

# Das CORC-Projekt von OCLC an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen

Monika Cremer, Heike Neuroth

## 1. OCLC und CORC

Der große amerikanische Bibliotheksverbund OCLC<sup>1</sup> (*Online Computer Library Center, Inc.*) hatte bereits vor einigen Jahren ein Projekt zur Katalogisierung von Internet-Dokumenten initiiert: INTERCAT, an dem sich viele amerikanische Bibliotheken beteiligt haben. Inzwischen ist die technische Entwicklung weitergegangen. OCLC startete im Januar 1999 das Projekt CORC<sup>2</sup> (*Co-operative Online Resource Catalog*) zur Erfassung von Internet-Ressourcen. Hier wurden bewusst schon in der Entwicklungsphase Bibliotheken beteiligt, und zwar auch auf internationaler Ebene.

Mit dem CORC-Projekt wird der Versuch unternommen, eine umfassende Datenbank zu Online-Ressourcen in internationaler Kooperation mit Bibliotheken aufzubauen. Dabei wird untersucht, welche Standards eingesetzt werden können und welche technischen Methoden und Instrumente benötigt werden, um Online-Ressourcen zukünftig in weitgehend automatisierter Form zu erfassen. Zur Zeit werden Online-Ressourcen auf verschiedene Arten katalogisiert. Einerseits werden sie mit Hilfe klassischer Regelwerke wie AACR2 z.B. unter Benützung von Formaten wie MARC erschlossen. Andererseits existieren schon diverse Ansätze, Online-Ressourcen mit Hilfe von Metadaten zu erfassen (z.B. im SonderSammelGebiets-FachInformationsführer-Projekt SSG-FI3).

Bei dem CORC-Projekt werden Katalogisate unterschiedlicher Herkunft und mittels verschiedener Regelwerke in einer Datenbank zusammengeführt. Die Datensätze können entweder in MARC 21 oder auch im Dublin Core Format eingegeben, angezeigt und für die lokale Datenbank heruntergeladen werden. Die Erstellung der bibliographischen und deskriptiven Angaben zu einer Online-Ressource kann über die Eingabe der URL erfolgen, aus der das CORC-System automatisch anhand der auf der HTML-Seite aufgeführten Informationen (Metadaten etc.) eine Basis-Titelaufnahme generiert. Die Vervollständi-

---

1 OCLC Online Computer Library Center, Inc. Home Page

<<http://www.oclc.org/oclc/menu/home1.htm>>

2 CORC Home page <<http://www.oclc.org/oclc/corc/index.htm>>

3 SSG-FI Homepage <<http://www.sub.uni-goettingen.de/ssgfi/>>

gung bzw. Korrektur dieser Basis-Titelaufnahme kann manuell wahlweise in der Dublin Core- oder MARC 21-Ansicht erfolgen. Das CORC-System bietet bei der teilautomatischen Erfassung von Online-Ressourcen auch die Generierung von Schlagworten, DDC etc. an. OCLC hält die gespiegelten Library of Congress Authority Records als Datenbank vor, so dass Personen- und Körperschafts-Normdaten kontrolliert werden können. Insgesamt besteht CORC aus vier verschiedene Datenbanken:

- Resource Record Database (*Record Creation*)
- Authority Database
- CORC *Pathfinder* Database
- Dewey Decimal Classification Database

Die *Niedersächsische Staats- und Universitätsbibliothek Göttingen* (SUB)<sup>4</sup> ist seit April 1999 Projektpartner. Seit dem 1. September 1999 fördert die DFG für ein Jahr die Kooperation mit dem CORC-Projekt. Mit dem noch laufenden Metadatenprojekt und dem abgeschlossenen SSG-FI Projekt hat die SUB in Bezug auf Metadaten und Katalogisierung von Online-Ressourcen fundierte Erfahrungen gesammelt, um diese in die internationale Projekt-Kooperation einbringen zu können.

Im OCLC-Verbund sind Bibliotheken aus aller Welt Mitglied, und das Internet ist ein global verfügbares Informationsinstrument. Durch die direkte Beteiligung der Bibliotheken soll erreicht werden, diesen neuen Dienst von vornherein so anwendungs- und benutzerfreundlich als möglich zu gestalten. Dazu wurden Teilnehmertreffen organisiert, die sowohl am Sitz von OCLC in Dublin/Ohio stattfanden, als auch auf den Kongressen der American Library Association.<sup>5</sup>

## 2. Teilnehmertreffen in Dublin/Ohio

### 2.1 April 1999

Das erste Treffen der Teilnehmer fand vom 27.-28. April 1999 in Dublin/Ohio statt. Zu diesen Zeitpunkt war die SUB Göttingen dem Projekt gerade beigetreten. Es fanden sich fast 90 Teilnehmer ein, in erster Linie aus den USA, doch auch Partner aus Großbritannien, Mexiko und Taiwan waren dabei. Die Deutsche Bibliothek nahm als Interessentin teil - inzwischen ist auch sie Partnerbibliothek geworden.

---

4 Informationssystem der Niedersächsischen Staats- und Universitätsbibliothek Göttingen <<http://www.sub.uni-goettingen.de/>>

5 An den Kongressen nahm Göttingen als Projektpartner nicht teil.

Damals waren 55 Institutionen eingetragene Projektteilnehmer (u.a. 30 akademische Institutionen, 3 Nationalbibliotheken, 6 OCLC-Networks in USA). In der Plenarversammlung wurde das Projekt vorgestellt; einige Erfolgsmeldungen durften dabei nicht fehlen: täglich wurden ca. 20-30 Sessions registriert und ca. 25-125 Datensätze angelegt (das ist ca. 1% der OCLC-Gesamtaktivitäten). Das System wird im laufenden Betrieb aufgebaut und soll bis ca. Juli 2000 voll funktionsfähig sein.

Hier wurden auch die neuen Werkzeuge vorgestellt, die das Katalogisieren in CORC erleichtern sollen:

- die Software „Mantis“<sup>6</sup> (basierend auf RDF und XML) ermöglicht die Auswertung der Homepages der elektronischen Ressourcen, sofern dort korrekte Metadaten vorhanden sind.
- Die Software „Scorpion“<sup>7</sup> wählt automatisch anhand des Dewey-Indexes die Begriffe auf der Homepage der Ressource (nach Häufigkeit) aus und ordnet sie den entsprechenden Dewey-Klassifikationen zu, i.a. werden drei Notationen angeboten, die Treffsicherheit ist allerdings noch nicht ausreichend hoch.

Die Arbeit fand in sogenannten BOFs (Break-Out-Forums oder Arbeitssitzungen) zu speziellen Themen statt. Zur Auswahl standen: Authority Control Use, Dublin Core Use, MARC Forum, Pathfinder Forum, Managing Digital Collections, Cartographic and Images, CORC System, International Scope, DDC/TEI/EAD/Thesauri, Government Documents and Serials, Envisioning CORC Public Services.

Die Ergebnisse aus den BOFs wurden am Nachmittag in summarischer Form (jeweils 10 Min.) vorgestellt. Am Abend lud OCLC zu einem gemeinsamen Essen ein, das die Möglichkeit bot, sowohl mit den Mitarbeitern des Projektes als auch mit den Teilnehmern ins Gespräch zu kommen und weitere Fragen zu klären.

