

„GEOL“ — eine freiformatige erdwissenschaftliche Proben-  
und Aufschlußdatei auf EDV-Basis

VON WOLFGANG SCHNABEL und LUDWIG BRÜSTLE\*)

Mit 4 Abbildungen

Inhalt

1.	Zusammenfassung — Summary — Résumé . . . . .	196
2.	Erdwissenschaftlicher Teil (W. SCHNABEL) . . . . .	197
2.1	Einleitung und Problemstellung . . . . .	197
2.2	Forderungen von seiten des Benützers . . . . .	198
2.3	Grundprinzipien des Aufbaues von GEOL . . . . .	198
2.4	Die Karteikarte als Dokument . . . . .	198
2.4.1	Erläuterungen der einzelnen Kategorien . . . . .	200
2.4.1.1	Ortsangaben . . . . .	200
2.4.1.2	Probenbezeichnungen . . . . .	201
2.4.1.3	Geologische und petrographische Angaben . . . . .	201
2.4.1.4	Angaben über die Aufsammlung . . . . .	202
2.4.1.5	Angaben über den Fundort (Fundort-Beschreibung) . . . . .	202
2.4.1.6	Angaben über die Analysen (Proben-Beschreibung) . . . . .	203
2.4.1.7	Literaturangaben . . . . .	203
2.4.1.8	Präparations- und Inventarangaben . . . . .	206
2.4.1.9	Sperrvermerk . . . . .	206
2.5	Verwaltung der Datei . . . . .	206
2.5.1	Kennzeichnung der Dokumente . . . . .	206
2.5.2	Wörterbuch (Thesaurus) . . . . .	206
2.6	Änderungsmöglichkeiten (Update) . . . . .	207
2.7	Erweiterungsmöglichkeiten . . . . .	208
2.7.1	Plotterausdrucke . . . . .	208
2.7.2	Statistische Auswertungen . . . . .	208
3.	Einsatz des IBM-Dokumentationssystems STAIRS (Storage and Information Retrieval System) (L. BRÜSTLE) . . . . .	208
3.1	Einleitung . . . . .	208
3.2	STAIRS-Konzeption . . . . .	209
3.2.1	Dialogsystem . . . . .	209
3.2.2	Maschinelle Indexierung . . . . .	209
3.2.3	Erweiterte Abfrage . . . . .	210
3.2.4	Datenbankkonzept . . . . .	210
3.2.5	Datenschutz . . . . .	211
3.3	Dokumenteingabe . . . . .	211
3.3.1	Maschinenlesbarkeit . . . . .	211
3.3.2	Maschinenerkennbarkeit . . . . .	211
3.4	STAIRS-Funktionen . . . . .	212
3.4.1	SEARCH . . . . .	212

\*) Anschrift der Verfasser: Dr. WOLFGANG SCHNABEL, Geologische Bundesanstalt, Rasumofskygasse 23, A-1031 Wien. Dr. LUDWIG BRÜSTLE, IBM-Osterreich, Obere Donaustraße Nr. 95, A-1020 Wien.

3.4.2	SELECT	212
3.4.3	BROWSE	213
3.4.4	RANK	213
3.4.5	PRINT	213
3.4.6	SORT	213
3.4.7	SAVE/EXEC	213
3.4.8	HELP	213
3.5	Dialogbeispiel	214
3.6	Programmablauf in der Datenbank GEOL	217
3.7	Nötige Maschinenausrüstung für den Betrieb von STAIRS	219
4.	Literaturhinweise	220

## 1. ZUSAMMENFASSUNG

Es wird die Organisation eines erdwissenschaftlichen Informationssystems „GEOL“ vorgestellt, die es gestattet, Probenresultate und Aufschlußbeschreibungen weitgehend unformatiert und im Volltext zu erstellen. Die zugrundegelegte Kartei erlaubt die Erfassung maschinenerkennbarer Informationen verschiedenster Probenresultate auf einem einzigen, einheitlich für alle Probenarten zutreffenden Formular.

Zum Aufbau und zur Abfrage dieser Datenbank wurde das IBM-Programmsystem STAIRS (Storage and Information Retrieval System) herangezogen. Dieses Dokumentationssystem ordnet den eingegebenen Dokumenten automatisch Suchbegriffe zu, die Abfrage der Datenbank erfolgt in Form eines Dialogs zwischen dem Benutzer und dem System.

Weitere Ausbau- und Verwendungsmöglichkeiten für die Zukunft werden angedeutet.

### Summary

In this paper the organisation of the geological Information System „GEOL“ is presented. GEOL makes it possible to handle analysis of different samples and outcrop-descriptions largely unformatted and in fulltext.

The original file enables the collection of computer readable informations concerning various, completely different samples. These informations are all included in an unique form fitting for all possible samples.

The IBM-Programm STAIRS (Storage and Information Retrieval System) was used for creating this geological database and for retrieving documents from it. This program assigns automatically descriptors to the entered document („Automatic Indexing“), the retrieval of the database is based on a dialogue between the users and the system.

### Résumé

Nous présentons l'organisation d'un système informatique géologique „GEOL“, qui permet d'établir, sous format largement libre et non abrégé, des descriptions d'analyses d'échantillons et d'affleurements. Le fichier de base permet la saisie d'informations-machine des différents résultats d'analyses à l'aide d'un formulaire unique, commun à tous les types d'analyses.

Pour l'établissement et l'accès de cette banque de données nous avons utilisé le système STAIRS (Storage and Information Retrieval System) d'IBM. Ce système de documentation associe automatiquement des cocepts-clé aux documents en entrée. L'accès à la banque de données se fait sous forme d'un dialogue entre l'utilisateur et le système. Des possibilités d'extension et d'utilisations ultérieures sont indiquées brièvement.

## 2. ERDWISSENSCHAFTLICHER TEIL

(W. SCHNABEL)

### 2.1 Einleitung und Problemstellung

Immer mehr geowissenschaftlich-geotechnisch arbeitende Institutionen sind durch die ständig wachsende Anzahl von anfallenden Daten gezwungen, elektronische Speichersysteme zu verwenden. Es existieren bereits unzählige Systeme, welche in den meisten Fällen auf ganz bestimmte, mehr oder minder eng begrenzte Problemkreise zugeschnitten sind.

An der Geologischen Bundesanstalt (GBA) fallen im Zusammenhang mit der geologischen Landesaufnahme und der Sammlung geologischer Unterlagen eine große Menge von Probandaten an, welche einerseits sehr heterogen zusammengesetzt sind und andererseits nicht nur von der GBA selbst, sondern von verschiedensten Instituten, Stellen und Einzelpersonen stammen. Es sind neben Geländebeobachtungen in erster Linie petrographische, chemische und paläontologische Gesteinsanalysen, deren einheitlicher elektronischer Erfassung sich bisher von ihrem grundsätzlich verschiedenen Aufbau und Umfang her Schwierigkeiten in den Weg gestellt haben. Besonders die vollinhaltliche Speicherung von Fossilbestimmungen sprengt jedes System, das festformatig organisiert ist und nur mit Symbolschlüsseln arbeitet.

Werden von vielen verschiedenen Stellen Daten angeliefert, ergibt sich bei einem festformatigen, mit Symbolschlüsseln arbeitenden System weiters das Problem, daß Belegunterlagen in großer Zahl nur von speziell eingeschultem Personal richtig erstellt werden können [siehe Literatur (3)].

Mit diesen Schwierigkeiten konfrontiert, wurde der Versuch unternommen, ein leicht zu handhabendes System für Analysendaten zu erstellen, bei dem die zu erfassenden Daten weitestgehend freiformatig und im Klartext der Datenbank zur Verfügung gestellt und auf Datenträger gebracht werden können.

Das hier vorgestellte System wurde mit Analysendaten aufgebaut und getestet, welche im Rahmen des Forschungsvorhabens „Wienerwaldflysch“ der erdwissenschaftlichen Institute der Universität Wien anfielen. Herr Prof. Dr. E. CLAR, der es ermöglichte, im Rahmen dieser Arbeitsgemeinschaft über den Fond zur Förderung der wissenschaftlichen Forschung die Mittel für die Programmerstellung flüssigzumachen, sei herzlich für seine

Unterstützung gedankt. Die Rechenzeit hat die IBM-Österreich kostenlos zur Verfügung gestellt, wofür ebenfalls sehr gedankt sei.

## 2.2 Forderungen von seiten des Benützers

Im vorhergegangenen Abschnitt (2.1) ist die Hauptforderung bereits angedeutet, nämlich die weitgehende Möglichkeit, die Daten möglichst ungekürzt und im Volltext eingeben zu können. Daraus ergibt sich zwangsläufig auch die Erstellung der einzuspeichernden Daten im freien Format. Eine weitere Forderung ist die nach einer möglichst einheitlichen Formularerstellung, d. h. nach wenigen Dokumentformularen. Eine Vielzahl verschiedener Vordrucke (z. B. für jede Analysenart anders organisiert) würde ja sowohl die Handhabung für den Benützer als auch die Übertragung auf einen elektronischen Datenträger erschweren.

