

Monika Cremer, Heike Neuroth

## Internationale Kooperation mit dem CORC-Projekt von OCLC an der SUB Göttingen



*Das Projekt „Cooperative Online Resource Catalog“ (CORC), von OCLC (Online Computer Library Center, Dublin/Ohio) bot der Niedersächsischen Staats- und Universitätsbibliothek (SUB) Göttingen im Rahmen eines DFG Projektes die Möglichkeit, neue Erschließungsmethoden für die kooperative Katalogisierung von Online-Ressourcen zu evaluieren. Dabei sind bibliotheksspezifische Gegebenheiten in Deutschland berücksichtigt worden. Der Artikel stellt die neuen Komponenten des CORC Systems vor, beschreibt detailliert die Arbeiten der SUB Göttingen in diesem Projekt und gibt einen Ausblick auf Entwicklungen im Bereich der kooperativen Erschließung.*

International cooperation with the CORC-project of OCLC at the Niedersächsische Staats- und Universitätsbibliothek (SUB) of Göttingen

*The project „Cooperative Online Resource Catalogue“ (CORC) was initiated by OCLC (Online Computer Library Center, Dublin/Ohio). In the context of a DFG project the SUB in Göttingen investigated and evaluated new methods of cooperative cataloguing of online resources developed for the CORC system. The DFG project considered the library specific issues of Germany. The paper describes in detail the new components of the CORC system as well as the work done at SUB and gives an outlook on developments in cooperative cataloguing.*

Coopération internationale avec le projet CORC de l'OCLC à la Niedersächsische Staats- und Universitätsbibliothek (SUB) de Göttingen

*Le projet „Cooperative Online Resource Catalogue“ (CORC) initié par OCLC (Online Computer Library Center, Dublin/Ohio) donne l'occasion d'étudier et d'évaluer des nouvelles méthodes du catalogage partagé des ressources en ligne. C'est ce qu'on fait actuellement à la SUB de Göttingen. Subventionné par la DFG on prend également en considération la situation spécifique allemande. L'article présente les nouveaux éléments du système CORC, décrit en détail les travaux exécutés à la SUB de Göttingen et discute les perspectives du catalogage partagé.*

## Inhalt

1	Einleitung: OCLC und CORC	262
2	Projektziel	262
3	Projektablauf	263
4	Projektarbeiten in Göttingen (mit genauer Beschreibung unserer Evaluationen)	266
5	Öffentlichkeitsarbeit für CORC in Deutschland	269
6	Ergebnisse des Projektes und zukünftige Entwicklungen	269
7	Literaturliste	270
8	Anhang	271

## 1 Einleitung: OCLC und CORC

Das Online Computer Library Center (OCLC)<sup>1</sup> in Dublin/Ohio ist das größte Non-profit-Katalogisierungsunternehmen der Welt. Mit dem WorldCat bietet OCLC die größte Datenbank mit bibliographischen Informationen: 48 Millionen bibliographischer Datensätze sind zugänglich<sup>2</sup>. Die Organisation bietet daneben u.a. auch Datenbanken, Fernleihdienstleistungen und den Zugriff auf archivierte Online Ressourcen an.

OCLC hatte schon vor Jahren mit INTERCAT<sup>3</sup> ein Projekt zur Katalogisierung von Online-Ressourcen begonnen. Doch inzwischen sind innovativere Lösungen gefragt, um digitale Medien und Informationsressourcen in die Informationsdienstleistungen der Bibliotheken zu integrieren. Als eine der wichtigsten Initiativen, internationale Metadatenstandards zur Erschließung elektronischer Dokumente zu erarbeiten und durchzusetzen, kann Dublin Core<sup>4</sup> angesehen werden. Dublin Core Metadaten können in bibliothekarische Katalogsysteme integriert und somit in vorhandene informationstechnische Strukturen eingebunden werden, eine wichtige Basis auch für den Datenaustausch.

Anfang 1999 startete OCLC das Projekt „Cooperative Online Resource Catalog“ (CORC) mit dem Ziel, einen Verbundkatalog digitaler Ressourcen in weltweiter Kooperation aufzubauen, der über das World Wide Web zugänglich ist. Die SUB Göttingen war die erste deutsche Bibliothek, die sich mit Unterstützung der Deutschen Forschungsgemeinschaft aktiv an diesem Projekt beteiligte<sup>5</sup>.

Das Projekt CORC wurde im Juni 2000 abgeschlossen<sup>6</sup> und in das Produktionsangebot von OCLC integriert, d.h. ab diesem Zeitpunkt steht es ausschließlich über eine Mitgliedschaft im OCLC-Verbund zur Verfügung.

## 2 Projektziel

Das CORC Projekt hatte den Aufbau eines Verbundkatalogs digitaler Ressourcen in weltweiter Kooperation zum Ziel, der über das World Wide Web zugänglich ist. Dafür wurde ein technischer Prototyp entwickelt, der spezifische Funktionalitäten besitzt und im Laufe des Projektes durch die teilnehmenden Bibliotheken getestet und von OCLC weiterentwickelt wurde. Wesentliche Funktionen dieser über einen normalen Web-Browser zu nutzenden Datenbanklösung sind:

– Import- und Exportmöglichkeit für Dublin Core, MARC, RDF/XML,

- Integration von Dublin Core und MARC-Format in einer Datenbank,
- Bearbeitungssicht sowohl für Dublin Core als auch für MARC,
- Integration einer Harvesterkomponente zur automatisierten Erstellung von Metadatenätzen,
- automatische Klassifizierung von DDC,
- automatisierte Sacherschließung,
- Pathfinder-Konzept zur automatisierten Verwaltung fachlicher Linklisten von Internetressourcen,
- automatisierte Linküberprüfung.

Eine der wichtigsten technischen Innovationen aus bibliothekarischer Sicht ist die Entwicklung einer Harvesterkomponente für die Erstellung von Metadaten, d.h. die (teil-)automatische Katalogisierung von Internetressourcen. Bei der Katalogisierung einer Internetressource wird durch den Bearbeiter zunächst einmal nur die URL des Dokuments oder des thematischen Sites angegeben. Danach wird durch das CORC-System ein Gatherer zu dieser Adresse geschickt, der daraufhin automatisch einen Metadatenatz erstellt. Für englischsprachige Ressourcen wird dabei zugleich die automatisierte Vergabe von DDC-Notationen durchgeführt. Dieser Metadatenatz muß anschließend vom Bearbeiter auf seine Richtigkeit überprüft, entsprechend korrigiert und ergänzt werden.

Für die SUB Göttingen, die neben der DDB Frankfurt der einzige aktive deutsche Teilnehmer am CORC-Projekt war, ergaben sich noch weitere Gründe für die Beteiligung am CORC-Projekt. Einer der wichtigsten war das Testen und Evaluieren der (teil-)automatisierten Erschließung von Internetressourcen ausgewählter, evaluierter Adressbereiche über die „*Harvesting Technology*“, wie sie auch im neu begonnenen Projekt „*Virtuelle Fachbibliothek angloamerikanischer Kulturraum*“ entworfen wurde („Flexible harvesting of resources, assisted classification and subject heading, automatic keyword extraction“). Diese technisch innovativen Konzepte des CORC-Projektes dokumentieren, daß die Projektziele weit über eine klassische bibliothekarische Katalogisierungskooperation digitaler Ressourcen hinausgehen.

Die Teilnahme am Projekt war auch ein Test für Möglichkeiten und Aufwand einer stärkeren Integration einer deutschen Bibliothek in internationalen Konsortien und Verbänden sowie für den Einsatz und die Teilnahme an der Entwicklung internationaler bibliothekarischer Standards. Die internationale Kooperation sollte auch eine bessere Prüfung neuer technischer Methoden und Instrumente und deren Vorteile ermöglichen.

1 <<http://www.oclc.org/>>.

2 Stand Oktober 2002.

3 Das Projekt lief von 1991 bis 1996.

4 <<http://dublincore.org/>>.