Am folgenden Tag wurden die Teilnehmer in sechs Gruppen aufgeteilt. Jede der Gruppen sollte die wichtigsten Prioritäten und Ziele für CORC erarbeiten. Es ging dabei um folgende Punkte: *Searching, Resource description creation/editing, Pathfinder creation/editing, General, Long-term focus for the project.*

Natürlich hatte bei *Searching* der Nutzerzugang über Z39.50 bei allen Priorität.

---

6 Shafer, Keith: Mantis, A flexible cataloging toolkit. Oct. 22, 1998

<<http://orc.rsch.org:6464/toolkit.html>>

7 Shafer, Keith: A brief introduction to Scorpion

<<http://orc.rsch.oclc.org:6109/binintro.html>>

Für die *Resource description/editing* wurden einheitliche Standards und Richtlinien gewünscht, ebenso eine effektive Dublettenkontrolle sowie zusätzliche Thesauri.

Bei der *Pathfinder creation/editing* wurde die Angabe von Druckwerken auf lokaler Ebene gewünscht, außerdem ein einfacheres Editieren der Pathfinder.

Beim Thema *General* gab es den einhelligen Wunsch nach Einsatz des Unico-de für Diakritika, ebenso zum Normdaten-Linking, einem Userguide mit Beispielen und einer Einführungsseite (Tutorial) im Internet, sowie einem verbesserten Datenimport.

Als *long-term focus* für das Projekt stand das Archivieren an erster Stelle, gefolgt von Standards, Partnerschaft (DC<sup>8</sup> Coalition), der Einbeziehung von CORC in den WorldCat. Viele dieser Wünsche sind inzwischen realisiert; die CORC-Mitarbeiter sind hochmotiviert und zeigen sich stets aufgeschlossen für Anregungen.

Auch für die Teilnehmer brachte das Treffen viele positive Ergebnisse. Der Erfahrungsaustausch untereinander hinsichtlich der Arbeitsweise in den einzelnen Bibliotheken war für viele sehr wertvoll.

## 2.2 November 1999

An dem Treffen nahmen 87 Personen teil, darunter zwei Kollegen aus dem Institute of Information Science in Maribor (Slowenien) sowie Kollegen von der University of Strathclyde (Schottland) und vom Monterrey Institute of Technology (Mexico), zwei Mitarbeiter der SUB Göttingen sowie je ein Vertreter der Deutschen Bibliothek (DDB) und des Bibliotheksservice-Zentrums Baden-Württemberg (BSZ).

Am Projekt waren damals 114 Institutionen beteiligt, davon 75 aus dem akademischen Bereich, 28 Regierungsinstitutionen, 23 ausländische Institutionen, 5 Bibliotheksschulen und einige öffentliche und Spezialbibliotheken. Das Anwendertreffen war wie das vorherige organisiert: Plenarsitzung, Arbeitsgruppen, Ergebnis-Diskussion in der Schluss-Plenarsitzung.

Diesmal war es allen Teilnehmern ein großes Anliegen, Näheres über die zukünftige *Preisstruktur* zu erfahren, doch dafür wurden gerade erst Pläne erstellt. Ein ausgereiftes Preismodell wird wohl erst gegen Projektende vorliegen. Bei einem Gespräch (Focus Group), das ein Consultant im Auftrag von OCLC mit einigen Teilnehmern zu dieser Frage führte, wurden verschiedene

---

8 Dublin Core Metadata Initiative <<http://purl.org/DC/index.htm>>

Preismodelle diskutiert. Die meisten Teilnehmer wollten dieselbe Preisstruktur wie bei WorldCat beibehalten.

Es zeigte sich, dass Dublin Core von den amerikanischen Teilnehmern bisher kaum benutzt wird. Fast alle Teilnehmer arbeiten mit MARC, da ihre Systeme Dublin Core nicht verarbeiten können.

Ein weiterer allgemeiner Wunsch war die Angabe der notwendigen Hard-/Software-Ausstattung der Arbeitsplätze, da es auch in USA öfters Budget-Schwierigkeiten gibt.

OCLC hat die CORC-Anwendungen auf Internet Explorer ausgerichtet, fast alle Bibliotheken nutzen aber Netscape, weswegen es teilweise zu Anwendungsschwierigkeiten kommt (vor allem bei älteren Netscape-Versionen).

In Zukunft soll CORC nicht nur für elektronische Ressourcen, sondern auch für Objekte genutzt werden. Die CORC-Daten sollen in den WorldCat eingebracht werden, d.h. sie erhalten dafür auch neue ID-Nummern.

OCLC möchte folgende Standards bedienen: MARC 21, Dublin Core, TEI, GILS, EAD.

Die meisten Titelaufnahmen wurden von der Texas State Library gemacht. Viele Anwender haben bisher nur eine geringe Anzahl von Titeln (unter 100) eingegeben, da auch in USA dieses Projekt neben dem normalen Geschäftsgang läuft und nur selten bereits integriert ist.

Die Verbesserung der *Authority Control* soll in den kommenden Monaten erfolgen.

Die Teilnehmer sollen auch vorläufige Normsätze (in OCLC) anlegen können, falls keine vorhanden sind. Es soll ein Minimal-Niveau festgelegt werden. Man wird mit Multi-Scheme-Authorities (Ansetzung je nach Land<sup>9</sup>) beginnen, die Indexierung soll verbessert werden. Die Z39.50-Suche wird gerade getestet.

Für den CORC *Import/Export* wünschten die Teilnehmer, dass mehr als zehn Datensätze im-/exportiert werden können, nur gibt es hier Schwierigkeiten mit den Down-/Upload-Zeiten.

Geplant ist auch, die CORC-Titelaufnahmen dann sofort im WorldCat anzuzeigen, umgekehrt sollen WorldCat-Datensätze über Nacht in die CORC-Datenbank eingespielt werden.

Für die *URL-Pflege* bei CORC wird mittels eines Harvest-Programms die URL (nur HTML-Seiten) überprüft. Die fehlerhaften URLs werden der Bibliothek an-

---

9 Lt. Frau Hengel-Dittrich (Die Deutsche Bibliothek) gibt es bei den Namensansetzungen ca. 10% Abweichungen.

gezeigt, die die Aufnahme gemacht hat. Die Häufigkeit der URL-Kontrolle wurde noch nicht festgelegt

Ein Problem sind immer noch die Antwortzeiten. Sie sind nicht nur bei uns in Göttingen langsam, sondern auch in den USA; speziell die Library of Congress, inzwischen auch Projektpartner, berichtete von ähnlichen Schwierigkeiten.

Durch die laufende Anpassung der Datenbank kommt es immer wieder zu Systemausfällen. Eine Trennung der Server für Entwicklung und Nutzung wird seit langem gewünscht. Sie ist für ca. Mai 2000 vorgesehen.

Die Dokumentation für CORC soll verbessert werden, einmal durch eine kontext-sensitive Online-Hilfe, aber auch durch verbesserte Unterlagen. Speziell zu den Pathfindern wurde eine bessere Dokumentation von der Mehrheit als notwendig befunden.

Auch diesmal wurde wieder die Festlegung von Minimalstandards gewünscht, ebenso die Verbesserung des Suchindexes.