Oft können nach dem augenblicklichen Kenntnis- und Forschungsstand gewisse Angaben nicht exakt gegeben werden. Eine wichtige Forderung an das System ist daher, keinen Zwang auszuüben, Beobachtungen genauer angeben zu müssen als sie in Wirklichkeit sind.

## 2.3 Grundprinzipien des Aufbaus von GEOL

Die Grundprinzipien der Organisation von GEOL werden diesen Forderungen gerecht. Die Datenerstellung erfolgt durch voll ausgeschriebene Angaben auf nur einem, einheitlich für die verschiedensten Probenarten gestalteten Formular als Dokument. Nach Überschreiben des Inhaltes dieser Dokumente auf einen Datenträger können sie als Handkartei verwendet werden. Symbolisiert wurde im vorliegenden Versuch nur dort, wo es im Interesse der Klarheit von fachlicher Seite her wünschenswert war. Es ist dies bei den schon aus der Fachliteratur her bekannten und allgemein gebräuchlichen Abkürzungen für Gefügedaten, Mineralien, chemischen Formeln und Elementen, Parametern und ähnlichem der Fall.

Die Angaben können, wenn sie unsicher sind, entweder ganz weggelassen oder ihrem unscharfen Inhalt nach eingegeben werden, indem Stichworte und Angaben mit Zusatzzeichen (etwa Fragezeichen), aber auch zusätzlichen Angaben besonders hervorgehoben werden. Die einheitlich für alle Proben gleich organisierte Karteikarte ist als Dokument die Informationseinheit für das System; sie ist in einzelne Kategorien (logische Untereinheiten) untergliedert, in denen die einzelnen Angaben ihrer Bedeutung nach aufgeschlüsselt werden. Diese wieder sind in einzelne Worte oder alphanumerische Zeichenketten untergliedert, die abrufbare Begriffe bilden.

## 2.4 Die Karteikarte als Dokument

Die Karte besteht aus 3 Hauptteilen, dem Kopf, der Fundort-Beschreibung (zugleich Aufschlußbeschreibung) und der Proben-Beschreibung (Abb. 1). Jeder dieser Teile ist in eine Anzahl von Kategorien (im Folgenden mit Kat. abgekürzt) gegliedert, die laufend

**GEOLOGISCHE BUNDESANSTALT WIEN - DOKUMENTATION**

1	K-Nr.	Land	Land	Proben-Codes	Verlässl.	Syst.	Länge	Breite	Höhe	von	bis	10	EDV
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22-25													
26													
27													
28													
29													
30													
31													
32													
33													
34													
35													
36													
37													
38													
39													
40													
41													

Formular 2

Abb. 1.: Die Karteikarte als Dokument von GEOL.

numeriert sind. Die zu beschreibende Einheit ist das Gestein (*Kat. 15*), das von einer bestimmten Stelle entnommen wurde. Es können mehrere, ganz verschiedene Analysen auf eine Karteikarte geschrieben werden, wenn die Angaben im Kopf für alle diese Analysen zutreffen, d. h., daß sie von der gleichen Stelle und vom gleichen Gestein stammen müssen. Reicht der Platz im Feld „Probenbeschreibung“ nicht aus, kann die Rückseite der Karteikarte beschrieben werden, wobei der Kopf für die Probenbeschreibung (*Kat. 26 bis 30*) mit einem Stempel fortlaufend neu dazugedruckt wird, wenn es sich um verschiedene Analysen handelt. Es muß nicht jede Kategorie beschrieben werden, doch erweist es sich als zweckmäßig, neben der Kartenblattnummer (*Kat. 1*) die Koordinatenangaben (*Kat. 6, 7*) und die Proben-codes auf jeden Fall anzugeben, um auf die erfahrungsgemäß häufigsten Zugriffswünsche, nämlich denen nach regionalen Gesichtspunkten und nach bestimmten Analysengruppen möglichst erschöpfend Auskunft geben zu können.

#### 2.4.1 Erläuterung der einzelnen Kategorien

Die Einzelbeispiele sind in den folgenden Abschnitten nur zur Erläuterung der Eingabeprinzipien gedacht, nicht als eine genaue Gebrauchsanweisung. Eine solche ist in Ausarbeitung und wird zum gegebenen Zeitpunkt bekanntgemacht werden.

##### 2.4.1.1 Ortsangaben

(*Kat. 1, 2, 4—11*)

*Kat. 1, Kartenblatt-Nummer:* Hier wird im Regelfall die Nummer des Kartenblattes der Österreichischen Karte 1:50.000 eingetragen (1- bis 3-stellige Zahl) oder die amtliche Numerierung der Österreichischen Karte 1:25.000 als Unterteilung der Karten 1:50.000 (z. B. 161/4, 68/2).

*Kat. 2, Land:* Hier wird im Regelfall das Autokennzeichen des Bundeslandes eingetragen (W, N, B, O, S, T, V, ST, K). Liegt eine Probe aus dem Ausland vor, wird ein A mit Schrägstrich und das Autokennzeichen des jeweiligen Staates geschrieben (z. B. A/CH = Schweiz, A/D = Deutschland).

*Kat. 4—7, Koordinatenangaben:* Im Testbeispiel wurden die Gauß-Krüger-Koordinaten (Landesvermessung) verwendet. *Kat. 6* beinhaltet die Länge, *Kat. 7* die Breite, denen in *Kat. 5* das System (in Österreich M 28, M 31 und M 34) vorangesetzt wird. Die Art der Koordinatenfindung als ein wesentlicher Faktor der Genauigkeit wird in *Kat. 4, Verlässlichkeit* eingetragen, wofür ein Abkürzungsschlüssel verwendet wurde (z. B. E = eingemessen, A 1 = abgegriffen auf einer Karte 1:10.000).

*Kat. 8, Seehöhe:* Hier kann bei Oberflächenproben die Seehöhe (Normalnull) eingetragen werden. Bei Relativangaben wird als Relationsbezug ein Buchstabe vor die Höhenangabe gesetzt (z. B. W = Wiener Null). Bei Angaben aus Bohrungen muß das Feld Seehöhe mit der Höhe des Bohr-ansatzpunktes versehen werden.

*Kat. 9, 10, Tiefe von — bis:* Hier wird der Tiefenbereich angegeben, falls eine Probe aus einer Bohrung oder Grabung stammt.

*Kat. 11, Geographische Einheit:* Hier kann die Landschaftsbezeichnung im Volltext eingegeben werden (z. B. Wienerwald, Ötztaler Alpen, Inntal). Zur Erzielung einer Einheitlichkeit ist ein Landschaftsverzeichnis für Österreich mit allen hier verwendbaren Begriffen ausgearbeitet worden.

#### 2.4.1.2 Probenbezeichnungen

(*Kat. 3*)

Diese erfolgen in *Kat. 3, Proben-Codes*. Es wurde ein Verzeichnis von Abkürzungen für die Probenarten entworfen. Ein einzelner Buchstabe bezeichnet das Vorhandensein nicht weiter untersuchter Rohproben (z. B. H = Handstück, S = Schlammprobe), zwei Buchstaben bezeichnen die Analysenart, wobei der erste den Fachbereich bezeichnet (z. B. M = Mikropaläontologie, P = Petrographie), der zweite die Probenart (z. B. N = Nannofossiluntersuchung, D = Dünnschliff).

Beispiele: MN = Nannofossiluntersuchung  
 PD = Petrographischer Dünnschliff  
 PC = Chemische Analyse eines Gesteins  
 HC = Chemische Analyse von Wässern

In Sonderfällen wurden weitere Untergliederungen vorgenommen, wobei am Ende des zweistelligen Codes ein Bindestrich gesetzt und ein weiterer Buchstabe hinzugesetzt wurde (z. B. PC-R = Röntgenanalyse, PC-M = aus einer chemischen Analyse errechneter Modalbestand). Werden an einer Rohprobe mehrere Untersuchungen vorgenommen, werden die Codes in *Kat. 3* fortlaufend durch Zwischenräume getrennt geschrieben.

#### 2.4.1.3 Geologische und petrographische Angaben

(*Kat. 12—15*)

*Kat. 12, Tektonische Einheit:* Hier wird die tektonische Einheit volltextlich angegeben, in der Regel die Decken- oder Teildeckeneinheit oder Analoges (z. B. Hallstätter Zone, Greifensteiner Decke, Moravikum).

*Kat. 13, Stratigraphische Einheit:* Hier wird der Schichtenname (engl. Formationsbegriff) im Volltext angegeben, im Regelfall also der mit einem Lokalnamen versehene stratigraphisch-lithologische Begriff (z. B. Greifensteiner Schichten, Schöcklkalk, Innsbrucker Quarzphyllit). Bei nicht mit Lokalnamen benannten Einheiten muß die analoge Bezeichnung gewählt werden. Diese schließt meist das Alter ein (z. B. Gaultflysch, Neokom [neokome Schiefer der Kalkalpen], Mittelkreideschiefer), oder die petrographische Bezeichnung (Glimmerschiefergruppe, Zentralgneis).