5 Später beteiligten sich weitere deutsche Bibliotheken am Projekt, so Die Deutsche Bibliothek, die UB Regensburg, das Bibliothekservice-Zentrum Baden-Württemberg.

6 Zu diesem Zeitpunkt waren 489 Bibliotheken und Institutionen Projektteilnehmer.

### 3 Projektablauf

#### 3.1 Eigene Arbeiten

Das CORC Projekt bietet die Möglichkeit, Ressourcen mit unterschiedlichen Regelwerken bzw. Metadatenformaten zu beschreiben und die so erstellten Metadaten über verschiedene Formate wie MARC 21, Dublin Core etc. über eine Schnittstelle zu recherchieren und herunterzuladen. Dabei können die Metadaten im MARC 21 oder Dublin Core Format eingegeben werden. Für die Zukunft ist die Implementierung weiterer Formate wie GILS, TEI etc. geplant.

Das Innovative an CORC ist, daß durch die Eingabe der URL die Erstellung der Metadaten (z.B. bibliographische Beschreibung, Sacherschließung) durch verschiedene Tools unterstützt wird. Das CORC System versucht dabei, Metadaten aus der HTML Datei zu harvesten (einzusammeln). Folgende verschiedene Tools stehen bei der Erstellung eines Metadatensatzes zur Verfügung:

- Mantis Tool<sup>7</sup> zur automatisierten Erstellung von bibliographischen Metadaten (Titel, Autor etc.),
- Scorpion Tool<sup>8</sup> für die automatische Sacherschließung nach der Dewey Decimal Classification (DDC),
- WordSmith Tool<sup>9</sup> für die automatisierte Sacherschließung für freie Schlagworte (keywords).

Daneben besteht die Möglichkeit, anhand der gespiegelten Library of Congress Authority Records Personen- und Körperschafts-Normdaten zu kontrollieren und einzubinden.

Das CORC System besteht insgesamt aus folgenden verschiedenen Datenbanksystemen:

- Resource Record Database für die Erstellung von Metadatensätzen (**record creation**),
- Authority Datenbank (gespiegelt von der Library of Congress),
- **Pathfinder** Datenbank zur Generierung von spezifischen Linklisten (s.u.) und damit zur fachlichen Verwaltung von Internetressourcen,
- **DDC** Datenbank mit verschiedenen „search and browse“ Funktionen,
- WordSmith Tool<sup>10</sup> für die automatisierte Sacherschließung für freie Schlagworte (keywords).

Die Teilnahme der SUB Göttingen an dem international ausgerichtetem CORC Projekt hatte mehrere Gründe. Zum einem werden in dem noch laufenden Metadatenprojekt<sup>11</sup> und dem abgeschlossenen SSG-FI Projekt<sup>12</sup> fundierte Erfahrungen mit Metadaten und Katalogisierung von Online-Ressourcen gesammelt, die in die Projekt-Kooperation eingebracht werden konnten. Auf der anderen Seite erscheint gerade bei der Erfassung von Internet-Ressourcen die internationale Kooperation dringend erforderlich; CORC bietet Chancen, das Internet-Katalogisierungssystem für Bibliotheken weltweit zu werden, so daß eine frühzeitige Teilnahme nützlich erschien.

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

Als Beitrag der SUB Göttingen in beide Datenbanken wurden folgende Datenpools nachgenutzt bzw. für Testzwecke (s.u.) verwendet:

#### 3.1.1 Nachnutzung der SSG-FI Guides für „record creation“

Das SSG-FI Projekt (**SonderSammelGebiets-FachInformationen** im Internet) startete an der SUB Göttingen im Jahre 1996. Ziel war die Entwicklung eines Konzeptes, daß die Katalogisierung von fachrelevanten Internetressourcen und deren strukturierte Bereitstellung ermöglicht. Das SSG-FI Konzept basierte dabei schon als eine der ersten Anwendungen auf dem Metadatenkonzept von Dublin Core<sup>13</sup>. Das zugrunde liegende Datenbanksystem basiert auf Allegro<sup>14</sup>, einem für Import- und Exportroutinen sehr flexiblen System. Mittels eines Avanti-Servers als http-Schnittstelle werden die Metadaten dynamisch in HTML-Seiten generiert. Der strukturierte Zugang erfolgt anhand eines Klassifikationssystems (Subject Catalog) oder durch die Auswahl nach Dokumenttypen wie z.B. Universitäten, Datenbanken, Journals (Source Type Catalog).

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.). Auch im seit Januar 2000 laufendem RENARDUS-Projekt<sup>15</sup> werden die SSG-FI Guides der SUB Göttingen eingebunden.

In Göttingen werden zur Zeit vier Fachinformationsführer für folgende Fachgebiete angeboten: Geo-Guide<sup>16</sup> (Earth Sciences, Geography, Thematic Maps, Mining), MathGuide<sup>17</sup> (Mathematics) und für den anglo-amerikanischen Kulturraum der History Guide<sup>18</sup> sowie der Anglistik Guide<sup>19</sup>. Ein Fachinformationsführer Forestry Guide wird zur Zeit vorbereitet, ein weiterer für Buch- und Bibliothekswesen ist in Planung.

Insgesamt stehen ca. 6 700 Datensätze zur Nachnutzung für „**Record Creation**“ zur Offline-Einspielung bereit. Dazu war es erforderlich, ein Mapping zwischen dem SSG-FI Format, das Dublin Core basiert ist, und dem MARC 21 Format<sup>20</sup> 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. Auch wenn Dublin Core auf dem besten Wege ist, internationaler Standard zu werden (vgl. Europäische Normungsinstitution CEN<sup>21</sup> und National Information Standards Organization mit dem

7 Mantis Homepage, <<http://orc.rsch.oclc.org:6464/>>.

8 The Scorpion Project, <<http://orc.rsch.oclc.org:6109/>>.

9 The WordSmith Project, <<http://orc.rsch.oclc.org:5061/>>.

10 The WordSmith Project, <<http://orc.rsch.oclc.org:5061/>>.

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

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

13 Dublin Core Metadata Initiative (DCMI), <<http://dublincore.org/>>.

14 Allegro-C, Bibliothekarisches Datenbanksystem der Universitätsbibliothek Braunschweig, <<http://www.biblio.tu-bs.de/allegro/>>.

15 RENARDUS, <<http://www.renardus.org/>>.

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

17 MathGuide Homepage, <<http://www.MathGuide.de/>>.

18 History Guide, <<http://www.HistoryGuide.de/>>.

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

20 MARC Standards, <<http://www.loc.gov/marc/marc.html>>.

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

Z39.85 draft<sup>22</sup>), so gibt es doch eine Reihe von Metadaten, die sich nur über ein „grobes“ Crosswalk auf Dublin Core abbilden lassen. So ist z.B. die komplette Agent-Diskussion bei Dublin Core<sup>23</sup> seit geraumer Zeit im Gange und es gibt bisher noch keine Empfehlungen darüber, wie z.B. Personennamen und Körperschaften inklusive Zusatzinformationen (z.B. Email-Adresse) angesetzt werden können. Der Versuch, ein sogenanntes DC-Library Application Profile<sup>24</sup> zu entwickeln, könnte neuen Schwung in die Diskussion bringen. Zum anderen läßt 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.