Prioritäten bei der **Weiterentwicklung** von CORC:

An erster Stelle stand bei allen Gruppen die *URL-Pflege*. In der Zusammenfassung von OCLC wurde dann folgenden Themen erste Prioritäten eingeräumt: URL-Pflege, Normdaten-Pflege und -Einsatz, Unicode-Einführung, Dublin Core Version 1.1, Integration in WorldCat, Z39.50-Suche (zumindest im Test), bessere Dokumentation, Qualität der Aufnahmen, Dublettencheck.

### 3. Das CORC Projekt an der SUB Göttingen

Die Teilnahme der SUB Göttingen an dem international ausgerichteten CORC Projekt hat mehrere Gründe. Zum einem konnten mit dem noch laufenden Metadatenprojekt<sup>10</sup> und dem abgeschlossenen SSG-FI Projekt<sup>11</sup> in Bezug auf Metadaten und Katalogisierung von Online-Ressourcen fundierte Erfahrungen gesammelt werden, um diese in die Projekt-Kooperation einbringen zu können. Auf der anderen Seite könnte CORC das Internet-Katalogisierungssystem in Bibliotheken weltweit werden, so dass hier eine frühzeitige Teilnahme nützlich sein kann.

Das CORC-System bietet zwei wesentliche Datenbanken an: Die erste Datenbank enthält die Katalogisate (*Record Creation*) und in einer gesonderten Datenbank können fachspezifische Linklisten (sogenannte *Pathfinder*) entweder aus den Katalogisaten generiert oder neu erstellt werden.

---

10 SUB Göttingen Metadata Server <<http://www2.sub.uni-goettingen.de/>>

11 SSG-FI Homepage <<http://www.sub.uni-goettingen.de/ssgfi/>>

Als Beitrag der SUB Göttingen in beide Datenbanken werden folgende Datenpools nachgenutzt bzw. für Testzwecke (siehe unten) verwendet:

#### • **Nachnutzung der SSG-FI Guides**

In den SSG-FI Guides werden weltweit recherchierte, fachlich relevante Informationsquellen, die im Internet bereitgestellt werden, gesammelt und nach einem international anerkannten Metadatenstandard (Dublin Core Metadata Element Set) katalogisiert. Die Titelaufnahmen werden in einer Datenbank (Allegro/Avanti-Server) nach fachlichen und formalen Kriterien erschlossen, evaluiert und dynamisch generiert im Internet strukturiert angeboten. Somit dienen die SSG-FI Guides dem Wissenschaftler als Anlaufstelle für jeweils aktuelle und relevante fachspezifische Informationen.

Die SSG-FI Guides stellen im internationalen Kontext sogenannte „High Quality Subject Gateways“ dar, die weltweit genutzt werden, da sie durchweg in englischer Sprache vorliegen (Interface, Metadaten etc.). In Göttingen werden zur Zeit vier Fachinformationsführer für folgende Fachgebiete angeboten: *Geo-Guide*<sup>12</sup> (Earth Sciences, Geography, Thematic Maps, Mining), *MathGuide*<sup>13</sup> (Mathematics) und für den Anglo-Amerikanischen Kulturraum der *History Guide*<sup>14</sup> sowie der *Anglistik Guide*.<sup>15</sup> Ein Fachinformationsführer *Forestry Guide* wird zur Zeit vorbereitet. Die Subject Gateways sind in der SUB Göttingen im Rahmen des Projektes SSG-FI (SonderSammelGebiets-FachInformationen im Internet) entwickelt und bereitgestellt worden. Das SSG-FI-System wird im Rahmen des DFG-Projektes „Virtuelle Fachbibliotheken (Virtual Libraries)“ von anderen Bibliotheken nachgenutzt.

Insgesamt stehen mehrere 1000 Datensätze zur Nachnutzung für „Record Creation“ zur Offline-Einspielung bereit. Dazu ist es erforderlich, ein Mapping zwischen dem SSG-FI Format, das Dublin Core basiert ist, und dem MARC 21 Format zu erstellen. Obwohl natürlich ein Mapping von SSG-FI nach Dublin Core vorliegt, erweist es sich als erfolgversprechender, die Fülle der Metainformationen, die in SSG-FI erstellt wurden, in MARC 21 abzubilden. Dies hat mehrere Gründe: Zum einen liegt Dublin Core immer noch nicht in der endgültigen Fassung vor und ist damit noch kein internationaler Standard, auch wenn sogar bei der Europäischen Normungsinstitution (CEN)<sup>16</sup> zur Zeit Bestrebungen unternommen werden, Dublin Core zum europäischen Standard

---

12 Geo-Guide Homepage <<http://www.Geo-Guide.de/>>

13 MathGuide Homepage <<http://www.MathGuide.de/>>

14 History Guide <<http://www.HistoryGuide.de/>>

15 Vlib-AAC Anglistik Guide <<http://www.AnglistikGuide.de/>>

16 CEN – The European Committee for Standardization <<http://www.cenorm.be/>>

zu erklären (CWA - Metadata<sup>17</sup>). Besonders die Qualifier von Dublin Core unterliegen stetiger Diskussion. Dementsprechend ist das Mapping zwischen Dublin Core und Marc 21 im CORC System nicht aktuell und perfekt, genauso wenig wie das Mapping von SSG-FI nach Dublin Core nicht dem aktuellsten Stand entspricht. Zum anderen lassen sich eine Vielzahl von Metainformationen nicht (und wahrscheinlich auch in Zukunft nicht) in Dublin Core abbilden, wohl aber in MARC 21. Zum Beispiel gibt es dort ein Feld „General Notes“, in dem nach Rücksprache mit Experten der Library of Congress Informationen über Zugangsvoraussetzungen wie Software, Kosten etc. untergebracht werden können. In Dublin Core bietet sich dafür nur DC.Rights an, das allerdings eher für Informationen über Copyright bestimmt ist.

Auch bei der automatischen Erstellung von zahlreichen „Pathfindern“ kann das SSG-FI-System nachgenutzt werden. Durch die Fachklassifikation werden aus der Allegro-Datenbank für jeden Guide sogenannte Subject Catalogs erzeugt, die sich hervorragend als Ausgangsbasis für die *automatische* Erzeugung von Pathfindern eignen. Das dafür nötige Allegro-Script ist erstellt worden, die ersten Testläufe liefen sehr zufriedenstellend ab. Letzte Ungeheimheiten konnten bei dem 2. CORC-Participation Meeting im November mit dem dafür zuständigen Entwicklungsteam geklärt werden.

#### • Diverse Dokumente vom Web-Doc Server

Um wichtige, ausgewählte HTML Dokumente archivieren und damit langfristig vorhalten zu können, hat die SUB Göttingen einen Web-Doc-Server eingerichtet. Die monographischen Dokumente, die auf diesem Server bereit liegen, werden in CORC ebenfalls mittels der automatischen Harvest-Methode erfasst. Hierbei handelt es sich vielfach um deutsche Texte, die, wenn überhaupt, nur über „Meta-Tags<sup>18</sup>“ verfügen. Sie sind dadurch gut für die Evaluierung des CORC-Systems geeignet. Auch liegen diese Dokumente vorwiegend als PDF-Datei vor und sind deshalb aus Testgründen mit herangezogen worden.