*Kat. 14, Stratigraphische Untereinheit:* Unterabteilung der stratigraphischen Einheit (engl. Member-Begriff) bei möglicher weiterer Unterteilung (z. B. Obere Serie, Basisquarzit).

*Kat. 15, Petrographie:* Bezeichnung bzw. Kurzbeschreibung des Gesteins im Volltext,

- z. B. Glimmerschiefer
- Mergel, rot
- Sandstein, gradiert, fossilführend
- Granit, stark verwittert
- Sand, fluviatil, grau, kreuzgeschichtet

Es können beliebig viele Termini eingegeben werden, wobei aus Gründen der Klarheit die angegebene Reihenfolge: *Gestein* — *Genese* — *Farbe* — *Strukturen* — *Diverses* eingehalten werden soll.

#### 2.4.1.4 Angaben über die Aufsammlung

(*Kat. 16—18*)

*Kat. 16, Kollektor:* Der Name des Probenaufsammlers, eventuell Bezeichnung einer Arbeitsgruppe.

*Kat. 17, Zweck, Anlaß:* (z. B. Landesaufnahme GBA, U-Bahnbau Wien, Forschungsauftrag).

*Kat. 18, Datum:* Im Regelfall wird die Jahreszahl genügen, je ein Feld für Monat und Tag ist vorgesehen (wichtig z. B. bei manchen Wasseranalysen).

#### 2.4.1.5 Angaben über den Fundort (Fundort-Beschreibung)

(*Kat. 20—25*)

*Kat. 20, Probennummer des Kollektors:* Hier kann die individuell gehandhabte Proben- oder Aufschlußnumerierung eingetragen werden (als Folge von Buchstaben, Nummern, Sonderzeichen ohne Zwischenraum, z. B. 187, 367/1973, F635L, WIENERWALD/388).

*Kat. 21, Lagerung, Gefüge etc.:* Hier werden die Gefügemessungen des Aufschlusses eingetragen. Es erweist sich als zweckmäßig, die Messung der Fallrichtung und des Fallwinkels nach der Methode von CLAR (1954) anzugeben (z. B. 240/35, 075/20), die für die programmgesteuerte Auswertung der Daten besonders günstig ist. Den Meßdaten vorangesetzt wird das Symbol für die jeweilige Gefügeart, wie es allgemein gehandhabt wird (z. B. SS = Sedimentäre Schichtung). Ein für die Zukunft bindender Schlüssel wird ausgearbeitet. Stammen die Proben nicht aus dem Anstehenden, kann dies durch Wörter wie „verrutscht“, „Rollstück“ etc. angegeben werden.

*Kat. 22—25, Zusätzliche Angaben:* Hier ist Raum für Angaben gelassen, die entweder nicht elektronisch verarbeitet werden (z. B. eine Aufschlußskizze), oder es können nähere Angaben über den Fundort, soweit sie über die im Kopf gegebenen Ortsangaben hinausgehen, gemacht werden (z. B. „Probe aus Brunnaushub 60 m SW des Hofes Sieghartsberg“), oder der Aufschluß näher beschrieben werden (z. B. verfallener Steinbruch,

Prallhang). Zusätzlich freigehaltene Kategorien erlauben den Ausbau dieses auf die Probendokumentation zugeschnittenen Dokuments auf eine Aufschlußkartei.

#### 2.4.1.6 Angaben über die Analysen (Proben-Beschreibung)

(Kat. 26—36)

*Kat. 26, Code:* Wiederholung des in Kat. 3 (Kapitel 2.4.1.2) gegebenen Symbols, auf das die Analyse zutrifft.

*Kat. 27, Methode:* Kurzbezeichnung der Analysenmethode (z. B. Lichtmikroskop, Mikrosonde).

*Kat. 28, Jahr:* Jahresangabe der Analysenerstellung.

*Kat. 29, Analysennummer:* Laufende Analysennummer des Labors.

*Kat. 30, Bearbeiter:* Der Bearbeiter der wissenschaftlichen Bestimmung.

*Kat. 31—35, Probenergebnis:* Dieses Feld ist gänzlich frei organisierbar. Als *Kat. 32* wird das Analyseergebnis in der allgemein üblichen Form übersichtlich eingetragen. Logisch miteinander verknüpfte Teilangaben werden durch einen Beistrich zusammengefaßt,

z. B. bei chemischen Analysen: CaO 36 ‰,  
MgO 9 ‰,  
FeO Spuren,  
bei Fossilbestimmungen: Globotruncana stuarti,  
Neocomites neocomiensis D'ORBIGNY,  
Nannoconus sp. umgelagert,

und so fort. Dabei ist zu beachten, daß innerhalb der in dieser Weise logisch verknüpften Teilangaben kein Beistrich gesetzt werden darf sondern nur an deren Ende.

Für weitere allgemeine Zusatzangaben zum Probenergebnis sind die *Kat. 32—36* vorgesehen. *Kat. 32 und 33* beinhaltet die *Altersangaben*, der aus dem Probenergebnis zu folgernde Altersbereich (z. B. Pannon, Lias, Campan — Maastricht). Unter *Kat. 32* werden Einzelbegriffe, z. B. Lias, Campan, Coniac) eingespeist, unter *Kat. 33* Zweitbegriffe bei Reichweiten (für das obige Beispiel z. B. Maastricht). *Kat. 34* ist für Zusatzangaben zur *Probengüte* vorgesehen. Für solche Güteangaben ist ein Symbolschlüssel in Ausarbeitung. *Kat. 35* ist für *allgemeine Zusatzangaben* reserviert (z. B. „Näheres siehe Literatur“, „graphische Darstellung im Bericht“ usw.). *Kat. 36* ist für einen weiteren Ausbau reserviert.

Über die Möglichkeit, mehrere Probenergebnisse auf einem Dokument festzuhalten, wurde schon im Kapitel (2.4) gesprochen.

#### 2.4.1.7 Literaturangaben

*Kat. 37)*

In *Kat. 37* können eine oder mehrere Literaturzitate oder Berichte angegeben werden, in denen die betreffende Probe veröffentlicht ist oder zu deren gedanklichen Untermauerung sie beigetragen hat. Die Angabe kann

GEOLOGISCHE BUNDESANSTALT WIEN - DOKUMENTATION

57	Land	199	Wien-Codes	199	199	21820	37010	Höhe	von	bis	EDV		
Geographische Einheit		Kahlenberger Decke											
Lithologischer Schichtname		Mergel hellgrau											
Stratigraphische Unterinheit													
Kollektor		Landesaufnahme GBA											
Schubel		1971 M T V											
FUNDORT - BESCHREIBUNG						PROBEN - BESCHREIBUNG							
57977 - Nr des Kollektors						MN		LICHTMIKROSKOP		1971		Analysen-Nr	
58°150/60						SERADNER						EDV	
Zusätzliche Angaben (ev Skizze):						Probenergebnis							
						<p>Arkangelskiella cymbiformis,                  Prediscosphaera sp.,                  Micula staurophora,                  Cretarhabdus crenulatus,                  Watznaueria barnese,                  Cribrosphaerella numerosa,                  Marcalius inversus,                  Cretarhabdus conicus,                  Cylindralithus gallicus,                  Tetralithus gothicus fa trifida.</p> <p><u>ALTER:</u> Maastricht</p>							
LITERATURANGABEN						PRÄPARATIONS- u. INVENTARANGABEN							
Verh. Geol.B.-A, 1972, A73 FF.						Code		Labor-Nr.		Präparator		Kiste A (Reste)	

Formular 2

Abb. 2 a: Eine mit den Ergebnissen einer Nannoplanktonanalyse ausgefüllte Karteilkarte.

C00000053  
 57  
 NIEDERÖSTERREICH  
 ANALYSEN: MN  
 NANNOFOSSILIEN  
 GAUSS-KRUEGER-KOORDINATEN: L=21820, B=37010 (GENAUIGKEIT: A 1)  
 GEOGRAPHISCHE EINHEIT: WIENERWALD  
 TEKTONISCHE EINHEIT: KAHLENBERGER DECKE  
 STRATIGRAPHISCHE EINHEIT: ALTLENGBACHER SCHICHTEN  
 LITHOLOGIE: MERGEL, HELLGRAU.  
 AUFTRAGGEBER: SCHNABEL  
 ANLASS: LANDESAUFNAHME GBA  
 1971  
 PROBENUMMER DES BEARBEITERS: 57/97  
 AUFSCHLUSSDATEN:  
 AUFRECHTE LAGERUNG 150/60  
 PROBE BEFINDET SICH: PROBENRESTE KISTE A  
 LITERATUR: VERH. GEOL.B.-A. 1972, A73 FF.  
 ANALYSEN: MN  
 NANNOFOSSILIEN  
 ANALYTIKER: STRADNER (1971)  
 METHODE: LICHTMIKROSKOP  
 PROBENERGEBNIS:  
 ARKANGELSKIELLA CYMBIFORMIS,  
 PREDISCOOPHAERA SP.,  
 MICULA STAUROPHORA,  
 CRETARHABDUS CRENULATUS,  
 WATZNAUERIA BARNESE,  
 CRIBROSPHAERELLA NUMEROSA,  
 MARCALIUS INVERSUS,  
 CRETARHABDUS CONICUS,  
 CYLINDRALITHUS GALLICUS,  
 TETRALITHUS GOTICUS FA TRIFIDA.  
 ALTER: MAASTRICHT

Abb. 2 b: Der vollständige Ausdruck dieser Angaben. In gleicher Weise können diese Angaben auf einem Bildschirm angezeigt werden.