Eine aktualisierte Fassung des MARC/Dublin Core Crosswalks ist seit Februar 2001 veröffentlicht<sup>25</sup>. Dementsprechend ist das Mapping zwischen dem SSG-FI Format und dem MARC Format überarbeitet worden. Es wurde mit Rebecca Günther von der Library of Congress abgestimmt. Für das Mapping sind in allen SSG-FI Guides die Kategorien Author/Editor bzw. Publisher/Distributor nach normierten, AACR2 basierten Katalogisierungsregeln überarbeitet worden. Mehrere tausend Datensätze sind nach den neuen SSG-FI Regeln nach folgender Syntax überarbeitet worden:

[Verfasserangabe]\_Nachname,\_Vorname\_<E-Mail>\_(Zusatzinformationen)

Die Verfasserangabe ist optional und kann z.B. aus *Editor in Chief, Founding Editor, Advisory Editor, Project Leader* etc. bestehen. Bei mehreren Vornamen wird auf die Syntax **Nachname,\_Vorname\_Vorname** und bei mehreren Nachnamen auf die Syntax **Nachname\_Nachname,\_Vorname** zurückgegriffen. Damit ist es jederzeit möglich, einzelne Bausteine des Namens getrennt aus der Allegro Datenbank in entsprechende Felder eines anderen Formats zu exportieren. Akademische Titel werden nicht mitkatalogisiert. Bei Namen mit Präfixen wird nach den Regeln von AACR2<sup>26</sup> gearbeitet, wie folgende Beispiele verdeutlichen sollen:

#### Englisch:

Das Präfix wird an den Anfang des Namens gestellt:

Beispiel: VanBuren,\_Martin

#### Französisch:

Wenn das Präfix ein Artikel oder eine Verschmelzung von Artikel und Präposition ist, wird es an den Anfang des Namens gestellt.

Beispiel: LeRouge,\_Gustave

In allen anderen Fällen wird das Präfix nach dem Vornamen aufgeführt.

Beispiel: Musset,\_Alfred de

Die Angabe über die Email Adresse ist optional, es wird bei SSG-FI aber versucht, sie immer mit zu erheben. Dies ist besonders hilfreich, wenn die HTML Seite über die katalogisierte URL nicht mehr aufzufinden ist. Mittels Email-Kontakt konnte sie auf diese Weise schon das ein oder andere Mal wieder gefunden werden. An Zusatzinformationen können noch z.B. Arbeits- oder Forschungsort des Autors (Institut, Universität, Stadt

und Land) aufgeführt werden, falls diese nicht identisch sind mit den Angaben unter Publisher.

Mit der Überarbeitung des SSG-FI Formates und dem Mapping nach MARC 21 stehen der SUB Göttingen mehrere tausend Dokumente für die offline Einspielung in das CORC System zur Verfügung. Für die Zukunft sind jedoch noch einige Workflow-Routinen zu entwickeln, die ein Update der Metadaten aus dem SSG-FI Datenbanken in das CORC System erleichtern. Zwar zeigt das CORC System nicht mehr gültige URLs/Links an, jedoch hat die Pflege und Aktualisierung der Metadaten in den SSG-FI Guides natürlich Priorität und sollte auch aus diesen in andere Systeme erfolgen.

Für die Evaluierung des CORC Systems, besonders im Hinblick auf die Bedürfnisse der deutschen Bibliotheken (deutschsprachige WWW-Seiten, vorwiegend Dateien, deren Inhalt im PDF-Format vorliegt etc.) sind folgende Dokumente herangezogen worden:

#### 3.1.2 Diverse Dokumente vom Web-Doc Server der SUB Göttingen

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 Harvestmethode erfaßt. Hierbei handelt es sich vielfach um deutsche Texte die, wenn überhaupt, nur über „Meta-Tags“<sup>27</sup> verfügen. Sie sind dadurch gut für die Evaluierung des CORC-Systems geeignet. Auch liegen diese Dokumente vorwiegend als PDF-Dateien vor.

#### 3.1.3 Digitalisate von Reiseliteratur zu Nordamerika (offline Lieferung)

Das Göttinger Digitalisierungs Zentrum (GDZ)<sup>28</sup> hat Digitalisierungen aus der bedeutenden Itineraria-Sammlung der SUB für die Offline-Einspielung in das CORC-System bereitgestellt. Eine abschließende Lieferung erfolgte im November 1999.

#### 3.1.4 Ausgewählte HTML Seiten

Für die Evaluierung des CORC-Systems wurden unterschiedlichste Arten von Dokumenten, speziell HTML-Dokumente ausgewählt. Als Auswahlkriterien galten dabei folgende Aspekte: unterschiedliche Sprache, Vorliegen von (DC-)Metadaten, von Metadaten und „Meta-Tags“ oder nur von „Meta-Tags“ bzw. das Fehlen sowohl

22 Draft Standard Z39.85, <<http://www.niso.org/Z3985.html>>.

23 DCMI Agents Working Group, <<http://www.dublincore.org/groups/agents/>>.

24 DCMI Libraries Working Group, vgl. unter Milestone den Punkt „Develop a library application profile“, <<http://www.dublincore.org/groups/libraries/>>.

25 MARC to Dublin Core Crosswalk, <<http://lcweb.loc.gov/marc/marc2dc.html>>.

26 Siehe auch: AACR2 German Translation Project, <<http://www.loc.gov/loc/german/AACR2/AACR2translation.html>>.

27 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="...">).

28 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)>).

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.

Mit Hilfe dieser Auswahl an Dokumenten sind die verschiedenen Harvesting Tools, die im CORC System zur Unterstützung bei der Erstellung von Metadatenätzen zur Verfügung stehen, systematisch evaluiert worden. Dabei galt es, den zeitlichen Arbeitsaufwand bei der Nachbesserung und/oder Korrektur der Titelaufnahmen bezüglich aller relevanten Felder wie Titel, Autor, Beschreibung (Mantis Tool), Schlagworte (WordSmith Tool) und der Klassifikation (Scorpion Tool) abzuschätzen und so insgesamt die Unterstützung durch die teilautomatische Erschließung zu testen. Es wurden dabei auch speziell europäische bzw. deutsche Dokumente berücksichtigt, um Empfehlungen für eine weitere zukünftige Zusammenarbeit mit CORC aussprechen zu können.

### 3.1.5 *Nachnutzung der SSG-FI Guides für die automatische Erzeugung von fachspezifischen Pathfindern*

Bei der automatischen Erstellung von zahlreichen „Pathfindern“ kann das SSG-FI System ebenfalls nachgenutzt werden. 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.

Durch die Fachklassifikation werden aus der Allegro-Datenbank für jeden Guide sogenannte kleinere Subject Catalogs erzeugt, die sich hervorragend als Ausgangsbasis für die *vollautomatische* Erzeugung von Pathfindern eignen. Das dafür nötige Allegro-Script ist erstellt worden, die ersten Testläufe liefen sehr zufriedenstellend ab. Letzte Ungereimheiten konnten bei dem 2. CORC-Participation Meeting im November mit dem dafür zuständigen Entwicklungsteam geklärt werden. Zur Zeit sind in der CORC Pathfinder Datenbank ca. 100 Pathfinder automatisch zu speziellen Themen der Mathematik und der Geowissenschaften aus den SSG-FI Guides MathGuide und Geo-Guide erstellt worden.

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 ohne weiteres möglich und könnte mit in die Beschreibung als Zusatzinformation übernommen werden.

Die Pathfinder werden im XML/HTML Format ausgegeben und können auch für Webseiten 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 Veränderung (Einbinden von Style Sheets etc.) kopiert werden. Zum Herunterladen 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 Herunterladen 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. die URL eines Links verändert und ist der Pathfinder diesbezüglich aktualisiert worden, so ist man mit dieser Download-Möglichkeit ebenfalls auf dem neuesten Stand.

Für die SSG-FI Guides bietet CORC damit eine hervorragende Möglichkeit, die Guides besonders in den Bibliotheken der USA und damit den dortigen Nutzern bekannt zu machen.

## 3.2 CORC Anwendertreffen

Unterstützt wurden die eigenen Arbeiten durch die Anwendertreffen, auf denen Erfahrungen ausgetauscht und Probleme besprochen werden konnten. OCLC veranstaltete für die teilnehmenden Bibliotheken mindestens ein jährliches Anwendertreffen mit der Möglichkeit, Erfahrungen auszutauschen und aktiv an der Weiterentwicklung des CORC Systems mitzuarbeiten.