---

17 CWA – Metadata: CEN Workshop Agreement - Metadata for multimedia information  
<[http://www.cenorm.be/news/press\\_notices/metadata.htm](http://www.cenorm.be/news/press_notices/metadata.htm)>

18 hier verwendet im Sinne der Meta-Angaben, die üblicherweise verwendet werden, damit die HTML Seiten von Suchmaschinen, WWW-Servern etc. notiert werden  
(z.B. <meta name=„description“ content=„....“>)

**• Digitalisierungen von Reiseliteratur zu Nordamerika (offline Lieferung)**

Das *Göttinger Digitalisierungs-Zentrum* (GTZ)<sup>19</sup> hat Dokumente für die Offline-Einspielung in das CORC-System bereitgestellt. Eine erste Lieferung erfolgte bereits im November, der Rest soll in Kürze nachgereicht werden (insgesamt ca. 600 Datensätze).

**• Ausgewählte HTML Seiten**

Für die Evaluierung des CORC-Systems wurden unterschiedlichste Arten von Dokumenten, speziell HTML-Dokumente, ausgewählt. Als Auswahlkriterien galten dabei folgende Punkte: unterschiedliche Sprache, Vorliegen von (DC-) Metadaten, von Metadaten und „Meta-Tags“ oder nur von „Meta-Tags“ bzw. das Fehlen sowohl von Metadaten als auch von „Meta-Tags“ im Dokument, Vorhandensein von Java-Script auf der Einstiegsseite, PDF-Datei mit den eigentlichen Informationen zu dem Thema der Seite.

Die SUB Göttingen untersucht das teilautomatische Erschließungsverfahren in CORC im Hinblick auf:

- die automatische Erfassung der DDC (Scorpion) und Schlagworte (Word Smith keywords),
- die automatische Erfassung der bibliographischen Angaben,
- den zeitlichen Arbeitsaufwand bei der Nachbesserung und/oder Korrektur der Titelaufnahmen bezüglich aller relevanten Felder (Titel, Autor, Beschreibung, Schlagworte, Klassifikation etc.).

Sie evaluiert dabei die in CORC aufgenommenen Katalogisate nach sachlichen und formalen Erschließungskriterien.

Alle Datensätze, die in der SUB Göttingen in das CORC System eingegeben werden, werden ausschließlich durch die automatische Harvest-Methode zur Datenerfassung unterstützt. Das bedeutet, dass die Erfassung elektronischer Ressourcen mittels URL im MARC 21 Format erfolgt, wobei jedes von CORC bereitgestellte Tool zur weitgehend automatischen Erschließung genutzt und jedes einzelne Modul getestet wird: Neben den bibliographischen Daten (Titel, Autor etc.) und der sachlichen Erschließung (Beschreibung, Keywords) sind es die Tools DDC Scorpion zur automatischen DDC-Erschließung und Word Smith (WS) zur Schlagworterfassung. Getestet und für jeden Datensatz extra evaluiert werden sowohl die automatische *sachliche* und *formale* Erschließung

---

19 Göttinger Digitalisierungs-Zentrum GDZ)

<[http://www.sub.uni-goettingen.de/gdz/en/main\\_en.html](http://www.sub.uni-goettingen.de/gdz/en/main_en.html)>

als auch das zusätzliche *Harvesting* von DDC (Scorpion) und Schlagworten (Word Smith = WS keywords). Dazu wird eine Access Datenbank gepflegt, in der zu jedem Datensatz folgende Informationen vermerkt werden:

- Art und Umfang von Metainformationen: „Meta-Tags“ und/oder Metadaten im Sinne von Dublin Core bzw. keine Metainformationen,
- Sprache des Dokumentes,
- Format des Dokumentes (HTML, PDF etc.),
- Vorhandensein von Java-Script auf der Einstiegsseite,
- Bewertung des automatischen Harvesting Prozesses bezüglich jedes relevanten MARC 21 Feldes, d.h. wie hoch ist der zeitliche Arbeitsaufwand bezüglich der Nachbearbeitung des betreffenden Feldes. Dies bezieht z.B. auch Fragen mit ein wie: Ist der Titel des Dokumentes vollständig und korrekt wiedergegeben? Stimmen die „geharvesteten“ Schlagwörter mit dem Inhalt des Dokumentes überein, wenn nicht, wie hoch ist der zeitliche Aufwand für Korrekturen, Ergänzungen?

Die Teilnahme am CORC-Projekt ermöglicht es der SUB Göttingen, neue Erfahrungen in der Katalogisierung von Online-Ressourcen zu sammeln. Es können Erfahrungen bei der Normierung von Inhalten einer Kategorie (Personenansetzung, Titel etc.) und in Fragen der Benutzung von Internationalen Standards bei der Erschließung von Online-Ressourcen gewonnen werden, sowie im Mapping von verschiedenen Formaten (Dublin Core zu MARC 21 und umgekehrt, SSG-FI zu MARC 21, MARC 21 zu PICA) und in Zusammenhang mit neuen Datenformate wie RDF und XML (zur Verwaltung von Metadaten). Durch das Herunterladen von in CORC bereitgestellten Datensätzen in das PICA-System können Export/Import Routinen in bestehende Datenbanken (PICA, SSG-FI) getestet werden.

### 3.1 Vorläufige Ergebnisse

OCLC hat das CORC-Projekt am 3.12.1999 den interessierten deutschen Bibliotheken in der SUB Göttingen vorgestellt. Bei dieser Gelegenheit wurde über die Erfahrungen berichtet, die die SUB im Rahmen des Projektes gesammelt hat. Anhand von zwei ausgewählten Beispielen wurden der Arbeitsprozess bei einer Online-Demonstration detailliert aufgezeigt und die Ergebnisse der verschiedenen Harvest-Prozesse konkret analysiert. Die Ergebnisse bei der automatischen Erfassung von Katalogisaten in CORC sollten die Stärken und Schwächen des Systems veranschaulichen.

Zum damaligen Zeitpunkt konnten noch keine Aussagen über die Pathfinder getroffen werden, da das Allegro-Script für die automatische Erzeugung der Pathfinder aus den SSG-FI Guides noch nicht einwandfrei arbeitete und letzte

Fehlerquellen von Seiten des CORC Systems erst noch vom amerikanischen Entwicklungsteam behoben werden mussten.

Abbildung 1 zeigt die CORC Einstiegsseite für Record Creation: Im linken Rahmen sind die Module für die einzelnen Arbeitsschritte wie z.B. Suche im Katalog, Titelaufnahme im Katalog, Anzeigen aller erstellten Datensätze unter Berücksichtigung der verschiedenen Statusanzeigen (neu, privat, komplett etc.) und Hilfsmodule wie „Authority File“, Dewey Decimal Classification etc. aufgelistet. Zusätzlich kann man jederzeit zu den Pathfindern gelangen, um z.B. eine erstellte Titelaufnahme einer bestimmten Linkliste (Pathfinder) zuzuordnen. Im rechten Rahmen befindet sich das jeweilige Arbeitsfenster, in diesem Fall das für die Erstellung einer Titelaufnahme. Dabei sind die folgenden Optionen aktiviert: Anzeige im MARC Format, Harvesting der Informationen aus der URL (hier URL der Geo-Guide Seite), Anzeigen aller schon in der Datenbank existierenden Aufnahmen mit dieser URL, automatische Generierung der DDC (Scorpion) und Schlagworte (WS keywords) und das Einbinden der in einer Standardaufnahme enthaltenen Informationen, die für jede Titelaufnahme relevant sind und somit als konstant betrachtet werden können.

