank hat derzeit mit dem Entwurf des konzeptionellen Schemas den ersten "Meilenstein" erreicht. Die eigentliche Testphase kann jedoch erst nach Abschluß der Hardware- und Software-Installation beginnen und eine offizielle Freigabe der Datenbank ist nicht vor September zu erwarten. Zu diesem Zeitpunkt steht eine Datenbank mit "elementaren" Funktionen zur Verfügung.

Nach den ursprünglichen Vorstellungen sollte die Vorbohrung zur Erprobung eines bereits implementierten Informationssystems dienen. Die gewonnenen Erfahrungen sollten dazu genutzt werden, Schwächen und Lücken bis zum Beginn der Hauptbohrung zu beseitigen, und Wissenschaftlern und Technikern zu diesem Zeitpunkt ein leistungsfähiges Arbeitswerkzeug zur Verfügung zu stellen.

Um dieses Ziel trotz der erheblichen zeitlichen Verzögerungen zu erreichen, sind konzentrierte, koordinierte Anstrengungen erforderlich. Diese müssen zwei Schwerpunkte haben:

- Die Unterstützung der Arbeitsfähigkeit der einzelnen Feldlaborgruppen durch zeitgemäße Hilfsmittel, wie z. B. CAD- und Desktop-Publishing- Software, außerdem leistungsfähige Statistik-Grafik-Tools.
- Die Realisierung des wissenschaftlich/technischen Informationssystems.

Es muß erkannt werden, daß die Implementierung und Inbetriebnahme des Datenbanksystem KTBBase zwar einen ersten und auch sehr wichtigen Schritt darstellt, daß aber noch wesentliche Entwicklungsarbeiten zum Erreichen des Gesamtziels geleistet werden müssen. Die Entwicklung einer leistungsfähigen Datenbank-Shell ist der Schwerpunkt dieser Arbeiten. Wegen der knappen Personalsituation und des engen zeitlichen Rahmens ist die DV-Gruppe auf optimale und universell einsetzbare (Programmier-) Werkzeuge, auf eine flexible, schnelle Beschaffung und auf eine gezielte Vergabe von Software-Entwicklungsaufträgen nach außen angewiesen.

Das KTB ist ein Projekt mit einer extrem breiten Basisdatengewinnung an der geowissenschaftliche und technische Disziplinen ge-

meinsam arbeiten. Die Verwaltung und die wissenschaftlich/technische Nutzung dieser umfassenden Datenbestände kann nur mit Hilfe einer leistungsfähigen Datenverarbeitung erfolgen. Das zu entwickelnde Informationssystem muß mit sehr unterschiedlichen Datenobjekten wie "normalen Daten", komplex-strukturierten Daten, Text, Graphik und Bildern umgehen und soll unter einer einheitlichen, intuitiv beherrschbaren Benutzeroberfläche Eingabe- und Retrievalfunktionen, außerdem Spreadsheet-, Statistik und Grafikanwendungen zugänglich machen. Das KTB-Projekt stellt also nicht nur für Wissenschaft und Technik, sondern auch für die Datenverarbeitung eine Herausforderung dar und kann möglicherweise Beiträge und Anregungen dazu liefern, Datenbankanwendungen stärker in die geowissenschaftliche Forschung zu integrieren.

I.2.9 Literatur

DATE, C.J. (1986): An Introduction to Database Systems, Vol. 1.- Addison Wesley Publ. Co., Reading, MA.

VOSSEN, G. (1987): Datenmodelle, Datenbanksprachen und Datenbank-Management-Systeme.- Addison Wesley Verlag, Bonn.

LOCKEMANN, P.C. & J.W. SCHMIDT (1987) [Hrsg.]: Datenbank-Handbuch.- Springer Verlag, Berlin, Heidelberg.