Das erste Anwendertreffen fand im April 1999 in Dublin/Ohio, dem Hauptsitz von OCLC statt. Die 90 Teilnehmer kamen hauptsächlich aus den USA, daneben waren noch Anwender aus Deutschland<sup>29</sup>, Grossbritannien, Mexiko und Taiwan vertreten. Zu diesem Zeitpunkt nahmen ca. 30 akademische Institutionen, 3 Nationalbibliotheken und ca. 5 OCLC Networks aus den USA an dem Projekt teil.

Ein weiteres Treffen fand im November 1999 ebenfalls in Dublin/Ohio statt. Auch diesmal waren knapp 100 Teilnehmer vertreten, auch von dem Institute of Information Science in Maribor (Slowenien), von der University of Strathclyde (Schottland) und vom Monterrey Institute of Technology (Mexico). Fast doppelt so viele Institutionen nahmen am CORC Projekt teil. Die Library of Congress war zum ersten Mal als Mitglied des CORC Projektes vertreten.

Während der ALA (American Librarian Association) Konferenz in Chicago im Juli 2000 ergab sich für die SUB nochmals die Möglichkeit, an einem CORC Anwendertreffen teilzunehmen.

Waren am Anfang der Treffen die wenigsten Bibliotheken mit Dublin Core und der routinierten Anwendung und Integration in den Bibliotheksalltag vertraut, zeigte sich bei dem letzten Treffen, daß einige Bibliotheken<sup>30</sup> zumindest die Online Dokumente mit Dublin Core und nicht mit MARC 21 katalogisierten. Mehrere Work-Flow Analysen der Integration von CORC in das bibliothekarische Alltagsgeschäft und Erfahrungsberichte verdeutlichten den Trend, daß Dublin Core durch das CORC Projekt an allgemeiner Akzeptanz gewonnen hat.

Generell verliefen die Anwendertreffen nach folgendem Schema: Als erstes wurden allgemeine Informationen zu Daten/Zahlen und Strategiepläne wie z.B. Anzahl der

29 Aus Deutschland waren die DDB (Frankfurt/Main) und die SUB Göttingen vertreten.

30 University of Chicago, University of Michigan und University of Vermont.

Mitglieder in CORC, Anzahl der Katalogisate und Pathfinder im CORC System, Informationen zu Neuinstallationen, Updates (z.B. von Normdaten wie die der Library of Congress Name Authority Files) etc., Implementierung von neuen Standards (z.B. Update des Crosswalks von MARC zu Dublin Core) etc. vorgestellt. Danach trafen sich die Teilnehmer in kleineren Gruppen zu speziellen Aspekten des CORC Systems wie z.B. „Record Creation“, „Authority Files“, „Collection Development“, „Pathfinder“ etc., um zukünftige Entwicklungen mit zu gestalten. Besondere Probleme konnten in diesen kleineren Kreisen („breakout session“) diskutiert werden ebenso wie Wünsche für die zukünftige Gestaltung des CORC Systems. Diese Ergebnisse und Anregungen sind in sogenannten „group reports“ in einer „panel discussion“ zusammengetragen worden und führten gegen Ende eines jeden CORC Anwendertreffens zu einer Taskliste, die von dem CORC Entwicklungsteam vorrangig bearbeitet worden ist. Einige Punkte tauchten dabei wiederholt auf und sind auch heute nicht gelöst.

Daneben gab es immer wieder Vorträge von Mitgliedsbibliotheken über ihre Erfahrungen mit CORC wie z.B. Mitarbeiterschulung in Dublin Core, Katalogisierung mit Dublin Core, Erarbeitung von Arbeitsroutinen zur Integration von CORC in den Alltag etc.

Im folgenden werden einige Punkte aufgelistet, die zu Beginn der Anwendertreffen bei den Partnern eine große Rolle gespielt haben, vorrangig bearbeitet werden sollten und inzwischen gelöst sind:

#### Allgemeine Punkte:

- Userguide mit Beispielen,
- Einführungsseiten für Partner, die mit CORC gerade angefangen haben (Tutorial),
- bessere Dokumentation,
- Klärung der Beziehung zwischen CORC und World-Cat,
- Preispolitik,
- optische Verbesserung der CORC Benutzeroberfläche,
- einfacheres Editieren bei „record creation“,
- Diskussion um „constant data“, z.B. das lokale Anlegen eines sogenannten Templates in dem schon bestimmte Felder vorausgefüllt sind.

#### Spezielle Punkte:

- Einbindung von Dublin Core 1.1, Anpassung an neue Mappings von MARC 21 zu Dublin Core simple und qualified,
- Erweiterungen im Exportieren von Datensätzen,
- Erweiterung der DDC (direktes Browsing), Mapping zu anderen Klassifikationssystemen/kontrolliertem Vokabular,
- Pflege der URLs (d.h. z.B. Hinweis auf nicht mehr gültige URLs),
- Dublettenkontrolle, auch bereits vor der Neukatalogisierung eines online Dokumentes,
- Normdaten-Linking z.B. zum LoC Name Authority File,
- bezüglich der Pathfinder: URL-Check, Datenbankzugriff schneller und sicherer gestalten, Duplikate herausfiltern, Anlegen von „constant data“ Templates auch hier, Preispolitik überdenken.

Seit Juli 2000 sind die CORC Datensätze in WorldCat integriert. Ein Update der Authority Files von der Library of Congress erfolgt nächtlich.

Weitere geplante Verbesserungen bzw. Implementierungen umfassen folgende Punkte:

- Unicode-Einführung, Diakritika,
- Z39.50 Schnittstelle (zumindest im Test),
- CORC soll in Zukunft weitere Formate unterstützen wie z.B. TDI, DOI, GILS, TEI, EAD etc.,
- das Harvesting von PDF-Dokumenten soll möglich sein,
- die Möglichkeit, mehrere sogenannte *constant data* Templates anlegen und benutzen zu können,
- sobald weitere Metadatenformate in CORC implementiert sind, soll das Exportieren von Katalogisaten in andere Formate als z.B. MARC 21 möglich sein (one record, many views),
- Leitfaden bezüglich der Politik von „collection development“ (z.B. welche Dokumente werden warum katalogisiert, wann werden die Katalogisate gelöscht, wie und nach welchen Kriterien überarbeitet etc.). In diesem Zusammenhang ist die Frage nach einem „citation index“ für Internet Ressourcen wichtig. Kann und wenn ja, wie könnte ein solcher Index definiert werden?

Auf dem letzten Treffen im Juli in Chicago wurde beschlossen, eine sogenannte „CORC User Group“ zu bilden. Den Vorsitz dieser Gruppe sollte jemand aus den CORC Anwendern übernehmen (mittels online Wahl<sup>31</sup>), der als Vermittler zwischen den CORC Anwendern und dem CORC Entwicklungsteam fungierte. Nachdem CORC aus dem Projektstatus entlassen wurde, war auf diese Weise eine intensive Kommunikation zwischen den Partnern einerseits und den Partnern und dem OCLC bzw. CORC Team andererseits gewährleistet, und bestimmte Themen oder Probleme konnte für zukünftige Treffen besser vorbereitet werden. Vorsitzende der User Group<sup>32</sup> ist derzeit Jackie Shieh von der University Library of Michigan.

Zur Zeit sind ca. 673 259 Katalogisate und ca. 2 000 Pathfinder in CORC katalogisiert (Stand Juni 2002).