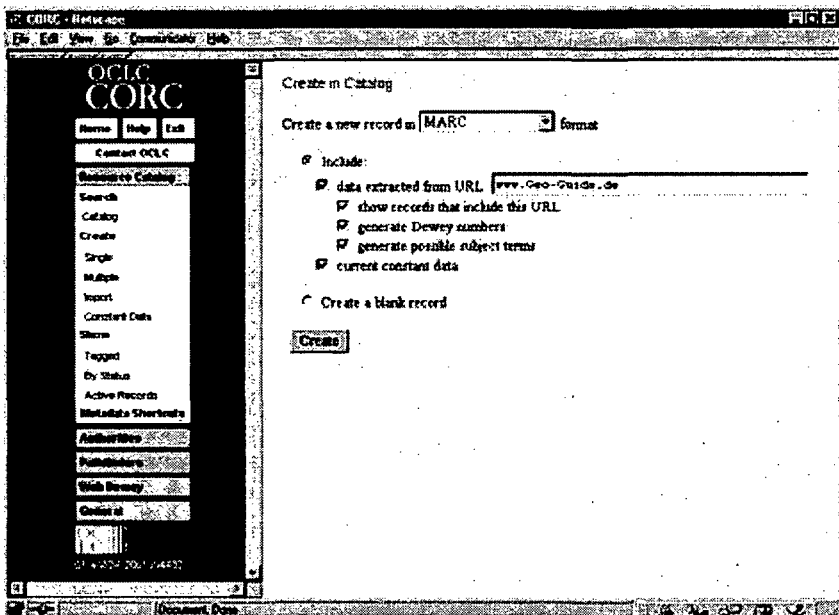


Abb. 1: CORC Bearbeitungsseite für die Erstellung einer neuen Titelaufnahme

### 3.1.1 Record Creation

Als positives Beispiel wurde die *Geo-Guide* Einstiegsseite ([www.Geo-Guide.de](http://www.Geo-Guide.de)) gewählt. Die Dokumente liegen in englischer Sprache vor, der Quelltext verfügt neben „Meta-Tags“ über Dublin Core codierte Metadaten und es handelt sich ausschließlich um HTML-Dateien.

Bei dem zweiten Beispiel, Forschungsbericht 99 („*Fober*“)<sup>20</sup> handelt es sich um ein deutschsprachiges Dokument, das im Quelltext weder über Metadaten noch über „Meta-Tags“ verfügt und dessen Hauptinformationen in einer PDF-Datei abgelegt sind. Dieses HTML-Dokument ist auf dem Web-Doc-Server der SUB archiviert.

Der Source Code der *Geo-Guide* Einstiegsseite enthält im HTML-Header folgende Metadaten und Meta-Tags (aus Gründen der Übersichtlichkeit ist der Text der Kategorien-Inhalte stark gekürzt worden):

```
<HTML>
<HEAD
<TITLE>GeoGuide - Subject Gateway...</TITLE>
Meta-Tags
<META NAME=„keywords“ CONTENT=„GeoGuide; SUB....“>
<META NAME=„description“ CONTENT=„The Lower....“>
Metadaten:
<META NAME=„DC.Title“ CONTENT=„GeoGuide - Subject....“>
<META NAME=„DC.Creator“ CONTENT=„Heike Neuroth....“>
<META NAME=„DC.Publisher“ CONTENT=„Lower Saxony....“>
<META NAME=„DC.Subject“ SCHEME=„GOKverbal“ CONTENT=„ maps....“>
<META NAME=„DC.Subject“ CONTENT=„GeoGuide; SUB...“>
<META NAME=„DC.Description“ CONTENT=„The Lower...“>
<META NAME=„DC.Type“ CONTENT=„WWW“>
<META NAME=„DC.Date“ CONTENT=„19991105“>
<META NAME=„DC.Format“ CONTENT=„text/html“>
<META NAME=„DC.Language“ CONTENT=„en; de“>
<META NAME=„DC.Identifier“ CONTENT=„http://www....“>
<META NAME=„DC.Rights.Access“ CONTENT=„free“>
<META NAME=„DC.Rights.Restriction“ CONTENT=„none“>
<META NAME=„SSG.Country“ CONTENT=„DE“>
<META NAME=„SSG.Format“ CONTENT=„ff“>
<META NAME=„SSG.Size“ CONTENT=„over 1,800 entries“>
</HEAD>
```

Das Ergebnis des automatischen Harvesting der *Geo-Guide* Einstiegsseite mit allen von CORC bereitgestellten Tools ist in den folgenden Abbildungen (Abb.

---

20 Forschungsbericht 99 <<http://webdoc.sub.gwdg.de/ebook/a/1999/fober>>

2: Ansicht MARC 21 Format) dargestellt. *Abbildung 3* zeigt die überarbeitete Titelaufnahme in der Dublin Core Ansicht.

Actions:	Displaying Record	1 2 ... 22	View:	Dublin Core (2)
CORC:	210963 (7)	Created:	ZKW 1999-11-29	
Status:	Complete	Modified:	ZKW 2000-02-09	
<b>Title</b>	Geo-Guide - Subject Gateway in the Earth Sciences			
<b>Identifier.URL</b>	<a href="http://www.Geo-Guide.de">http://www.Geo-Guide.de</a>			
<b>Publisher</b>	Lower Saxony State and University Library Göttingen			
<b>Creator.PersonalName</b>	Neuroth, Heide			
<b>Date.Issued</b>	1999-11-29			
<b>Description</b>	The Lower Saxony State and University Library Göttingen (Germany) offers an internet-based subject gateway to scholarly relevant information in the Earth Sciences, called Geo-Guide. The page is the starting point to an earth science internet information center. Users can browse by a subject or source type catalog or search for all metadata categories (title, editor, keywords, description etc.), which is available for each resource.			
<b>Format.MIME</b>	text/html			
<b>Language</b>	english			
<b>Relation.Requires</b>	Mode of access: World Wide Web			
<b>Rights</b>	Geo-Guide, SUB Göttingen, 1997-2000			
<b>Subject.Keyword</b>	Geo-Guide; SUB; virtual library, metadata, internet resources,			
<b>Subject.Keyword</b>	geography, maps, earth sciences, geophysics, geodesy, cartographical research, hydrology, oceanography, .....			
<b>Subject.Keyword</b>	910.912; 550; 551.46; 333.914; 915; 200.9; 409; 700; 793; 509; .....			
<b>Subject.Keyword</b>	74.00; 01.99; 38.10; 38.70; 10.99; 11.99; 15.00; 17.00; 20.99; .....			
	1 2 ... 22			

Abb. 2: Ergebnis des automatischen Harvestings der Geo-Guide Einstiegsseite im MARC 21 Format

