Michaela Mundt 02.10.1998

Weiterbildung zur Wissenschaftlichen Dokumentarin
Feldseminar im Daten- und Rechenzentrum des Geoforschungszentrums Potsdam
20. Juli-2. Oktober 1998

Der DOI (digital object identifier)
ein verlagsorientiertes Indexierungswerkzeug auch anwendbar auf Datensätze?

Internetstudie zur möglichen Anwendbarkeit des DOI
für die im ICDP-Clearinghouse angebotenen Daten

Erster Teil	Einführung in das Thema	1
Zweiter Teil	Allgemeines zu Identifiern Funktion, identifizierte Objekte, Anforderungen in elektronischen Medien, Limitationen herkömmlicher Systeme und von Internet-Adressierungen	3
Dritter Teil	Der DOI Entwicklungsgeschichte, Entwicklung und Beschreibung des Identifiers, Aufbau des DOI-Systems, Kritik an Identifier und System	6
Vierter Teil	Urheberrecht im Internet (digitales Urheberrecht) Schutz von Datenbanken (EU-Datenbankrichtlinie u.a.)	12
Fünfter Teil	Weiteres Vorgehen im Rahmen des Clearinghouse zu klärende Punkte: Datengüte, Akzeptanz	15
	Anmerkungen und Literaturverzeichnis	18

Der DOI (digital object identifier)

ein verlagsorientiertes Indexierungswerkzeug auch anwendbar auf Datensätze?

Erster Teil:
Einführung in die Problematik

Das Internet hat in den letzten Jahren zunehmend an Bedeutung für kommerzielle Zwecke gewonnen; aus dem unentgeltlichen, freien Zugang zu Informationen im Netz wird Handel, die einzelne Information dabei zur Ware. Aus dieser Art der Internetnutzung ergeben sich einige Probleme:

- die Internet-Adressierungen sind oft nicht mehr aktuell („Error 404"), die entsprechende Information nicht mehr wiederzufinden. Ergebnis: der Nutzer/Kunde ist enttäuscht und im schlimmsten Fall (für den Verkäufer) nicht mehr am Produkt interessiert.
- die kommerzielle Nutzung des Internet erfordert ein ausgefeiltes Managementsystem zur internen und externen Abwicklung der verschiedenen Prozesse (Stichwort Electronic Cash, Verknüpfung von Bestellung und Bezahlung, Copyright).

Grundlage zur Lösung der genannten und weiterer Probleme aus diesem Zusammenhang ist zuallererst die eindeutige Identifizierung der gehandelten digitalen Objekte. Im Verlagswesen ist die Verwendung von Bezeichnercodes für unterschiedlichste Produkte schon länger üblich und hat sich im bisher üblichen kommerziellen Rahmen bewährt. Die bisher verwendeten Bezeichner wie ISBN, ISSN, PII usw. sind jedoch im Internet-Kontext nicht oder nur sehr eingeschränkt einsetzbar. Aufgrund der enormen Wachstumsquoten dieses Marktes auch und gerade für Verlagsprodukte (Online-Zeitschriften und Multimediamarkt!) ist es nicht verwunderlich, daß der Anstoß zur Entwicklung eines neuen Identifier-Systems für digitale Objekte aus dem Verlagsbereich gekommen ist. Der DOI (digital object identifier) wurde auf der Frankfurter Buchmesse 1997 zum ersten Mal der Öffentlichkeit präsentiert.

In dieser Studie soll untersucht werden, ob der ursprünglich für den Verlagsbereich entwickelte DOI auch für die Codierung von elektronischen Datensätzen im Kontext des ICDP-Clearinghouse angewendet werden kann. Dafür waren ausführliche Internet-Recherchen notwendig: Informationen zum aktuellen Stand der Entwicklung des DOI sind nur hier zu finden; die Diskussionen um aktuelle Problemfelder, die mit der Entwicklung oder den möglichen Anwendungen verknüpft sind, werden hier geführt und nicht in Fachzeitschriften: die Entwicklung wird mit hoher Geschwindigkeit vorangetrieben, Artikel wären schon lange vor Drucklegung veraltet.

Die Homepage der DOI-Foundation dient als Ausgangspunkt der Recherchen. Hier äußern sich in erster Linie die an der Entwicklung des DOI beteiligten Parteien, daneben gibt es weitere Links zu weiteren Interessierten. Die Entwicklung und Einführung des DOI wird allgemein begrüßt, das Potential des DOI als sehr hoch eingeschätzt. Grundsätzliche Kritik am DOI selbst und am Aufbau des verwaltenden Systems gibt es nicht, einige wenige Beiträge befassen sich mit Teilaspekten, die mittel- bis langfristig geklärt werden sollten.