## GEOLOGISCHE BUNDESANSTALT WIEN - DOKUMENTATION

40	MD	PS	PD	PM	GK	H	Blatt	15068	88806	Höhe	von	bis	153	
Wienerwald Einheit							Nordzone							
Wolfganginger Schichten							Stratigraphische Untereinheit							
Kalksandstein [Genese, Farbe, Strukturen, (sedimentär, tektonisch), Diverses]														
Arge Wienerwald Uni Wien							Forschungsauftrag			1969 M T				
FUNDORT - BESCHREIBUNG							PROBEN - BESCHREIBUNG							
5100/H18; Kollektors							PS: 0,4-0,063			Jahr		Analysen-Nr.		
Lagerung, Gefüge ect							Bearbeiter			EDV				
Zusätzliche Angaben (ev. Skizze)							Probenergebnis							
							%SM 0.15, OP 54, BI 6, CL 1, GL 8, DM 31, GR 6, ZI 26, TU 58, RU 8, AP 1, CR X, ST X.							
LITERATURANGABEN							PRÄPARATIONS- u. INVENTARANGABEN							
Jahrb. Geol. B.-A. 1972, S. 103-186.							Code Labor-Nr. Präparator Kista D (H) (Nr. i)							

Formular 2

GK	Sehnenschnitt	1970	180
Schnabel			

Q1 0.09,  
MD 0.17,  
Q3 0.27,  
SO 1.72,  
SK 0.80.

ERGÄNZENDE ANGABEN:

Weitere Parameter und Kurve Lit. S 112

FM	MD-Integr. Zlg.	1970	
Schnabel			

KAL-BM 25,  
QU 58,  
KF 1,  
PL X,  
SM X,  
GB X,  
KA 8,  
FK 2,  
TS 2,  
GL 3,  
BG X,  
DIV X.

Abb. 3: Vorder- und Rückseite einer mit petrographischen Daten ausgefüllten Karteikarte. Sie beinhaltet die Detailergebnisse einer Schwermineralanalyse, einer Korngrößenanalyse und einer Integrationszählung eines Kalksandsteins.

entweder im Volltext erfolgen, oder — bei Vorhandensein einer Titelaufnahme — durch eine laufende Nummer. Mit dieser Kategorie kann vom Programm her in Zukunft auch eine Verknüpfung mit einer allfälligen Literaturkartei hergestellt werden.

#### 2.4.1.8 Präparations- und Inventarangaben

(Kat. 38—41)

*Kat. 38:* Angabe des Probencodes entsprechend Kat. 26.

*Kat. 39:* Laufende Labor-Nummer.

*Kat. 40:* Name des Präparators.

*Kat. 41:* Inventarisierungsangaben des Präparates oder des Probenmaterials, eventuell Inventarisierungsnummer.

#### 2.4.1.9 Sperrvermerk

(Kat. 19)

Durch Angaben von Kennzahlen in *Kat. 19*, *Vertraulichkeit*, kann entweder der Inhalt des gesamten Dokuments oder einzelner Teile daraus gesperrt werden. Die Angabe der Vertraulichkeit in dieser Kategorie ergänzt die im STAIRS-Standard vom Programm her vorgesehenen Maßnahmen des Datenschutzes (siehe auch Kap. 3.2.5).

## 2.5 Verwaltung der Datei

Das vorliegende System ist bisher als ein Demonstrationsmodell erarbeitet worden, das sich als brauchbar erwiesen hat und derzeit ausgebaut wird. Für die zukünftige Verwaltung der Datei ist ein Rahmen festgelegt worden, innerhalb dessen der Betrieb und der weitere Ausbau vollzogen werden wird.

### 2.5.1 Kennzeichnung der Dokumente

Sowohl im Kopf- als auch im Analysenteil des Kartenvordruckes ist je ein Feld „EDV“ zur laufenden Numerierung sowohl der Dokumente als auch der Analysen belassen, die bei der Einspeisung mit laufenden Nummern versehen werden.

### 2.5.2 Wörterbuch (Thesaurus)

Die im Abschnitt (2.4) näher beschriebene Karteikarte stellt ein Dokument dar, in dem ein mehr oder minder ungegliedert vorliegender Bericht in Einzelbegriffe zerlegt und diese dem ihnen zukommenden Überbegriff (= Kategorie) zugeteilt wurde. Es sind also Deskriptoren angefallen, die eine weitere Inhaltsanalyse des Dokumentes erübrigen und als selbstständige Suchbegriffe Verwendung finden. Um dem Benutzer der Datei, sei es bei der Erstellung von neuen Daten als auch beim Auffinden von bereits eingespeicherten Angaben zu unterstützen, wird derzeit ein Thesaurus erarbeitet, der die logischen Suchbegriffe für jede einzelne Kate-

gorie übersichtlich aufzeigen wird. In einer ersten Ausbaustufe werden nur die in der Datei bereits vorkommenden Begriffe angeführt sein, in einer weiteren ist geplant, alle möglichen Begriffe für die einzelnen Kategorien für das gesamte Bundesgebiet in den Thesaurus aufzunehmen. Es werden also die manuelle und maschinelle Indexierung ineinandergreifen (siehe auch Kapitel 3.1).

Der Thesaurus wird sinngemäß durch die Ortsangaben in den Kategorien 1 (Kartenblatt Nr.) und 2 (Land) ergänzt sowie durch die in Kat. 15 aufscheinenden petrographischen Begriffe. Entsprechend der Suchlogik, wie sie STAIRS zugrundliegt, ist jede durch Zwischenräume (blanks) getrennte alphanumerische Zeichenfolge ein möglicher Suchbegriff. Zur Aufnahme in den Thesaurus werden sich sinngemäße Stichworte bei der Fundortbeschreibung in den Kategorien 22—25 ergeben, wo Begriffe wie „Steinbruch“, „Brunnenbegrabung“ ect. als Deskriptoren in Frage kommen, besonders nach einer Erweiterung der Probenkartei auf eine Aufschlußkartei. Analog dazu werden die in der Probenbeschreibung in symbolisierter oder abgekürzter Form aufscheinenden Minerale, chemischen Elemente, Parameter ect. in den Thesaurus aufgenommen werden.

## 2.6 Änderungsmöglichkeiten (Update)

Eine Berichtigung bzw. Ergänzung von bereits auf Datenträger gespeicherten Karteikarten kann durch programmierte Verfahren an jeder gewünschten Stelle vorgenommen werden. Auf dieselbe Weise können generelle Änderungen überholter Begriffe, wie sie sich z. B. bei den Kategorien der tektonischen und stratigraphischen Einheiten entsprechend dem wissenschaftlichen Fortschritt ergeben werden, am gesamten gespeicherten Datenbestand vorgenommen werden. Von solchen Korrekturmöglichkeiten sind allerdings die als Handkartei weiterverwendeten Dokumentoriginale ausgenommen, bei denen solche Änderungen händisch bewerkstelligt werden müßten, was bei einem umfangreichen Datenbestand praktisch undurchführbar ist; bei Berücksichtigung der Tatsache, daß es sich dabei um ein Original des betreffenden Bearbeiters handelt, braucht dies aber auch gar angestrebt zu werden, hier sei nur auf die Möglichkeit solcher genereller Änderungen im gespeicherten Datenbestand hingewiesen.

Wenn auch die Organisation dieses Systems die Verwendung von Ausdrücken im Volltext ermöglicht, so wird es sich im Laufe der Zeit doch ergeben, daß Abkürzungen oder Symbolisierungen bestimmter Begriffe Vorteile ergeben. Abkürzungsschlüssel für bestimmte Begriffsgruppen (bei petrographischen Beschreibungen etwa die Farben, bei geologischen Angaben die stratigraphischen und tektonischen Einheiten) können unter Zuhilfenahme von bereits gespeicherten Begriffen allmählich erarbeitet werden. Es ist geplant, solche Schlüssel entsprechend bekanntzumachen, worauf diese schon auf den Karteikarten selbst verwendet werden können, wobei prinzipiell die Möglichkeit offen bleiben wird, weiter den Volltext zu

gebrauchen. Eine entsprechende Vereinheitlichung erfolgt vom Programm her. Ein Schlüssel ist z. B. für die Bundesrepublik Deutschland erstellt worden [siehe Literatur (3)].