## 4 Projektarbeiten in Göttingen

### 4.1 Evaluierung der Harvesting Tools des CORC-Systems

Für die Evaluierung der Harvesting Tools des CORC-Systems wurden die in CORC aufgenommenen Katalogisate nach sachlichen (Beschreibung, Schlagworte) und formalen Erschließungskriterien getrennt evaluiert. Folgende Vorgehensweise war für die Evaluierung der verschiedenen Dokumente vereinbart worden:

- ausschließliche Anwendung der automatischen Harvest-Methode zum Gathern der Metadaten, d.h. die Erschließung der elektronischen Ressourcen erfolgt im MARC Format durch die Eingabe der URL;
- Evaluierung der Ergebnisse, d.h. in einer gesonderten Access-Datenbank wurde getrennt nach formalen und sachlichen Kriterien der zeitliche Nachbear-

31 CORC User Group Voting Form, <<http://www.sil.si.edu/staff/corc-user-group/corcvote.htm>>.

32 Seit Januar 2002 nennt sich die Gruppe Connexion User Group <<http://www.sil.si.edu/staff/CORC-User-Group/content.htm>>.

Automatische Erschließung (1=sehr gut):											total	Nachbesserungen bei:					Zeitaufwand
Sprache	Title	Aut./Publ.	Formal	Keyw.	DDC	Sach Descr.	Metadata				Title	Aut./Publ.	Keyw.	Descr.	DDC	Nachbess.	
6																	
7																	
8																	
9																	
10																	
11																	
12																	
44	en	4	5	4,5	4	3	3,5	5	21	nv	4,2		nv		mittel	mittel	
45	en	1	1	1	2	5	3,5	1	10	Dublin Core	2				alles	mittel	
46	en	5	5	5	2	3	2,5	5	20	nv	4		nv		mittel	mittel	
47	en	5	4	4,5	3	4	3,5	5	21	nv	4,2				mittel	mittel	
48	de	1	5	3	5	5	5	5	21	nv	4,2	Um	Um		alles	mittel	
49	en	2	4	3	5	1	3	5	17	nv	3,4					wenig	
50	de	2		2	5	1	3	5	13	nv	2,6	Um					wenig
51	en	4		4	5	2	3,5	5	16	nv	3,2				wenig	mittel	
52	de	4	5	4,5	5	5	5	5	24	nv	4,8	Um	Um		alles	mittel	
53	en	5	5	5	2	2	2	5	19	Metadata	3,8		nv			mittel	
54	en	1	2	1,5	2	2	2	5	12	Dublin Core	2,4			nv		mittel	
55	en	1	2	1,5	2	3	2,5	5	13	Dublin Core	2,6				wenig	mittel	
56	no	1	1	1	2	3	2,5	5	12	Metadata	2,4				wenig	wenig	
57	en	2	5	3,5	2	1	1,5	3	13	Metadata	2,6		E			wenig	
58	en	1	1	1	1	1	1	3	7	Dublin Core	1,4					wenig	
59	en	1	4	2,5	3	2	2,5	5	15	Metadata	3		E		wenig	wenig	
60	en	1	2	1,5	3	2	2,5	3	11	Dublin Core	2,2					wenig	
61	en	1	4	2,5	3	3	3	5	16	Metadata	3,2		E		wenig	mittel	
62	en	1	1	1	3	5	4	5	15	Dublin Core	3				alles	mittel	

Tab. 1 Übersicht über den Nachbearbeitungsaufwand jeder in CORC generierten Ressource

beitungsaufwand für jedes Metadatenfeld erhoben und eine „Gesamtnote“ für diesen Metadatenatz generiert. Dabei wurde auch die Beschaffenheit der zu harvestenden Seite festgehalten, d.h. ob es sich um eine PDF- oder HTML-Datei gehandelt hat, welche Art von Metadaten im Header vom Source Code („meta-tags, DC Metadaten etc.) enthalten waren und in welcher Sprache das Dokument vorlag. Auch die Implementierung von Java-Skript konnte sich störend auf die Harvesting Prozesse auswirken, so daß dieser Fall ebenfalls vermerkt wurde;

- aus diesen Ergebnissen konnten Trends abgeleitet werden, unter welchen (optimalen) Bedingungen die jeweiligen Harvesting Tools am besten funktionierten. Die Charakterisierung des zeitlichen Nachbearbeitungsaufwandes erfolgte mittels einer Skala von 1 bis 6, wobei 1 bedeutete, daß in diesem Metadatenfeld weder etwas korrigiert noch hinzugefügt werden mußte. Dies bezieht z.B. auch Fragen 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? Tab. 1 gibt in der Zusammenfassung (letzte Spalte) beispielhaft wieder, wie hoch der Nachbearbeitungsaufwand für die Erstellung jeder Ressource in CORC war. Gesondert wurde auch festgehalten, worin jeweils das Problem für Korrekturen oder Ergänzungen bestand, z.B.

mussten bei deutschsprachigen Seiten immer die Umlaute korrigiert werden, manchmal sind durch das WordSmith Tool gar keine Schlagwörter geharvestet worden, so daß diese manuell nachgetragen werden mußten, etc.

Tab. 2 zeigt den Zusammenhang zwischen dem zeitlichen Nachbearbeitungsaufwand und dem Vorhandensein von Metadaten auf einer Internetseite auf.

Generell läßt sich sagen, daß das Vorhandensein von Dublin Core Metadaten, integriert im Source Code im Body einer HTML Seite, den zeitlichen Aufwand pro geharvestete Seite erheblich reduziert. Während die Gesamtbewertung bei den Dokumenten ohne Metadaten zwischen 3 und 4 liegt (vgl. Spalte Evaluation in Tab. 2 obere Hälfte), so liegen die Noten der Gesamtbewertung bei Dokumenten, die über Dublin Core Metadaten verfügen, im Durchschnitt zwischen 2 und 3 (vgl. Spalte Evaluation in Tab. 2 untere Hälfte). Insgesamt sind etwas über 100 Dokumente auf ihre „Harvesting-Tauglichkeit“ evaluiert worden.

#### 4.2 Crosswalk SSG-FI zu MARC 21 bzw. SSG-FI zu Dublin Core

Bei dem Austausch von Metadaten ist es sinnvoll, sich entweder auf ein gemeinsames Metadatenformat mit genauer Semantik und Syntax für jedes Element zu einigen, wie es z.B. bei Renardus und der Virtuellen

Titel	Sprache	Arbeitsaufwan	Evaluation	Metadaten
Die Ausstattung von Hochschulbibliotheken mit lokalen Bibliothekssys	de	mittel	4,2	nicht vorhanden
Jahresforschungsbericht Universität Göttingen	de	mittel	3,6	nicht vorhanden
Weissbuch zur Foerderung von Frauen	de	mittel	3,6	nicht vorhanden
Zen and the Art of the Internet	de	wenig	3,6	nicht vorhanden
THE DESCRIPTION AND INDEXING OF IMAGES:	en	mittel	4	nicht vorhanden
International Conference on the Principles and Future Development of	en	mittel	3,6	nicht vorhanden
Guidelines for Robot Writers	en	mittel	3	nicht vorhanden
Lizenzierung digitaler Ressourcen	de	mittel	3,8	nicht vorhanden
Licensing Principles	en	wenig	3,2	nicht vorhanden
Project Reuse	en	mittel	4	nicht vorhanden
Proceedings of the First Summer Galway Topology Colloquium	en	mittel	3	nicht vorhanden
Proceedings of the Second Galway Topology Colloquium at Oxford	en	mittel	3	nicht vorhanden
The Complete Works of William Shakespeare	en	mittel	4,2	nicht vorhanden

  

Titel	Sprache	Arbeitsaufwan	Evaluation	Metadaten
Digitising History	en	mittel	2	Dublin Core
Metadata	no	wenig	2,4	Dublin Core
Metadata: Enabling the Internet	en	mittel	2,4	Dublin Core
Metadata for the masses	en	mittel	2,6	Dublin Core
Dublin Core Management	en	wenig	1,4	Dublin Core
Handling MARC with Perl	en	wenig	2,2	Dublin Core
CrossROADS and Interoperability	en	mittel	3	Dublin Core
What is... RDF?	en	mittel	3	Dublin Core
An Approach for Packaging Dublin Core Metadata in HTML 2.0	en	wenig	2,4	Dublin Core
"Warwick framework and Dublin core set provide a comprehensive	en	wenig	2,6	Dublin Core
Guidelines for Embedding Metadata in State Library's Web Pages usir	en	mittel	2	Dublin Core

Tab. 2 Zusammenhang zwischen der „Güte“ der CORC tools bei der automatischen Generierung von Ressourcen und Metadaten

Fachbibliothek<sup>33</sup> entwickelt wurde bzw. wird, oder ein Mapping von einem Metadatenformat zu einem anderen zu erarbeiten.