Action:	Displaying Record		12..22	View:	MARC			
CORC:	210963 (3)	Created:	ZKW 1999-11-29					
Status:	Private	Modified:	ZKW 1999-11-29					
<u>100</u>	1	<u>Neuroth, Heike</u>						
<u>245</u>	00	GeoGuide - Subject Gateway in the Earth Sciences (SUB G) in [computer file]						
<u>246</u>	00	GeoGuide - Subject Gateway in the Earth Sciences (SUB G						
<u>260</u>		in Lower Saxony State and University Library G :o 1999-11-29						
<u>500</u>		The Lower Saxony State and University Library G						
<u>500</u>		The Lower Saxony State and University Library G						
<u>540</u>		free						
<u>540</u>		none						
<u>500</u>		GeoGuide - Subject Gateway in the Earth Sciences (SUB G) in Göttingen, Germany) <META						
		NAME="description" CONTENT="The ...						
<u>540</u>		GeoGuide, SUB Göttingen, 1997-99						
<u>653</u>		GeoGuide; SUB; virtual library, metadata; internet resources; ...						
<u>653</u>		geography, maps, earth sciences; geophysics; geodesy, extraterrestrial research, hydrology,						
		oceanography, ...						
<u>653</u>		910; 912; 550; 551.46; 333; 914; 915; 200.9; 409; 700; 793; 509; ...						
<u>653</u>		74.00; 01.99; 38.10; 38.70; 10.99; 11.99; 15.00; 17.00; 20.99; ...						
<u>653</u>		GeoGuide; SUB; virtual library, metadata; internet resources; ...						
<u>653</u>		WWW						
<u>690</u>		records, author, offering, sub, source						
<u>699</u>	1	<u>025.317</u>	Conversion and coarctation					
<u>699</u>	1	<u>623.0155</u>	Military geophysics					
<u>699</u>	1	<u>346.08688</u>	Title insurance--law					
<u>856</u>	40	*q text/html :u <a href="http://www.SUB.Uni-Goettingen.de/raff/geo55ez.html">http://www.SUB.Uni-Goettingen.de/raff/geo55ez.html</a>						
<u>856</u>	40	*q text/html :u <a href="http://www.sub.uni-goettingen.de/raff/geo">http://www.sub.uni-goettingen.de/raff/geo</a>						
<u>###</u>								
Type:	ELv1:	E	freq:	d	Auth:	Ctrl:	Lang:	xx
ELv1:	Form:		Cont:	O	Blag:	ORac:	Clry:	
	Cont:		CPub:		Flot:	O	Inde:	O
Desc:	ILL:		Feat:	O	Dist:	s	Date:	1999, 1129

Abb. 3: Titelaufnahme Geo-Guide in der überarbeiteten Fassung im Dublin Core Format

Auf den ersten Blick ist die Fülle der Informationen bzw. die Dichte der Belegung der notwendigen Kategorien auffällig. Auch wenn einige Kategorien

doppelt (vgl. Kat. 500 = description, Kat 653 = aus Metadaten generierte keywords) bzw. sogar z.T. dreifach belegt sind, so sind doch alle notwendigen Informationen zu der Seite überliefert worden. Eine echte Verlinkung zu dem Personen-Normsatz der LoC Normdatei ist jetzt möglich (vgl. Kat. 100). Die automatisch generierten WS keywords ergeben keinen bzw. kaum Sinn (vgl. Kat. 690: records, author, offering, sub, source) und spiegeln nicht den Inhalt der Geo-Guide Seiten wider. Bei den automatisch generierten DDC Scorpion sieht es ähnlich aus, sie sind inhaltlich nicht dem Thema der HTML-Seite entsprechend (vgl. Kat. 699: z.B. Military Geophysics, Title insurance-law). Der Sprachcode konnte wegen abweichendem ISO Code im SSG-FI Projekt nicht erkannt werden (vgl. Lang: xx). Der Unicode ist noch nicht fehlerfrei implementiert (vgl. die dritte 500-er Kategorie: Gttingen statt Göttingen). Insgesamt beläuft sich trotz einiger Kleinigkeiten der Nachbearbeitungsaufwand hier auf wenige Minuten, auch wenn das Feld description (Kat. 500) neu geschrieben werden muss (vgl. Abb. 3). *Abbildung 3* zeigt die überarbeitete Fassung mit der neuen URL, auch hier ist deutlich zu sehen, dass der Unicode nicht funktioniert (vgl. z.B. DC.Publisher).

Der Source Code der „Fober“ Einstiegsseite sieht bezüglich des HTML-Headers wie folgt aus:

```
<head>
<!-- <link REL=„STYLESHEET“ HREF=„.../ftel4.css“ type=„text/css“ --> -->
<style type=„text/css“>
<!--
.nav { font-family: Arial, Helvetica, sans-serif; font-size: 10pt; font-weight: bold; color: #FFFFFF; text-decoration: none }
.link { font-family: Times New Roman, serif; font-size: 12pt; font-weight: bold; color: #993300; text-decoration: none }
-->
</style>
<meta name=„GENERATOR“ content=„Microsoft FrontPage 3.0“>
<meta HTTP-EQUIV=„Content-Type“ CONTENT=„text/html; charset=windows-1252“>
<meta NAME=„GENERATOR“ CONTENT=„Microsoft FrontPage 3.0“>
<title>Forschungsbericht 1999</title>
</head>
```

Im Source Code sind weder „Meta-Tags“ noch Metadaten vorhanden, deshalb sind durch das automatische Harvesting sehr wenig Informationen über

die HTML-Seite abrufbar. Lediglich der wenig aussagekräftige Titel für die Browser-Leiste wird im Source Code angegeben. Dementsprechend dürrig ist der durch das Harvesting in CORC generierte Datensatz: *Abbildung 4* (Ansicht MARC 21 Format). Besonders bei diesem Beispiel ist zu sehen, dass das automatische DDC Harvesting bei deutschsprachigen Seiten, die auch noch im PDF-Format vorliegen, erwartungsgemäß nicht funktioniert (vgl. Kat. 699: Gems-prospecting). *Abbildung 5* (überarbeitete Titelaufnahme im DC-Format) zeigt eindrucksvoll, wie hoch hier der Nachbearbeitungsaufwand ist, im Prinzip muss die komplette Titelaufnahme manuell neu erstellt werden. In diesem Beispiel müssen sowohl die vorhandenen Kategorien wie Titel (Kat. 245) und Beschreibung (vgl. Kat. 500) überarbeitet werden als auch eine Fülle von Kategorien ergänzt werden. Erstaunlicherweise hat das CORC-System hier die Sprache des Textes richtig erkannt und umgesetzt (vgl. Lang: ger). Allerdings zeigt auch dieses Beispiel deutlich, dass bisher die Unicode-Implementierung nicht funktioniert.