Prinzipiell besteht bei allen Begriffen, die sich mit anderen ganz oder teilweise decken (z. B. Hornblende — Amphibol, Sandstein — Arkose) die Möglichkeit, auf eine solche Identität vom Programm her durch Synonymbeziehungen hinzuweisen. Beim vorliegenden Modell sind solche programminterne Synonymbeziehungen bei den relativen Altersangaben hergestellt worden (siehe auch Kap. 3.6b).

## 2.7 Erweiterungsmöglichkeiten

Hier sollen Ausgabemöglichkeiten, die nicht in den Routinen von STAIRS vorgesehen sind (siehe Kap. 3.4), aufgezeigt werden. Solche bestehen in vieler Hinsicht und sollen am Beispiel zweier programmgesteuerter Ausgabemöglichkeiten hier nur auszugsweise gestreift werden; sie sind z. T. von der Zugänglichkeit zu den entsprechenden Peripheriegeräten abhängig.

An einer Verwendung dieses freiformatigen, hier an einer Proben- und Oberflächenauflösungsbeschreibung angewandten Systems für Bohrlochbeschreibungen wird gearbeitet, da es aber für diese Zwecke zur Abänderung der hier zugrundegelegten Karteikarte (Abb. 1) kommen muß, soll hier nicht mehr davon die Rede sein.

### 2.7.1 Plotterausdrucke

Solche sind, gestützt auf die Angaben der Koordinaten (Kat. 6 und 7) nach Erarbeitung von Plotter-Programmen in vielfältiger Weise durch Symbole (Punkte, Fallzeichen ect.) und/oder Beschriftungen möglich. Prinzipiell kann jede Teilangabe der Kartei in Plotter-Ausdrucken dargestellt werden. Auf diese Weise können Punkt-Karten des Karteiinhaltes hergestellt werden, die eine wertvolle Ergänzung zur Literatur, zur geologischen Landesaufnahme und zur Tiefenerschließung bestehender geologischer Karten darstellen.

### 2.7.2 Statistische Auswertungen

Die Möglichkeit von statistischen Auswertungen ist naturgemäß bei entsprechend umfangreichem Datenbestand in vielfältiger Weise möglich.

## 3. EINSATZ DES IBM-DOKUMENTATIONSSYSTEMS STAIRS (STORAGE AND INFORMATION RETRIEVAL SYSTEM)

(L. BRÜSTLE)

### 3.1 Einleitung

Zur Erstellung der erdwissenschaftlichen Datenbank „GEOL“ wurde das IBM-Programmprodukt STAIRS (Storage and Information Retrieval System) herangezogen. STAIRS ist ein Programmsystem, das in den großen

Bereich der Informationssysteme, speziell der nichtnumerischen Informationssysteme gehört. Die wesentlichste Forderung an ein nichtnumerisches Informationssystem ist es, die im Text der Dokumente enthaltenen Informationen so aufzubereiten und wegzuspeichern, daß sie als Antwort auf Anfragen an dieses Informationssystem wieder zur Verfügung gestellt werden können.

Die inhaltliche Erschließung von Dokumenten kann auf verschiedene Arten realisiert werden: In der traditionellen Dokumentation dienen als Beschreibungselemente im wesentlichen die Notationen eines Klassifikationssystems oder die Deskriptoren eines Thesaurus. Da diese Beschreibungselemente intellektuell gewonnen, und damit von außen zugeordnet werden, bezeichnet man sie auch als externe Beschreibungselemente (manuelle oder intellektuelle Indexierung).

Der Inhalt kann aber auch mit Wortfragmenten oder Stichwörtern aufbereitet werden, d. h. mit Bestandteilen, die im „Dokumenttext“ enthalten sind. Diese internen Beschreibungselemente spielen vor allem in der maschinellen Dokumentation eine Rolle (maschinelle Indexierung).

### **3.2 STAIRS-Konzeption**

#### **3.2.1 Dialogsystem**

Die Abfrage von STAIRS-Datenbanken basiert auf einem Dialog zwischen Fragesteller und Datenverarbeitungssystem von einer Datenstation aus. Als Datenstationen dienen Bildschirmgeräte mit Tastatur, Schreibmaschinenterminals usw. Ein Dialogsystem gibt dem Benutzer im Allgemeinen Gelegenheit, sich auf zweierlei Weise über die Qualität seiner Fragestellung zu orientieren: Einmal zeigt es, durch eine Zahlenangabe (Statistik) an, wieviele Dokumente sich auf Grund seiner Frageformulierung qualifiziert haben, z. B. indem es sie auf einem Bildschirm aufzeigt. So kann der Benutzer prüfen, inwieweit die gefundenen Dokumente die gewünschte Information enthalten (Relevanzkontrolle).

Entspricht die Zahl der Dokumente bzw. deren Relevanz nicht den Vorstellungen des Benutzers, so kann er unmittelbar auf die bereits formulierten Teile seiner Anfrage Bezug nehmen und diese Anfrage entweder durch eine ODER-Verknüpfung durch zusätzliche Stichwörter erweitern oder z. B. durch eine UND- oder UND NICHT-Verknüpfung einengen. Diese Kommunikation zwischen Benutzer und Dokumentationssystem über eine Datenstation stellt in jedem Fall sicher, daß der Benutzer ein zufriedenstellendes Ergebnis erhält.

#### **3.2.2 Maschinelle Indexierung**

Zur maschinellen Analyse und Indexierung sprachlich vollständiger Dokumententexte (z. B. Befunde, Produktbeschreibungen, Zeitschriftenartikel, Verträge, Gesetze, Gerichtsentscheidungen, Patente, Beschreibung von Vorträgen oder in unserem Fall Proben- und Aufschlußbeschreibungen) ver-

wendet STAIRS die sogenannte Stichwortmethode. Diese Methode — auch maschinelle Indexierung genannt — geht davon aus, daß sich in dem zur Darstellung eines Sachverhaltes verwendeten Text so aussagekräftige Wörter oder Wortgruppen finden, daß sie als Beschreibungselemente und damit auch als Suchelemente unmittelbar verwendet werden können. Ein alphabetisches Wörterbuch dieser maschinell gewonnenen Stichwörter muß ferner ein ausreichendes Ordnungsmittel darstellen. Eine Reihe von Wörtern (die sogenannten Stopwörter) besitzen so geringe Aussagekraft, daß ihre Aufnahme in das Wörterbuch nicht sinnvoll erscheint, wie z. B. Artikel, Präpositionen, Konjunktionsworte, aber auch bestimmte Stichwörter, die in einzelnen Fachgebieten so oft Verwendung finden, daß sie für das betreffende Gebiet keine Aussagekraft mehr besitzen (im vorliegenden Fall z. B. „Schichten“). Diese Stopwörter können vom Benutzer — abhängig vom jeweiligen Dokumentationsgebiet — definiert werden und finden im Zuge des „Maschinellen Indexing“ keine Aufnahme in das Kontextwörterbuch.

### 3.2.3 Erweiterte Abfrage

Ziel einer Recherche ist es, aus den in der Datenbank gespeicherten Informationen diejenigen herauszufinden, auf die eine bestimmte Anfrage zutrifft. Dazu ist es erforderlich, die zu klärende Frage so zu formulieren, daß sie vom maschinellen Dokumentationssystem interpretiert und „verstanden“ werden kann. Voraussetzung hierfür ist, daß sich der Fragesteller an bestimmte Regeln hält.

STAIRS besitzt eine Abfragesprache mit entsprechenden Sprachelementen, mit denen zur Qualifizierung der relevanten Dokumente bei der „Maschinellen Indexierung“ erkannte Textzusammenhang die Unterscheidung zwischen formatierten und nichtformatierten Daten und die Gliederung der Dokumente in einzelne Abschnitte (Kategorien) genutzt werden.

Die Operatoren drücken aus, welche Bedingungen an die einzelnen Stichwörter geknüpft sind, damit sich ein der Abfrage entsprechendes Dokument qualifiziert.

Neben den logischen Operatoren (UND, ODER, UND NICHT, ENTWEDER ODER) bietet STAIRS die Möglichkeit, mit Hilfe der sogenannten „Kontext-Operatoren“ strukturelle (Grammatik) und bedeutungsmäßige (Kontext) Abhängigkeiten abzufragen (näheres in Punkt 3.4.1. „SEARCH“). Eine Suche in formatierten Dokumentabschnitten (z. B. Datumsangaben) wird durch die Vergleichsoperatoren — z. B.  $>$ ,  $<$ ,  $=$ , und einige andere gewährleistet (Näheres in Punkt 3.4.2 „SELECT“).

### 3.2.4 Datenbankkonzept

Mit STAIRS können bis zu 16, jeweils unter einem Namen ansprechbare Datenbanken geführt werden. Das erweist sich vor allem dann als sehr nützlich, wenn Dokumente sehr unterschiedlicher Struktur (z. B. Gesetze, Presseartikel, Patente, Probenergebnisse) zu verarbeiten sind, wenn

Dokumente zeitlich getrennt werden sollen oder Dokumente heterogener Sachgebiete zu speichern sind und sich Anfragen nie auf alle diese Gebiete zugleich beziehen. Infolgedessen sind Anfragen sowohl von mehreren Datenstationen an eine einzelne Datenbank als auch von jeder einzelnen Datenstation an mehrere Datenbanken möglich.