Für die Offline-Lieferung der SSG-FI Metadaten in das CORC-System mußte zuerst ein Mapping von dem SSG-FI Format entweder zu Dublin Core oder zu MARC 21 erstellt werden. Das CORC Team hat sich dafür entschieden, ein Crosswalk nach MARC 21 zu entwickeln. Obwohl das Dublin Core Format<sup>34</sup> über 40 Qualifier<sup>35</sup> zu den 15 Grundelementen anbietet, so reicht das Angebot nicht aus, um die Vielzahl und Vielfalt an SSG-FI Metadaten-elementen auf DC abzubilden. Für einige SSG-FI spezifische Metadaten-elemente gibt es (noch) keine Entsprechung in Dublin Core. Dies trifft vor allem für die Vielzahl an Zusatzinformationen bei den SSG-FI Elementen Author bzw. Editor, Publisher und Distributor. Email Adressen, Dienstadressen, URL etc. können in Dublin Core nicht in Qualifiern, sondern nur ohne Empfehlungen für eine Syntax in die entsprechenden „Agent“ Elemente (DC.Creator, DC.Publisher, DC.Contributor) gemappt werden. Damit möglichst wenig Informationen bei dem Crosswalk von SSG-FI zu einem in dem CORC-System verwendeten Format verloren gehen und das Mapping damit möglichst vollständig und eindeutig ist, wurde dem MARC 21 Format gegenüber dem Dublin Core Format der Vorzug gegeben. Allerdings ist das Mapping für jedes Element mit Hilfe der

Crosswalks von SSG-FI zu Dublin Core<sup>36</sup> und von MARC 21 zu Dublin Core<sup>37</sup> bzw. von Dublin Core zu MARC 21<sup>38</sup> gegengeprüft worden.

Der im Anhang in Auszügen zur Veranschaulichung vorgestellte Crosswalk ist nur für das Mapping von SSG-FI zu MARC 21 gültig und nicht für MARC 21 zu SSG-FI. Im folgenden wird kurz auf einige problematische Bereiche bei der Entwicklung des Crosswalks eingegangen.

#### 4.2.1 Crosswalk SSG-FI zu Dublin Core

Bei dem Mapping von dem Metadatenformat SSG-FI zu Dublin Core ergaben sich folgende Fragestellungen, die z.T. noch nicht vollständig gelöst sind:

33 Renardus <<http://www.renardus.org>>; Virtuelle Fachbibliothek <<http://www.virtuellefachbibliothek.de/>>.

34 Dublin Core Metadata Element Set, Version 1.1, <<http://www.dublincore.org/documents/dces/>>.

35 Dublin Core Qualifiers, <<http://www.dublincore.org/documents/dcmes-qualifiers/>>.

36 Vgl. auch Anhang; das ausführliche Mapping von SSG-FI zu Dublin Core findet sich z.B. auf den Geo-Guide Seiten unter <<http://www.Geo-Guide.de/info/tools/metadata.html>>.

37 MARC to Dublin Core Crosswalk [February, 2001], <<http://lcweb.loc.gov/marc/marc2dc.html>>.

38 Dublin Core/MARC/GILS Crosswalk [2001-03-12], <<http://lcweb.loc.gov/marc/dccross.html>>.



- In SSG-FI wird zwischen „Author“ und „Editor“ unterschieden. Solange die Diskussion in der „Agents Working Group“<sup>39</sup> noch anhält, werden beide Elemente zu DC.Creator gemappt.
- Bei Dublin Core sind detaillierte Angaben zu Creator nicht in eigenen Feldern vorgesehen, dafür aber in MARC 21 in dem Feld „statement of function“ (700 \$e): alle Teilmeldungen zu Creator/Publisher/Contributor können in MARC Kategorien detaillierter in einzelnen Feldern abgebildet werden (z.B. auch Email Adresse in 700 \$g für miscellaneous information nach Aussage von Rebecca Günther oder auch „Publisher Place“ in 260 \$a, Place of publication, distribution, etc.).
- In SSG-FI gibt es ein zusätzliches Feld für die Unterscheidung zwischen online, offline und print. Dafür gibt es in Dublin Core keine Entsprechung, in MARC 21 hingegen schon (Feld 256 \$a, Computer data or programs), allerdings gibt das Feld 007 mit dem Eintrag „cr“ ebenfalls einen Hinweis darauf, daß es sich um Computer-Dateien mit „Remote Access“ handelt.
- In Dublin Core gibt es leider kein Metadatenelement für „General Notes“, d.h. z.B. daß die Informationen über die Evaluierung einer Ressource, wie z.B. in SSG-FI vorhanden, keine Entsprechung in DC haben. Man könnte sie mit zu der Beschreibung (DC.Description oder DC.Description.Abstract) mappen. In MARC 21 gibt es für General Notes ein extra Feld (500 \$a).
- In DC gibt es keine Qualifier zu DC.Rights, um z.B. genaue Aussagen über Copyright, Intellectual property rights etc. treffen zu können. In MARC 21 dagegen können die Metadatenelemente Access und Restriction des SSG-FI Formates nach dem MARC 21 Feld 540 \$a (Definition: Usually, the text of a legal or official statement of restrictions<sup>40</sup>) gemappt werden.
- Der Country Code von SSG-FI kann über Dublin Core nicht abgebildet werden, aber über ein MARC 21 Feld (008/15-17).

#### 4.2.2 Crosswalk SSG-FI zu MARC 21

Auch wenn das detaillierte SSG-FI Metadatenformat besser in MARC 21 abgebildet bzw. gemappt werden kann, so gibt es auch hier noch offene Punkte:

Eine kleine Ungereimtheit bei dem Mapping von SSG-FI nach MARC ergibt sich bei dem Element Last Update. Dieses Element kann mehr oder weniger eindeutig nach Dublin Core gemappt werden, allerdings nicht so eindeutig nach MARC: Nach dem Crosswalk von Dublin Core zu MARC (vom 12.03.2001<sup>41</sup>) wird der Qualifier Modified des DC Elementes Date zum MARC Feld 583 \$d mit \$a=modified gemappt. Während \$a mit dem Eintrag „modified“ sinnvoll erscheint, ist das Unterfeld \$d mit der Definition „A time period that cannot be expressed as a concrete time“ (vgl. Definitionen zu Feld 583<sup>42</sup>) etwas vage und spiegelt nicht den tatsächlichen Sachverhalt, nämlich eine Zeitangabe über das letzte Aktualisierungsdatum einer online Ressource, wider.

## 5 Öffentlichkeitsarbeit für CORC in Deutschland

Die Information anderer deutscher Bibliotheken über das Projekt erfolgte durch Bekanntmachung in der „Inet-

bib“-Mailingliste (22.10.1999) und über die Homepage der SUB Göttingen (Projekte). Eine Mailingliste für die deutschen CORC Anwender/Nutzer wurde im Dezember 1999 eingerichtet.

In Zusammenarbeit mit OCLC wurde am 3. Dezember 1999 in der SUB Göttingen ein Workshop für interessierte deutsche Bibliotheken abgehalten, auf dem das Projekt von OCLC (Nicholas Rawson und William Carney) vorgestellt wurde.

Die deutschen Projektteilnehmer (zu diesem Zeitpunkt DDB Frankfurt, speziell mit Online Dissertationen und Dublin Core Metadaten, und die SUB Göttingen) berichteten über ihre Erfahrungen mit CORC. Die UB Regensburg und das Bibliotheksservice-Zentrum Baden-Württemberg hatten Anträge auf Projektbeteiligung gestellt und stellten ihre Vorhaben kurz vor. Bei der UB Regensburg handelt es sich um die Digitalisierung einer Porträtsammlung, beim BSZ um Online Ressourcen aus dem Hochschulbereich.