Actaria: Display Record View: MARC

CORC: 104091 Created: EXW 1999-07-02  
 Status: Complete System: OCL 1999-10-27

001 1 1 m d  
 010 1 1 Scherwatzki, Bernhard  
 110 3 Bundesminister Hans J. Scherwatzki [aust] Österreich  
 245 0 Forschungsbereich 1999 - Schwerpunkt [aust] Österreichische Forschungsstrategie Phase I [aust] Bericht des Bundesministers Hans J. Scherwatzki [aust] Wissenschaft und Verkehr an den Nationalen  
 514 Computer document(s)  
 520 [aust] Bernhard Scherwatzki [aust] born at [aust] 1999-07-02  
 500 Der Schwerpunktbericht des Bundesministers Hans J. Scherwatzki enthält einen Bericht zum gegenwärtigen Lage und den Bedürfnissen der Forschung in [aust] Österreich samt den entsprechenden statistischen Daten  
 501 Title from Web page  
 514 Research report  
 514 System requirements: Adobe Acrobat Reader  
 514 Forschungsbereich, Forschung, Technologie, Entwicklung [aust] Österreichische Forschungsstrategie, [aust] Österreich, Nationalrat, Forschungsorganisationen gesetz

011 1 022 General libraries, archives, information centers  
 011 1 318 726 Information policy

516 40 [aust] text/html [aust] http://www.bnf.ac.uk/Recherche/fober/01/index.htm  
 516 40 [aust] text/html, text/pdf [aust] http://welder.tuh-graz.de/fober/1999/foberindex.htm

###  
 Type: n Rel: k Rec: d Ann: Crl: Lang: ger  
 Rel: m File: Cpb: Mkn: Ctr: g

Abb. 4: Ergebnis des automatischen Harvestings der „Fober“ Einstiegsseite im MARC 21 Format



betrachtet der Tatsache, dass schon aufgenommene Datensätze diesbezüglich alle überarbeitet werden müssen, sobald die Unicode-Implementierung erfolgreich ist, sollte diese Aufgabe von Seiten des CORC-Entwicklungsteam vorrangig angegangen werden.

- Bei der „Authority Control“ erfolgt die Sortierung leider nicht alphabetisch, sondern nach einem sogenannten „Relevance Ranking“ (Originalzitat aus einer Rückantwort des CORC-Entwicklungsteams). Allerdings bleibt unklar, um welche Relevanzen es sich dabei handelt.
- Besonders hilfreich ist, dass die DDC nun in das CORC System implementiert wurde, so dass ein Blättern und die Suche nach Schlagworten in der DDC möglich sind. Korrekturen bei der automatischen DDC-Scorpion lassen sich nun wesentlich weniger zeitaufwendig durchführen. Auch eine Konkordanz zu den LCSH (Subject Headings der LoC) und zu verwandten Systemen (z.B. NetFirst, OCLC WordCat) wird angezeigt.
- Beim Mapping zwischen MARC 21 und Dublin Core werden zusätzliche Qualifier verwendet, die noch nicht zum DC-Standard gehören, aber für die Abbildung der MARC 21 Daten wichtig sind. Trotzdem werden natürlich nicht alle MARC 21 Felder nach Dublin Core umgesetzt bzw. auch nicht immer vollständig (z. B. fehlt bei der Serienangabe die Zählung).
- Es fehlt ein Programm zum Harvesting von PDF-Dateien.
- Die geplante Unterstützung des Z39.50 Protokolls ist für den Datenaustausch zwischen CORC und anderen Systemen positiv zu bewerten.

Bezüglich der **formalen Erschließung** sieht das bisherige Ergebnis wie folgt aus:

Sobald Metadaten im Source Code vorhanden sind, ist i.A. der zeitliche Aufwand für die Nachbearbeitung der bibliographischen Daten gering.

Bei der Eingabe in MARC 21 ist die Anwendung der AACR2 Regeln notwendig, sowie die Interpunktionsregeln nach ISBD. Die Umsetzung der MARC-Daten für die Übernahme der Daten aus CORC in das lokale System muss geregelt sein, da in Deutschland mit anderen Formaten und Regelwerken gearbeitet wird und die Nachbearbeitung dementsprechend erheblich sein kann. Hilfreich wäre hier, wenn wie vom amerikanischen Entwicklungsteam angekündigt und auf dem 2. Kooperationsstreffen im November in Dublin/Ohio gefordert, die Definition eines Minimal-Standards bei der inhaltlichen Belegung der Kategorien festgelegt werden würde. Ein „Minimal Quality Standard“ für die Erstellung von Datensätzen würde garantieren, dass die von deutscher Seite erstellten Datensätze in CORC auch von anderen CORC-Teilnehmern nachgenutzt werden, die Experten im Umgang mit dem MARC 21 Format sind

und somit gewisse Ansprüche an einen Datensatz stellen. Um den Aufwand für die Nachbearbeitung der Datensätze bei der Übernahme von CORC Daten in deutsche Systeme möglichst gering zu halten, ist eine Annäherung der Regelwerke (AACR2, RAK-WB) wünschenswert und mit RAK 2 auch in Arbeit.

Die Normdaten der Library of Congress weichen z.T. in ihrer Ansetzung von den deutschen Normdaten ab.

Bezüglich der **sachlichen Erschließung** sieht das bisherige Ergebnis wie folgt aus:

Wenn ausreichend Metadaten im Source Code einer HTML-Seite vorhanden sind, wird der zeitliche Aufwand für die Nachbearbeitung spürbar reduziert. Die Felder/Elemente „description“ und „keywords“ müssen höchstens ergänzt oder in Kleinigkeiten korrigiert werden.

Das automatische DDC (Scorpion) und Keyword Harvesting (WS) funktioniert bei englischen Dokumenten zu ca. 50 %, bei deutschsprachigen Seiten kaum.

Der Nachbearbeitungsaufwand liegt bei den HTML-Seiten mit guten Voraussetzungen (Metadaten, englischsprachig etc.) bei maximal 10 Minuten, bei anderen Seiten kann er sich auf über 30 Minuten erstrecken. Besonders problematisch erweisen sich Web-Seiten, die im PDF-Format vorliegen. Hier gelingt es kaum, die notwendigen Informationen mittels Harvesting zu ermitteln. Der zeitliche Aufwand für Ergänzungen ist so hoch, dass die Titelaufnahme auch gleich manuell erstellt werden kann. Im Hinblick auf die große Menge an PDF-Ressourcen sollte die Entwicklung eines Harvesting-Programms von OCLC vorangetrieben werden.

### **3.1.2 Pathfinder**

Pathfinder sind Linklisten bzw. sogenannte „Webliographien“, die völlig unabhängig von der „Record Creation“-Datenbank manuell oder automatisch erzeugt werden können. Damit handelt es sich um eine gesonderte Datenbank, in die allerdings auch mit Hilfe der Titelaufnahmen aus der „Record Creation“-Datenbank Pathfinder erstellt oder bereits existierende ergänzt werden können. Dabei werden aus dem kompletten Katalogisat nur der Titel, die URL und die Beschreibung übernommen.

Die Zuarbeit der SUB Göttingen zu der „Pathfinder“-Datenbank besteht in der Nachnutzung des SSG-FI Systems: Unter Allegro wurde eine Exportroutine entwickelt, mit der aus den Daten der SSG-FI Guides weitgehend automatisch Pathfinder erzeugt werden, die dann in diese CORC Datenbank einge-

bracht werden können. Diese Routine wurde zuerst erfolgreich am Geo-Guide getestet, kann aber auch für alle anderen Guides verwendet werden. Im Vorfeld ist noch zu überlegen, welche Subject Cataloge aus dem jeweiligen Guide mit Hilfe der Klassifikation (überwiegend „Göttinger Online Klassifikation“ GOK, im MathGuide MSC) aus Allegro exportiert werden sollen. Damit die Pathfinder nicht zu viele Treffer aufweisen und dadurch die Ladungszeiten zu lang werden, sollten ca. 70 Einträge nicht überschritten werden.

Für die Pathfinder werden die Felder mit dem Titel, der URL und der Beschreibung der Ressource ausgewertet. Die Hinzunahme von weiteren Feldern z.B. die SSG-FI Kategorie Evaluation ist problemlos möglich und kann mit in die Beschreibung als Zusatzinformation übernommen werden.