### **3.2.5 Datenschutz**

Aus dem Datenbankkonzept ergibt sich zwangsläufig die Notwendigkeit, Inhalte von Datenbanken gegen unberechtigten Zugriff zu schützen, den Zugriff zur gesamten Datenbank also zu gestatten oder zu verhindern. Auch muß es möglich sein, wenn einzelne Daten nur ein einziges Mal gespeichert, aber in verschiedener Weise verarbeitet werden sollen, auch einzelne Teile einer Information vor unbefugtem Zugriff zu schützen. GEOL erfüllt diese Forderungen dadurch, daß es den Zugriffsschutz in Form eines Identifikationsschlüssels des Benutzers über den Schutz der Datenbank hinaus auf einzelne Dokumente, Abschnitte (Kategorien) und formatierte Felder ausdehnt (siehe Kap. 2.4.1.9).

## **3.3 Dokumenteingabe**

Zwischen den Eingabedaten eines Dokumentationssystems und ihrer Verarbeitung besteht eine unmittelbare Abhängigkeit. Einmal müssen die zu verarbeitenden Daten von der Maschine übernommen werden, d. h. maschinenlesbar sein. Zum anderen müssen die Daten bei der Erfassung dem Verarbeitungsziel entsprechend strukturiert werden, d. h. sie müssen maschinenerkennbar sein.

### **3.3.1 Maschinenlesbarkeit**

STAIRS setzt kein spezielles Datenerfassungsmedium voraus. Als Datenträger kommen Lochkarte, Lochstreifen, Magnetbänder und optische Belege in Frage. Weiters können Daten „On line“ über Bildschirmgeräte erfaßt werden. Bei der Maschinenlesbarkeit von Daten kommt es darauf an, daß deren Code bzw. typographische Form von der Maschine akzeptiert wird; sie ist also im wesentlichen hardwareabhängig.

Die Daten zum Aufbau der erdwissenschaftlichen Datenbank „GEOL“ wurden mittels Lochkarten erfaßt.

### **3.2.2 Maschinenerkennbarkeit**

Die Maschinenerkennbarkeit hat den Zweck, einzelne Segmente (Kategorien) eines Dokumentes durch ein Verarbeitungsprogramm identifizieren zu können; sie ist also softwarebedingt (näheres über die Kategorisierung der erdwissenschaftlichen Dokumente in Punkt 2.4).

Die Strukturierung der Dokumente wird erreicht durch Verwendung von Kategorienzeichen, aus denen das Verarbeitungsprogramm erkennt, ob der Inhalt einer Kategorie z. B. ein Autor, ein Titel oder ein Proben-

ergebnis ist. Dabei ist es ohne Bedeutung, ob dieses Kennzeichen aus alphabetischen, numerischen oder alphanumerischen Zeichen besteht. Wichtig ist allein, daß ein einmal gewähltes Kategorienkennzeichen, das z. B. die darauf folgende Zeichenkette als Autor definiert, immer dann gewählt wird, wenn bei einem Dokument ein Autor zu erfassen ist.

Der Vorteil dieser Dokumentstrukturierung liegt vor allem darin, daß der Benutzer sich bei der Abfrage auf spezielle Kategorien beziehen kann.

### 3.4 STAIRS-Funktionen

Die Recherche verläuft im Wesentlichen in zwei Schritten. Der erste Schritt läßt sich als Suchen nach relevanten Informationen bezeichnen, d. h. als Ermittlung, welche Dokumente auf Grund einer bestimmten Frageformulierung überhaupt relevant sind. Er umfaßt die beiden Funktionen SEARCH und SELECT. Innerhalb dieser Funktionen findet der eigentliche Dialog zwischen dem Benutzer und dem System statt.

Der zweite Schritt in einer STAIRS-Recherche besteht darin, daß die Dokumente, die sich während des ersten Schrittes als relevant qualifiziert haben, auf dem Bildschirm ausgegeben werden. Die Dokumentausgabe wird durch den BROWSE-Befehl bewerkstelligt.

#### 3.4.1 „SEARCH“

SEARCH verarbeitet Fragen, die aus den bei der maschinellen Inhaltsanalyse der Dokumente gewonnenen Elementen (Stichwörtern) und logischen bzw. Kontextoperatoren bestehen. Neben den logischen Operatoren ODER und UND verfügt die SEARCH-Funktion mit dem Operator NOT auch über die Möglichkeit der Negation in bezug auf eine Ausgangsmenge von Dokumentnummern. Mit XOR können ENTWEDER-ODER-Verbindungen hergestellt werden. Mit den Kontextoperatoren SAME, WITH und ADJ können UND-Sachverhalte weiter eingengt werden. SAME verlangt, daß durch diesen Operator verknüpfte Stichwörter im selben Abschnitt (Kategorie) eines Dokumentes vorkommen müssen (KOLLEKTOR SAME MAYER bedeutet, daß sich nur die Dokumente qualifizieren, bei denen das Stichwort MAYER in der Kategorie „Kollektor“ enthalten ist). Der Operator WITH verlangt, daß die verknüpften Stichwörter im selben grammatikalischen Satz vorhanden sein müssen. Bei ADJ (adjacent) müssen die Stichwörter unmittelbar aufeinander folgen, im gegenständlichen Fall z. B. Gattungs- und Artnamen eines Fossils (GLOBOTRUNCANA ADJ STUARTI).

Die Anzahl der durch Operatoren verknüpfbaren Stichwörter ist innerhalb eines Dialogs praktisch unbegrenzt. Dabei können zum Aufbau einer komplexen Fragestruktur auch Klammerungen verwendet werden.

#### 3.4.2 „SELECT“

Während sich die SEARCH-Funktion auf die inhaltlichen Angaben der Dokumente bezieht, wählt SELECT die sich qualifizierenden Dokumente

über die formalen Angaben aus, die ja Inhalt formatierter Felder sind. Dieses Auswählen geschieht ähnlich wie bei SEARCH mit Hilfe einer Reihe von Operatoren, wobei allerdings neben die logischen Operatoren der BOOLE'schen Algebra und Kontextoperatoren die Vergleichsoperatoren GLEICH, GRÖßER GLEICH, KLEINER GLEICH, GRÖßER, KLEINER, NICHT GLEICH und INNERHALB DER GRENZEN VON ... treten.

DATUM WL 1969, 1973, bewirkt z. B., daß sich alle Dokumente qualifizieren, die eine Datumsangabe zwischen 1969 und 1973 aufweisen (WL = within limits).

#### 3.4.3 „BROWSE“

Die BROWSE-Funktion erlaubt neben der Anzeige des gesamten Dokuments die Beschränkung auf einzelne Abschnitte („Browsing Kategorien“), das Hin- und Herblättern im einzelnen Dokument und von Dokument zu Dokument (näheres siehe in Punkt 3,5: Dialogbeispiel).

#### 3.4.4 „RANK“

Diese Funktion sortiert die sich qualifizierenden Dokumente nach ihrem Relevanzgrad. Die Relevanz eines Dokumentes wird im wesentlichen durch die Häufigkeiten bestimmt, mit denen die Stichwörter der Anfrage in der Antwortmenge vorkommen.

#### 3.4.5 „PRINT“

Durch die Druckfunktion PRINT kann die Ausgabe des Textes der relevanten Dokumente über den Schnelldrucker veranlaßt werden und zwar genau nach Anzahl und Umfang.

#### 3.4.6 „SORT“

Im Gegensatz zu RANK sortiert SORT die auszugebenden Dokumente nach dem Inhalt eines vom Fragesteller zu spezifizierenden formatierten Feldes in auf- oder absteigender Reihenfolge.

#### 3.4.7 „SAVE/EXEC“

Mit SAVE ist es möglich, eine einmal formulierte Anfrage unter einem vierstelligen Namen (Kennwort) zu speichern. Sie braucht bei Bedarf nicht erneut formuliert, sondern lediglich mit EXEC unter diesem Namen wieder abgerufen zu werden. Die SAVE-Funktion unterstützt im wesentlichen die selektive Informationsverteilung.

#### 3.4.8 „HELP“

Die HELP-Funktion ermöglicht es dem noch ungeübten STAIRS-Besitzer, sich während des Dialoges über die genaue Funktionsweise der ein-

zelen STAIRS-Befehle zu informieren. Bei einer fehlerhaften Befehls-eingabe stoppt STAIRS den laufenden Dialog und informiert den Benutzer über die Fehlerart und über den Weg, wie dieser Fehler zu beheben ist.

### 3.5 Dialogbeispiel

Annahme: Im Raum Preßbaum werden geologische Aufnahmen durchgeführt, der Bearbeiter sucht vorhandene Belege über die Kahlenberger Decke. Sein besonderes Interesse gilt dabei den leitenden Nannofossilien, u. a. der Form *Tetralithus goticus* DEFLANDRE forma *trifida* STRADNER.