In verschiedenen deutschen Bibliothekszeitschriften wurde über das Projekt berichtet (s. Literaturliste).

## 6 Ergebnisse des Projektes und zukünftige Entwicklungen

Die Teilnahme am CORC Projekt hat zu neuen Erfahrungen bei der Katalogisierung von Online Ressourcen geführt, was sowohl die Normierung von Kategorieinhalten (Personenansetzung, Titel usw.) als auch die Nutzung internationaler Standards bei der Erschließung von Online Ressourcen angeht. Unabdingbar sind dabei Kenntnisse entweder von Dublin Core oder MARC/AACR2.

Als nützlich hat sich das Mapping verschiedener Formate erwiesen, wie z.B. Dublin Core zu MARC 21 und vice versa, SSG-FI zu MARC 21, MARC 21 zu PICA, auch im Zusammenhang mit neuen Datenformaten wie RDF und XML zur Verwaltung von Metadaten. Durch das Herunterladen von in CORC erfassten Datensätzen in das PICA-System konnten Export/Import-Routinen in bestehenden Datenbanken getestet werden.

Der überwiegende Teil der Göttinger Datensätze wurde auch evaluiert (siehe Kapitel 4.1). Da es sich in der Regel um PDF-Dateien ohne Metadaten handelte, konnte das Mantis-Programm zur halbautomatischen Erstellung von Titelaufnahmen leider nur rudimentär ausgenutzt werden.

Aus dem Itineraria-Digitalisierungsprojekt wurden Datensätze offline eingespielt. Daneben wurden auch Pathfinder aus den Bereichen Geowissenschaften und Mathematik erstellt. Umgekehrt ist es nicht gelungen, in CORC vorhandene Pathfinder für die SSG-FI-Guides zu nutzen, da die Qualität der CORC Pathfinder damals

39 DCMI Agents Working Group, <<http://www.dublincore.org/groups/agents/>>.

40 MARC 21 Concise Bibliographic: Note Fields (Part 2: 536-58X) – 540, <<http://lcweb.loc.gov/marc/bibliographic/ecbdnot2.html#mrcb540>>.

41 Dublin Core/MARC/GILS Crosswalk, <<http://lcweb.loc.gov/marc/dccross.html>>.

42 MARC 21 Concise Bibliographic: Note Fields (Part 2: 536-58X) – 583, <<http://lcweb.loc.gov/marc/bibliographic/ecbdnot2.html#mrcb583>>.

nicht dem SSG-FI Qualitätsstandard entsprach und Personalressourcen zur qualitativ höherwertigen Aufbereitung fehlten.

Durch die Beteiligung am CORC Projekt konnte die SUB Göttingen als deutsche Bibliothek das internationale Umfeld für die Katalogisierung von elektronischen Ressourcen besser kennenlernen. Im Moment ist es für die Bibliotheken notwendig, entweder Dublin Core anzuwenden oder im MARC 21 Format nach den amerikanischen Katalogisierungsregeln (AACR2) zu katalogisieren. Weder Dublin Core noch AACR2 gehören derzeit in ausreichender Form zu den Ausbildungscurricula in Deutschland, Kenntnisse beider Datenstrukturen sind in den Bibliotheken selbst nur bei wenigen Spezialisten anzutreffen. Eine Implementierung von Dublin Core ist in den Verbänden bisher nicht Standard. Bei Dublin Core muß zudem die Verknüpfung zu Normdaten noch geklärt werden. Dies wird sicher Die Deutsche Bibliothek als besonderer Vertreter im Rahmen von Dublin Core befördern.

Als Erschließungsformat hat sich Dublin Core in den USA im Verlaufe des Projektes weiter durchsetzen können, was auch bei den CORC User Meetings deutlich wurde. Es ist damit zu rechnen, daß auch in Deutschland die Anwendung dieses Formates in Zukunft steigen wird, zumindest bei den Hochschulschriften<sup>43</sup> und hier speziell bei den Dissertationen, für die viele Universitäten Eingabeformulare bereitstellen (s.a. Diss-Online Eingabe Formular<sup>44</sup>).

Die Akzeptanz von Dublin Core als internationales web-basiertes Format wird steigen, wenn die Autoren ihren Dokumenten die entsprechenden Metadaten beifügen, wie es jetzt schon bei den Dissertationen geschieht. Die Bereitstellung und Nutzung von einfach zu handhabenden Werkzeugen und Eingabemasken sollte auch im Interesse der Universitäten und Forschungseinrichtungen liegen, um die Anwendung von Metadaten in breiter Form zu fördern und durchzusetzen. Das Dublin Core Mapping wird dann auch für die hiesigen Bibliothekssysteme als Standard gefordert werden. Unabhängig davon wird die steigende Nutzung von Dublin Core die Katalogisierung nach AACR2 und in MARC 21 in absehbarer Zeit nicht verdrängen.

Die fortschreitende Globalisierung geht auch an den Bibliotheken nicht spurlos vorbei. Auf der 5. GBV-Verbundkonferenz am 11./12. September 2001 wurde auf einer Podiumsdiskussion der Frage nachgegangen, ob „Deutschland reif für internationale Zusammenarbeit“ sei<sup>45</sup>. Die Podiumsteilnehmer zeigten sich insgesamt dieser Frage (d.h. AACR2 und MARC 21) gegenüber aufgeschlossen, befürchteten aber Widerstand von Seiten der Anwender (d.h. der Katalogisierer).

Inzwischen hat der Standardisierungsausschuß am 6.12.2001 beschlossen<sup>46</sup>, eine Studie in Auftrag zu geben, um die Pros und Contras sowie Kosten für einen eventuellen Umstieg prüfen zu lassen. In der bibliothekarischen Öffentlichkeit wird seither sehr heftig darüber diskutiert, ob solch ein Umstieg sinnvoll und wünschenswert ist. Eine von Bernhard Eversberg initiierte Umfrage im März/April 2002 brachte, wie nicht anders zu erwarten, eine deutliche Ablehnung des Umstiegs auf amerikanische Regeln zum Ausdruck<sup>47</sup>. Schon Ende der 80er Jahre hatte Helmut Vogt (damals Direktor der SUB Göttingen) für einen Umstieg auf amerikanische Regeln<sup>48</sup> geworben und war auf harsche Ablehnung ge-

stoßen, auch von Seiten anderer deutscher Bibliotheksverbände. Die Haltung der Verbände hat sich gewandelt<sup>49</sup>.

Mit der von Elmar Mittler in Zusammenarbeit mit dem OCLC initiierten REUSE-Studie wurden wichtige Grundlagen für eine sachliche Diskussion der Unterschiede der Regelwerke und deren Nutzung für die praktische Arbeit gegeben<sup>50</sup>. Die Publikation der deutschen Übersetzung der AACR2 wird eine weitere Versachlichung ermöglichen. Schließlich kann die von Elmar Mittler als Präsident von LIBER eingesetzte MARC-Arbeitsgruppe ein europaweit abgestimmtes Vorgehen erleichtern und die Position bei der weiteren Entwicklung von MARC21 und AACR2 als internationale Standards verbessern<sup>51</sup>. Nun heißt es, die Ergebnisse der Studie des Standardisierungsausschusses abzuwarten. In der Zwischenzeit wurde CORC weiterentwickelt und die Datensätze sind in den WorldCat integriert. Juli 2002 hat dieses Online Katalogisierungssystem einen neuen Namen innerhalb der OCLC Katalogdienste erhalten: OCLC Connexion – Integrated Cataloging and Metadata services<sup>52</sup>.