Ein weiterer Aspekt dieser Studie sind Fragen des Urheberrechtes im Internet, speziell bezogen auf wissenschaftliche Daten und Datenbanken. Das Urheberrecht stellt eines der das Internet am stärksten limitierenden Gesetze dar und wird deshalb intensiv diskutiert. Das Netz ist zur Zeit ein noch weitgehend ungeregelter Raum, aber es gibt auf allen Ebenen Bestrebungen, Reglementierungen einzuführen: die wichtigen international gültigen Verordnungen schränken die (unlimitierte) Verwendung geschützten Materials auch im „fair use“ oder in legalen Zusammenhängen wie beispielsweise für Bibliotheken, immmer stärker ein. So fragwürdig diese Bestrebungen zum Teil sein mögen, für einen möglichen Zugang auf Datensätze und Datenbanken über das Internet, wie im Clearinghouse angedacht, ist die Absicherung der „Datenproduzenten“ obligatorisch.

Zweiter Teil

Allgemeines zu Identifiern

Identifier = Bezeichner = Kennzeichner

Bezeichner finden im Kontext des Verlags- und Publikationswesens Anwendung als Identifizierungscodes für die verschiedensten Publikationsformen wie Bücher (ISBN), Zeitschriften (ISSN, PII), Serien (SICI: serial and item contribution identifier)[i]; aber auch im Nicht-Printbereich gibt es entsprechende Werkzeuge (z.B. Audioaufnahmen auf Schallplatten und anderen Tonträgern).

Die erste und wichtigste Funktion eines Identifiers ist die Möglichkeit der eindeutigen Identifizierung eines Werkes, unabhängig vom Umfang oder Kontext. Aus dieser eindeutigen Identifizierbarkeit ergeben sich u.a. folgende Vorteile:

- Erleichterung des Handels zwischen Verlagen und Kunden; zwischen Verlagen und anderen Rechteinhabern
- Sicherung von Urheberschaft und Verwertung
- Aussagen über verschiedene Eigenschaften des Werkes (Form, Inhalt, Verlag) z.T. indirekt aus dem Aufbau des Identifiers und ohne Kenntnis des Werkes möglich
- Vereinfachung des eindeutiges Zitierens
- im Bibliotheksbereich Verwendung für Bestellvorgänge, Zuordnungen, Dublettenfindung

Identifizierte Objekte sind immer „*Werke*" verschiedenster Art; herkömmliche Formen wären z.B. Bücher, Artikel, Fotos, Karten..., neuerdings auch „*digitale Objekte*" wie beispielsweise elektronischer Text, Grafik, Multmediaobjekte, *Datensätze* (in digitaler Umgebung).

„Werke" sind im Urheberrecht definiert als „persönliche geistige Schöpfungen; die durch ihren Inhalt oder durch ihre Form ... etwas Neues und Eigentümliches darstellen" (darunter fallen auch Sammelwerke mit einer neuer Zusammenstellung der in ihnen vereinigten Originale und Datenbanken).

Digitale Objekte: „Digital objects can be existing electronic products, newly planned digital offerings or even digital product annoncements, online services, or internal digital archives“[ii]

„document-like object“: intellektuell gehaltvoller Inhalt, textbasiert, aber auch Karten, oder Bilder, usw. (schwammige Definition, nach[iii])

Anforderungen, die Identifier in elektronischen Medien erfüllen müssen

- Notwendigkeit der möglichen Indexierung einer enorm großen Anzahl von Objekten, z.T. in mehreren Versionen (10^{14} Identifier allein für STM-Material otwendig[iv], diese Angabe beschränkt sich lediglich auf Textdokumente!)
- Gewährleistung der langfristigen Auffindbarkeit dieser Dokumente (Persistenz)
- Identifizierung von Teilen digitaler Objekte beliebiger Granularität muß möglich sein
- Erleichterung des Zugangs auf die identifizierten Objekte
  nach Auffinden der Information ist sofort der Zugang auf das Dokument (direkt oder indirekt über Antwortseiten des Verlages) möglich
- Abwicklung automatisierter Abfrage- und Bestellvorgänge
- automatisierte Verwaltung der mit diesen Vorgängen verknüpften Rechte

Limitationen der herkömmlichen Identifier-Systeme:

- z.T. nur begrenzte Anzahl Identifier möglich (abhängig von der Struktur der Identifier)
- Indexierung verschiedener Versionen eines Dokumentes wird in den verschiedenen Bezeichnersystemen unterschiedlich gehandhabt
- Beschränkungen der Indexierbarkeit beispielsweise auf bestimmte Dokumentformen (SICI: nur bereits im Druck erschienene Dokumente, Konstruktion des Index benötigt Angaben aus dem Kontext der Zeitschrift)
- Unterstützung eines automatischen Managementsystems für Bestellungen, Copyright u.ä. durch die herkömmlichen Systeme nur begrenzt möglich (technische u. logische Limitationen)
- Granularität von Informationen festgelegt (SICI: durch enge Verknüpfung mit Druckerzeugnissen)

Limitationen der URLs (uniform resource names) im Internet

- URLs benennen Lokationen bzw. Verknüpfungen zwischen Lokationen im Netz, kennzeichnen aber nicht deren Inhalte
- URLs sind angelegt als flexible, einfach implementierbare und leicht zu entfernende Verknüpfungen, also von vornherein nicht als längerfristige Adressierung
- URLs sind begrenzt auf bestimmte Zugangsservices (z.B. Gopher, WWW)