Im folgenden Dialogbeispiel (es ist mit Absicht sehr speziell gehalten), das sich zwischen dem Benutzer und dem System via Bildschirm abspielt, sind die Bildschirmaus-schnitte umrandet. Die vom Benutzer eingegebenen Befehle sind mit Großbuchstaben dar-gestellt, die vom System automatisch vorgegebenen Zeichen und die Antworten sind kursiv geschrieben.

Nachdem mit dem System Verbindung aufgenommen wurde, werden mit der ersten Frage Angaben über die „Kahlenberger Decke“ gesucht. Dies geschieht durch Angabe der Funktion SEARCH, worauf vom System mit 00001 die erste Fragezeile freigegeben wird, in der der Suchbegriff KAHLENBERGER ADJ DECKE angegeben wird (die Worte „Kahlen-berger“ und „Decke“ stellen ja ein Wortpaar dar, sie müssen in den Do-kumenten hintereinander stehen. Siehe auch Kap. 3.4.1).

SEARCH

00001 KAHLENBERGER ADJ DECKE

<i>Kahlenberger</i>	<i>750 Treffer</i>	<i>482 Dokumente</i>
<i>Decke</i>	<i>5628 Treffer</i>	<i>4217 Dokumente</i>
<i>Resultat</i>	<i>251 Treffer</i>	<i>227 Dokumente</i>

Es wurde mitgeteilt, daß das Stichwort „Kahlenberger“ 750 mal in 428 Dokumenten vorkommt, das Stichwort „Decke“ 4628 mal in 4217 Dokumenten und der gesuchte Begriff „Kahlenberger Decke“ 251 mal in 227 Dokumenten.

Dem Bearbeiter sind dies zu viele Unterlagen und er schränkt daher seine Frage weiter ein, indem er sich nur auf das Kartenblatt 57 bezieht, auf dem sein Untersuchungsgebiet liegt. Unter dem Befehl SELECT schränkt er also seine erste Frage ein, vom System wird damit die 2. Frage-zeile freigegeben, in der der neue Suchbegriff 1 KARTENBLATT EQ 57 angegeben wird (Einschränkung der Fragezeile 1 auf Dokumente, die sich auf das Kartenblatt (EQ=) 57 beziehen).

SELECT

00002 1 KARTENBLATT EQ 57

Resultat

114 Dokumente

Es wurde mitgeteilt, daß 114 Dokumente über die Kahlenberger Decke auf Blatt 57 vorhanden sind. Auch diese Anzahl ist dem Bearbeiter zu hoch und er schränkt die zweite Frage weiter ein, indem er das engere Interessengebiet durch Koordinaten weiter einengt. Dieses liegt zwischen den Längenkoordinaten —22000 und —20000 und innerhalb des Breitenkoordinatenstreifens 37000 und 38000 (WL = within limits). (Siehe auch Fußnote <sup>1</sup>).

SELECT

00003 2 LAENGE WL —22000, —20000 AND  
BREITE WL 37000, 38000

Resultat

21 Dokumente

In 21 Dokumenten findet der Bearbeiter also für ihn zutreffende Unterlagen.

Sein weiteres Interesse gilt der Nannofossilform *Tetralithus goticus* DEF. forma *trifida* STRADNER, von der unser Bearbeiter weiß, daß diese Form z. T. als Unterart *trifidus*, z. T. als forma *trifida*, also uneinheitlich, in der Literatur geführt ist. Diesem Umstand trägt er bei der Abfrage dadurch Rechnung, daß er nach Gattung und Art durch TETRALITHUS ADJ GOTICUS fragt und durch den Operator SAME, gefolgt von TRIFID \$ („Mask“-Funktion, siehe Fußnote <sup>1</sup>), zu erkennen gibt, daß er mit dem Wortpaar „Tetralithus goticus“ den Wortstamm „trifid“ sucht.

SEARCH

00004 (TETRALITHUS ADJ GOTICUS) SAME TRIFID \$

<i>Tetralithus</i>	228 Treffer	218 Dokumente
<i>Goticus</i>	89 Treffer	89 Dokumente
<i>Trifida</i>	21 Treffer	21 Dokumente
<i>Trifidus</i>	5 Treffer	5 Dokumente
Resultat	26 Treffer	26 Dokumente

<sup>1</sup>) Die MASK-Funktion, als Dollarzeichen (\$) dargestellt, erlaubt die Suche nach Wortstämmen. Es wird der sichere Wortstamm angegeben und mit \$ abgeschlossen, wodurch alle Begriffe gesucht werden, die mit dem angegebenen Wortstamm beginnen. Sie werden alphabetisch geordnet am Bildschirm aufgezeigt.

Ebenso können auch Zahlenfolgen gekürzt abgefragt werden. Auf Grund der MASK-Funktion kann die Frage in Fragezeile 00003 (s. o.) auch folgendermaßen gestellt werden:  
2 LAENGE WL —21 \$, —20 \$ AND BREITE 37 \$

Das gesuchte Fossil findet sich in 26 Dokumenten (5 mal als Unterart „trifidus“, 21 mal als „trifida“).

Nun möchte unser Bearbeiter noch wissen, ob diese Form aus der Kahlenberger Decke seines engeren Arbeitsgebietes beschrieben ist und kombiniert seine ursprüngliche Frage nach der Kahlenberger Decke in seinem Gebiet (Fragezeile 00003) nach der nach dem gerade gesuchten Fossil (Fragezeile 00004).

00005 3 AND 4

*Resultat*

*7 Treffer*

*7 Dokumente*

Diese 7 qualifizierten Dokumente möchte der Bearbeiter auf dem Bildschirm durchblättern, und zwar nach dem Aufsammlungsdatum aufsteigend (A = ascending) sortiert.

Nach Eingabe des Befehls SORT wird die Zeile 00006 aufgezeigt und der Befehl 5 DATUM A eingegeben, worauf die Vollzugsmeldung „Dokumente sortiert“ erfolgt. Daraufhin wird der BROWSE-Befehl gegeben (siehe auch Kap. 3.4.3), worauf einzelne Kategorien und Kategoriengruppen angeführt werden. Der Bearbeiter kann auswählen, ob er die vollständigen Dokumente sehen will (was mit dem Operator ALL veranlaßt werden würde) oder nur einzelne Teile derselben. Unser Bearbeiter will nur die Koordinaten, die stratigraphischen Einheiten, das Datum, den Analytiker, das Probenresultat und das Alter der Proben wissen. Er vermerkt die entsprechenden Kennzahlen auf dem Bildschirm.

SORT

00006 5 DATUM A

*Dokumente sortiert*

BROWSE

<i>1 Kartenblatt-Nr.</i>	<i>2 Land</i>	<i>3 Analysen</i>
<i>4 Koordinaten</i>	<i>5 Geogr. Einheit</i>	<i>6 Tekton. Einheit</i>
<i>7 Stratigr. Einheit</i>	<i>8 Lithologie</i>	<i>9 Auftraggeber</i>
<i>10 Anlaß</i>	<i>11 Datum</i>	<i>12 Aufschlußdaten</i>
<i>13 Analytiker</i>	<i>14 Probenergebnis</i>	<i>15 Alter</i>

4 7 11 13 14 15

Die gewünschten Teile der 7 Dokumente erscheinen nun, nach Datum geordnet, auf Abruf auf dem Bildschirm. Auf der folgenden Bildschirmdarstellung ist der 5. Dokumententeil gezeigt.

### 5. Dokument

#### 1. Seite

Gauss-Krüger-Koordinaten:

$L = -21820, B = 37010$  (Genauigkeit: A 1)

Stratigraphische Einheit: Altenglbacher Schichten  
1971

Analytiker: Stradner (1971)

Probenergebnis:

*Arkangelskiella cymbiformis*,  
*Prediscosphaera* sp.,  
*Micula staurophora*,  
*Cretarhabdus crenulatus*,  
*Watznaueria barnese*,  
*Cribrosphaerella numerosa*,  
*Marcalius inversus*,  
*Cretarhabdus conicus*,  
*Cylindralithus gallicus*,  
*Tetralithus goticus* *fa trifida*.