Die Übernahme von PICA durch OCLC wird auch für Europa zu neuen Plattformen für die Katalogisierung von Online Ressourcen führen und vermutlich ähnliche Funktionalitäten bieten können wie CORC. Eine PICA Arbeitsgruppe zur Katalogisierung elektronischer Ressourcen ist bereits aktiv. Erste Ergebnisse sind Ende 2002 zu erwarten.

## 7 Literaturliste

- CORC: New tools and possibilities for cooperative electronic resource description. Ed. by Karen Calhoun and John J. Riemer. New York: Haworth Press, 2000. 217 S.
- Cremer, Monika und Heike Neuroth: Eine Zukunftswerkstatt des Katalogisierens? Das CORC-Projekt: Erfahrungen an der SUB Göttingen. In: BuB 34 (2000) S. 245-248
- 43 Koordinierungsstelle Dissonline Der Deutschen Bibliothek <<http://www.ddb.de/professionell/dissonline.htm>>.
- 44 MyMetaMaker(v 1.31): <<http://physnet.uni-oldenburg.de/services/mmm/>>.
- 45 S. hierzu auch Klaus Franken: Ist Deutschland reif für die internationale Zusammenarbeit? In: Bibliotheksdienst 35 (2001) 10, S. 1342-1343.
- 46 <<http://www.ddb.de/professionell/afs.htm>>.
- 47 B. Eversberg: Fragen für die Studie zur AACR-Einführung <<http://www.biblio.tu-bs.de/allegro/formate/aacr-fr.htm>>.
- 48 Kolloquium zur retrospektiven Katalogisierung in Niedersachsen am 3. und 4. Mai 1988 in der Niedersächsischen Staats- und Universitätsbibliothek Göttingen. Göttingen 1988. (Niedersächsische Staats- und Universitätsbibliothek Göttingen. Bibliotheksrechenzentrum für Niedersachsen. Mitteilungen. Sonderheft 1).
- 49 Vgl. auch Elmar Mittler: Katalogkonversion – Wendepunkt für deutsche Bibliotheken? In: Zeitschrift für Bibliothekswesen und Bibliographie 36 (1989) S. 407.
- 50 Project REUSE. Aligning international cataloging standards <<http://webdoc.sub.gwdg.de/ebook/aw/reuse/index.htm>>.
- 51 Den Vorsitz in der LIBER-MARC-Arbeitsgruppe hat Elisabeth Niggemann übernommen; siehe auch Renate Gömpel, Elisabeth Niggemann: RAK und MAB oder AACR und MARC? In: Zeitschrift für Bibliothekswesen und Bibliographie 49 (2002) S. 3-12.
- 52 Die „CORC documentation“ wurde in die neue „Connexion documentation“ integriert: <<http://www2.oclc.org/connexion/documentation/>>.

Cremer, Monika und Heike Neuroth: Das CORC-Projekt von OCLC an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen. In: Bibliotheksdienst 34 (2000) S. 755-776  
 DDC deutsch: <[http://www.ddb.de/professionell/ddc\\_info.htm](http://www.ddb.de/professionell/ddc_info.htm)>  
 Dublin Core Metadata Initiative (DCMI) <<http://www.dublincore.org/>>  
 Dublin Core Metadata Element Set, Version 1.1: Reference Description [1999-07-02] <<http://www.dublincore.org/documents/dces/>>  
 Dublin Core Qualifiers [2000-07-11] <<http://www.dublincore.org/documents/dcmes-qualifiers/>>  
 DCMI Agents Working Group <<http://www.dublincore.org/groups/agents/>>  
 DCMI Libraries Working Group <<http://www.dublincore.org/groups/libraries/>>  
 Dublin Core/MARC/GILS Crosswalk [2001-03-12] <<http://lcweb.loc.gov/marc/dccross.html>>  
 Godby, C. Jean, Ray R. Reighart: The WordSmith toolkit. In: Journal of library administration 34 (2001) S. 307-326  
 Godby, C. Jean, Ray R. Reighart: The WordSmith indexing system. In: Journal of library administration 34 (2001) S. 375-384  
 Hickey, Thomas B.: CORC – Cooperative Online Resource Catalog. In: Journal of library administration 34 (2001) S. 317-323  
 Hickey, Thomas B., Diane Vizine-Goetz: The role of classification in CORC. In: Journal of library administration 34 (2001) S. 421-430  
 Hsieh-Yee, Ingrid and Michael Smith: The CORC experience: survey of founding libraries. P. 1. In: OCLC Systems & Services 17 (2001) 3, S. 133-140; P. 2. In: OCLC Systems & Services 17 (2001) 4, S. 166-177  
 Journal of Internet cataloging: Special theme issue on CORC. Vol. 4 (2001) N. 1/2 (abstracts: <<http://www.haworthpressinc.com/jic/jic4nr1-2.htm>>  
 MARC to Dublin Core Crosswalk [February, 2001] <<http://lcweb.loc.gov/marc/marc2dc.html>>  
 Medeiros, Norm: Peering over the fortress walls: the metadata invasion begins. In: OCLC Systems & Services 17 (2001) 4, S. 154-156  
 SSG-FI Metadata Concept <<http://www.Geo-Guide.de/info/tools/metadata.html>>

SSG-FI	Dublin Core	MARC21	Indic.	Denotation
City, Country Code	DC.Publisher.Place (open discussion relating to DC.Agent)	260 \$a		Place of publication, distribution, etc. (R)
E-Mail Address of Publisher	none	270 \$m	#0	E-Mail Address
	none	270 \$i	#7	Type of address (NR)
Distributor	DC.Publisher (open discussion relating to DC.Agent)	260 \$b		Name of publisher, distributor etc.
... (vgl. Publisher)				
Type	none	256 \$a		
		007 cr		Control field – Computer file
Language	DC.Language	041 \$a		Language Code
Country	none	008/15-17		Country Code
Keywords (controlled)	DC.Subject	650 \$a	10	Subject Added Entry – Topical Term
Keywords (controlled)	DC.Subject.LCSH	650 \$_0	10	Library of Congress Subject Headings
Description	DC.Description. Abstract	520 \$a		Summary, etc. note
Evaluation/Rating	none	500 \$a		General Note
URL	DC.Identifier Scheme=URL	856 \$u	4	Uniform Resource Locator
Source Type	DC.Type	655 \$a	#7	Index Term – Genre/Form
Last Update	DC.Date.Modified	583 \$d + \$a modified		Action note (action intervall)
Access	??DC.Rights	540 \$a		Terms governing the use and reproduction note
Restriction	??DC.Rights	540 \$a		Terms governing the use and reproduction note
...				

# = Blank

## 8 Anhang

Anhang 1: Auszüge aus dem Crosswalk SSG-FI zu Dublin Core und MARC 21 (für eine vollständige Version vgl. <<http://www.sub.uni-goettingen.de/ssgfi/corc/crosswalk/>>)

SSG-FI Metadatenkonzept, vgl. <<http://www.Geo-Guide.de/info/tools/metadata.html>>

SSG-FI	Dublin Core	MARC21	Indic.	Denotation
Title	DC.Title	245 \$a		Title statement
Author	DC.Creator (open discussion relating to DC.Agent)	700 \$a	1#	Added entry – personal Name
Author. Address	none	700 \$u	##	Address/Country
Country relating to Authors Address	none	see above	##	Country
E-Mail Address of Author	none	700 \$g	#0	E-Mail Address
Editor	DC.Creator (open discussion relating to DC.Agent)	700 \$a	1#	Added entry – personal Name
... (vgl. Author)				
Publisher	DC.Publisher.Name (open discussion relating to DC.Agent)	260 \$b		Name of publisher, distributor etc.

### Anschrift der Autorinnen:

Monika Cremer  
 Dr. Heike Neuroth  
 Niedersächsische Staats- und  
 Universitätsbibliothek  
 D-37070 Göttingen  
[cremer@sub.uni-goettingen.de](mailto:cremer@sub.uni-goettingen.de)  
[neuroth@sub.uni-goettingen.de](mailto:neuroth@sub.uni-goettingen.de)