Die Pathfinder werden im XML/HTML Format ausgegeben und können auch für Web-Seiten verwendet werden. Jeder Pathfinder kann unter Berücksichtigung der Copyright-Bestimmungen z.B. für Service-Leistungen einer Bibliothek heruntergeladen werden und zur weiteren Bearbeitung und/oder format-technischen Veränderungen (Einbinden von Style Sheets etc.) kopiert werden. Zum Downloaden werden zwei Möglichkeiten angeboten: „Export Pathfinder Link“ erzeugt entweder eine PURL- oder URL-Adresse (Dynamic Pathfinder PURL bzw. URL), die auf den aktuellen Link zu dem Pathfinder verweist und problemlos in eine Web-Seite eingebunden werden kann, „Export HTML“ bietet den Pathfinder zum Downloaden als komplette HTML Datei an. Der Vorteil der ersten Möglichkeit besteht darin, dass mit Hilfe der dynamischen PURL oder URL der Pathfinder stets den aktuellen Stand wiedergibt. Hat sich z.B. eine URL von einem Link verändert und ist der Pathfinder diesbezüglich aktualisiert worden, so ist man mit dieser Download-Möglichkeit ebenfalls auf dem neuesten Stand.

*Abbildung 6* zeigt ein Beispiel einer dynamischen Pathfinder-Erzeugung aus dem Geo-Guide, speziell aus dem Subject Catalog Geology im Unterpunkt „Endogene Processes“. Der Vorteil der automatischen Erfassung von Pathfindern liegt auf der Hand: Es muss nicht manuell zu jeder URL (die ebenfalls in das entsprechende Pathfinder-Feld kopiert werden muss) der Titel und die Beschreibung erfasst werden. Im Prinzip kann aus jeder Linkliste, die mit den HTML-Codes für Definitionen codiert ist (<DL>, <DT> etc.), automatisch ein Pathfinder generiert werden. Die Pathfinder Überschrift und Beschreibung können entweder manuell hinzugefügt werden, oder wie im Fall der SSG-FI Subject Cataloge mittels XML Befehlen teilautomatisch erzeugt werden.

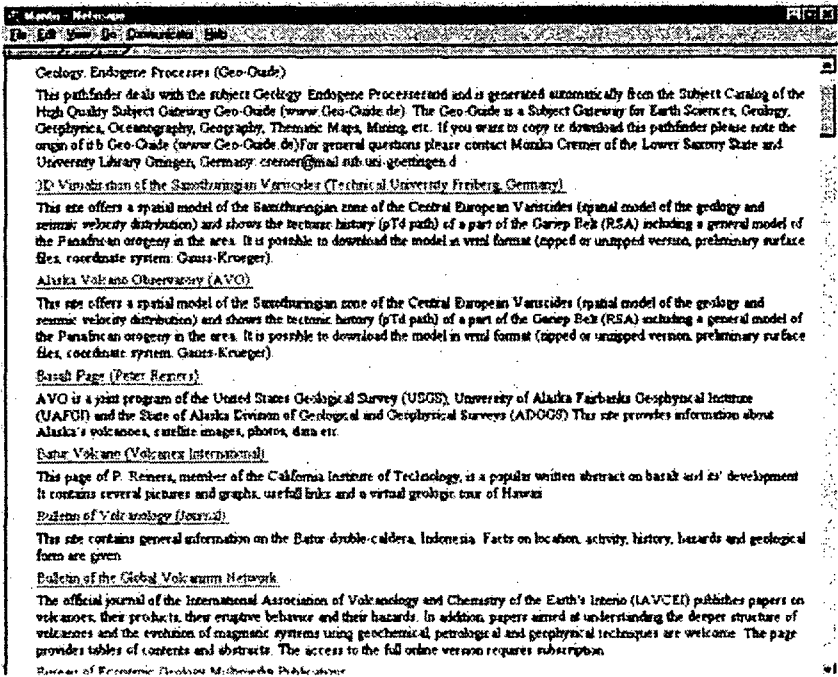


Abb. 6: Beispiel eines Pathfinders, automatisch generiert aus dem Subject Catalog des Geo-Guide (Geology: Endogene Processes)

Ergänzungen zu einem Pathfinder können direkt aus der CORC-Titel-Datenbank erfolgen. Wird z.B. eine Titelaufnahme zu einem speziellen Thema erstellt, zu dem es auch einen Pathfinder gibt, so kann diese Aufnahme markiert und in den Pathfinder direkt integriert werden. Titel, URL und Beschreibung werden dabei automatisch aus der Titelaufnahme entnommen.

Eine Reihe von Optionen wie z.B. zu Layout des Pathfinders (Trennlinien etc.), zu rechtlichen Fragen etc. können gewählt werden.

Insgesamt hat sich das CORC-System seit der letzten OCLC Präsentation im Dezember 1999 in einigen wichtigen Punkten entscheidend verbessert. So ist z.B. die Dokumentation zu den beiden Datenbanken Record Creation und Pathfinder umfangreicher geworden und liegt in einer aktuellen Version vor

(Stand März/April 2000).<sup>21</sup> Die umfangreiche Implementierung der DDC mit allen möglichen Hilfsmodulen und Konkordanzen erleichtert ein effektives Arbeiten.

#### 4. Ausblick

Heute beteiligen sich bereits 289 Bibliotheken und Institutionen rund um den Globus am CORC-Projekt, von Australien bis Südamerika. Vor allem Australien nutzt die Dublin Core-Plattform für die Eingabe. Die wahlweise Anwendung von MARC 21 und Dublin Core verleiht der CORC Datenbank eine große Flexibilität, die allerdings in Deutschland noch nicht ausreichend genutzt werden kann. Wenn Dublin Core sich als internationaler Standard durchsetzt - damit ist im World Wide Web durchaus zu rechnen - wird auch CORC weiterhin als Nachweisinstrument für die wachsende Menge von Online-Ressourcen an Bedeutung gewinnen. Allerdings wird Voraussetzung sein, daß den Online-Ressourcen bereits vom Verfasser brauchbare Metadaten beigegeben werden, um den Nachweis zu beschleunigen, was sich deutlich anhand der Göttinger Erfahrungen zeigt. Die Katalogisierer werden durch die neuen Funktionalitäten der CORC-Datenbank nicht abgeschafft, aber es wird den Bibliotheken ermöglicht, auch für Ressourcen im World Wide Web den Nachweis mit Hilfe der neuen Werkzeuge und durch globale Zusammenarbeit rasch und effektiv zusätzlich zu den gewohnten Dienstleistungen anzubieten.

Das Dublin Core Mapping wird in Zukunft auch für die hiesigen Bibliothekssysteme als Standard gefordert werden.



---

21 CORC System User Guide: Create, z.B. Import, and Export Records (April 2000, <[http://www2.oclc.org/oclc/pdf/create\\_recs\\_qr.pdf](http://www2.oclc.org/oclc/pdf/create_recs_qr.pdf)>), Create and Use Pathfinders (March 2000, <[http://www2.oclc.org/oclc/pdf/pf\\_qr.pdf](http://www2.oclc.org/oclc/pdf/pf_qr.pdf)>)