Alter: Maastricht

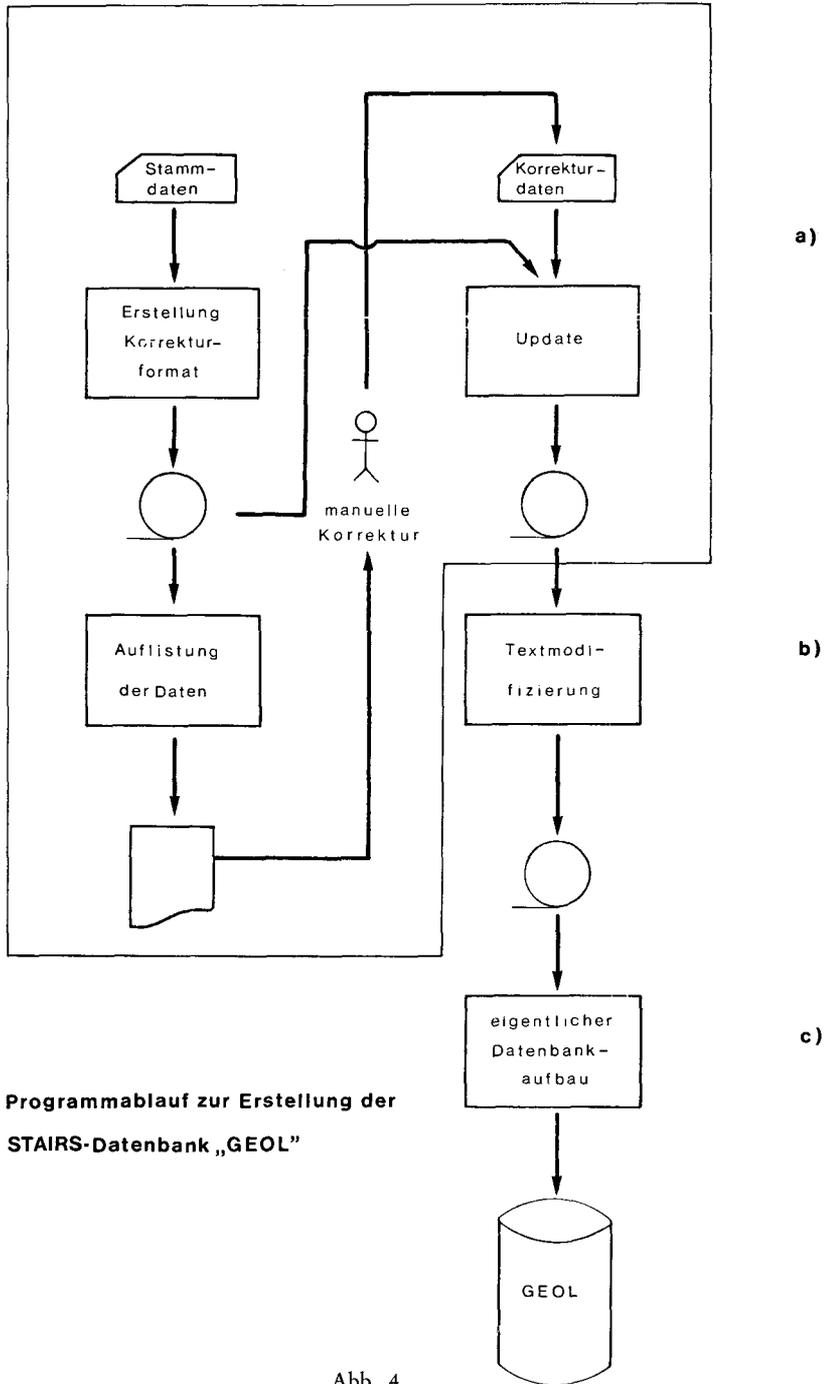
Diese Probe interessiert besonders, da die campane Leitform *Tetralithus goticus* forma *trifida* hier in einer Maastrichtprobe vorkommt. Dieses Dokument soll in seiner Gesamtheit (also nicht nur die auf Grund des BROWSE-Befehles gezeigten Dokumentteile) am Schnelldrucker ausgedruckt werden, was mit dem Befehl PRINT ALL veranlaßt wird. Daraufhin erfolgt der Ausdruck des Gesamtdokumentes (es ist das Dokument, das in Abb. 2, Seite 204 abgedruckt ist).

### 3.6 Programmablauf in der Datenbank GEOL

Der Aufbau einer STAIRS-Datenbank vollzieht sich in drei Stufen (Abb. 3):

- a) Korrektur der Daten
- b) Automatische Dokumentänderung
- c) Eigentlicher Datenbankaufbau

*ad a)* Die auf Lochkarten oder Band erfaßten Daten werden auf Grund von vorgegebenen Algorithmen auf formale Fehler geprüft. Das Korrekturprogramm stellt insbesondere sicher, daß keine unzulässigen Kategorien bzw. Kategoriengruppen in Dokumenten enthalten sind und daß zulässige Kategorien formal richtig ausgefüllt sind. Fehlerhafte Dokumentteile werden mit Angabe der Zeilennummer ausgedruckt. Die Formalfehler werden von Sachbearbeitern korrigiert. Die daraus resultierenden Streichungen, Änderungen und Einfügungen werden auf Lochkarten erfaßt und wieder mit den Stammdatenbestand zusammengespielt (Update-Lauf).



Programmablauf zur Erstellung der  
STAIRS-Datenbank „GEOL“

Abb. 4.

*ad b)* Die nun fehlerfreien Dokumente werden durch einen weiteren Programmablauf überarbeitet. Durch diese Programme werden die ursprünglich erfaßten Abkürzungen in den Volltext übersetzt (z. B. MN = Nannofossilien). Sowohl die Abkürzungscodes als auch die automatisch eingefügten voll ausgeschriebenen Worte bilden mögliche Suchbegriffe. Weiters werden die Dokumente durch automatische Generierung von zusätzlichen Stichwörtern erweitert. Mittels einer im System eingespeicherten stratigraphischen Tabelle werden z. B. die in den Kategorien 32 und 33 enthaltenen Altersangaben mit zusätzlichen Suchbegriffen versehen. Hat der Sachbearbeiter einem Probenergebnis z. B. den Suchbegriff „Sarmat“ zugeordnet, werden die Stichwörter „Miozän“, „Neogen“ und „Tertiär“ generiert (siehe auch Kap. 2.6).

*ad c)* Nach dieser applikationsbedingten Dokumentaufbereitung beginnt der eigentliche Datenbankaufbau über die STAIRS-Standardprogramme. Jede STAIRS-Datenbank besteht aus 4 physisch getrennten Dateien; dem *Wörterbuch*, der *invertierten Datei*, der *Textindexdatei* und der *Textdatei*. Das *Wörterbuch* enthält in alphabetischer Reihenfolge alle im Wege der maschinellen Indexierung gewonnenen Suchbegriffe sowie eventuelle Synonymbezeichnungen. Die *invertierte Datei* enthält für jedes Stichwort die Nummern der Dokumente, zu deren Beschreibung das Stichwort herangezogen wurde. Die Formatangaben aus den definierten formatierten Feldern sind in der *Textindexdatei* gespeichert. Das vollständige Dokument ist in der *Textdatei* enthalten. Die Grundüberlegung hierfür ist, als Antwort auf eine spätere Suche nach Informationen nicht allein einen Quellenhinweis nachgewiesen zu bekommen, sondern die ursprüngliche Information selbst, um diese z. B. von einem Bildschirm aus auf ihre Relevanz unmittelbar überprüfen zu können.

## 2.7 Nötige Maschinenausstattung für den Betrieb von STAIRS

Die Minimalkonfiguration einer beliebigen Installation ist abhängig von

1. den Fähigkeiten des vom Benutzer eingesetzten Betriebssystems.
2. den Speicheranforderungen des Datenverarbeitungs-Monitors, der die On-line Übertragungen steuert und
3. der Anzahl und dem Umfang der Datenbanken, die in das System aufgenommen werden sollen.

Zum Datenbankaufbau benötigt man einen Hauptspeicherbereich von maximal 128 K (K = Kilobytes = 1024 Bytes = 1024 Hauptspeicherstellen). Für die Datenbankpflege (Mischen, Update von Datenbanken) sind maximal 130 K nötig. Das On-line Dialogsystem von STAIRS (AQUARIUS) benötigt einschließlich des Datenfernverarbeitungssystems CICS (Customer Information Control System) bei ca. 15 Bildschirmgeräten etwa 180 K.

Die Minimalausrüstung erfordert eine Zentraleinheit IBM 360-40 (156 K) oder ein System IBM 370-135 (240 K), einen Plattenspeicher, zwei 4-Spur-Bandeinheiten, Kartenleser, Drucker und Abfrageterminals (z. B. Bildschirme IBM 3270).

#### 4. Literaturhinweise

CLAR, E., 1954: Ein zweikreisiger Geologen- und Bergmannskompaß zur Messung von Flächen und Linearen. — Verh. Geol. BA. 1954/4, 201—215, Wien 1954.

LANG, F. H., und BOCK, F., 1973: Wiener Beiträge zur elektronischen Erschließung der Information im Recht. — IBM Österreich, Veröffentlichung, Wien 1973.

LOOK, E. R., und VINKEN, R., 1971: Elektronische Datenverarbeitung bei der Aufnahme und Herstellung von geologischen Karten (mit einem Beitrag von S. BESSAU). — Herausgegeben von der Bundesanstalt für Bodenforschung und den Geologischen Landesämtern der Bundesrepublik Deutschland, Hannover 1971.

STAIRS-Handbuch, 1973: Teil A: Funktionen und Möglichkeiten. — IBM-Form E 12-1189-0. Teil B: Von der Datenerfassung zum betriebsbereiten System. — IBM-Form E 12-1190-0. Teil C: Datenbankpflege. — IBM-Form E 12-1191-0.

Bei der Schriftleitung eingelangt am 12. März 1974.