Identifier auf Basis des Konzepts der *URN*s (uniform resource name)

- Bezeichner bestehen aus zwei Teilen: Index für die benennende Stelle (zentrale Vergabe) und ein Index für jedes Objekt
- sind nicht an die Verwendung eines bestimmten Zugangsservices gebunden
- Verknüpfung mit „resolution service", der die URNs direkt in URLs für den direkten Datenzugang umwandelt; dadurch sind verschiedene Lokationen (mirror sites) oder in unterschiedlichen Formaten abgelegte Versionen erreichbar
- bei Änderung der Zugangsadressierung muß nur der Eintrag in der adressverwaltenden Datenbank geändert werden
- in das gewählte System lassen sich die bisher verwendeten Identifier integrieren.

Dritter Teil

Der DOI (digital object identifier)

Kurzer Abriß der Entwicklungsgeschichte

Initiierung der Entwicklung eines ergänzenden/neuen Identifiers erfolgt 1994 durch die Association of American Publishers (AAP)
In der Planungsphase: Bedürfnisse des Verlagswesens erkunden
Beschluß zur Entwicklung des DOI im September 1996; Entwicklung der notwendigen Technologie durch die „Corporation for National Research Initiatives" (CNRI) und R.R. Bowker Publ.(USA) unter Nutzung des Handle System® (URN-Netzwerk) des CNRI. Eine interne Vorstellung des ersten Prototyps erfolgte im Februar 1997 in Washington (AAP/PSP meeting), die erste Präsentation des DOI in der Öffentlichkeit auf der Frankfurter Buchmesse 1997, die zweite Präsentation ist angekündigt für den 7.10.98 auf der 50. Frankfurter Buchmesse.

Die „International DOI Foundation" (IDF)
wurde als Stiftung gegründet 1997.
Gründungsmitglieder / Träger der Entwicklung waren
- Association of American Publishers (AAP)
- Corporation for National Research Initiatives (CNRI)
- International Publishers Association (IPA)
- International Association of Scientic, Technical and Medical publishers (STM)

heutige Mitglieder, u.a.
- einige große wissenschaftliche Verlage (Academic Press, Elsevier, Wiley, Springer)
- verschiedene ländergebundene Zusammenschlüsse von Verlagen, Autoren,
- Buchhandelsvereinigungen verschiedener Länder
- Software-Produzenten (Microsoft)
- verschiedene wissenschaftliche Vereinigungen (American Chemical Society)

ständig neue Mitglieder

Ziele:

- Unterstützung der Verbreitung des DOI, schwerpunktmäßig im Verlagsbereich
- Unterstützung der technischen Entwicklung
- Entwicklung von Applikationen (zuerst beschränkt auf den Verlagsbereich)
- Bereitstellung von Informationen über den DOI für alle Interessierten durch Pflege einer überaus reichhaltigen Internetseite[v] u.a. mit technischen Informationen, Literaturverzeichnis, Mailinglisten zu verschiedenen Themen, Eintrittsformularen

Entwicklung des DOI

Einbeziehung verschiedener Einrichtungen in Teilaspekte der weiteren Entwicklung:
NISO[vi] (USA): Standardisierung der Syntax
EDItEUR[vii]: deskriptive Metadaten, Metadaten für Rechte

Zielvorstellungen zu Beginn der Entwicklung:

- persistente Indexierung elektronischer Dokumente
- Nutzung für elektronische Copyright-Management Systeme: Routing mit Zentralverzeichnis, welches die Anfragen an Adressdatenbanken weiterleitet
- unter Verwendung eines „einfachen/unintelligenten" Bezeichners ohne weitere Bedeutung in der Syntax
- im Internet (html-kompatibel)

Beschreibung des DOI-Identifiers

Aufbau:

- maximal 128 Zeichen
- einfache Reihe von Zahlen und Buchstaben
- ohne vorgegebene Bedeutung (einfacher oder „dumb" identifier, im Gegensatz zu „intelligenten" Bezeichnern[viii])
- fast alle Zeichen verwendbar (Limitationen durch html-Syntax)
- läßt sich jedem eigenständigen Objekt zuordnen

– zwei Teile:

Präfix: zwei Stellen kennzeichnen das DOI-Verzeichnis, vier Zahlen den Verlag bzw. den Eigentümer, Trennung durch Punkt [.]

Suffix stellt die eigentliche Identifikationsfolge dar

keine zwingenden Vorschriften für die Struktur des Suffix

Empfehlung: Verwendung vorhandener Dokumentbezeichner (ISBN, usw...) unter Einfügung des Bezeichner-Codes

über diese spezifische Zeichenkette ist die eindeutige Kennzeichnung sowohl des Objektes als auch des Verlegers bzw. des Eigentümers möglich.

Was sonst noch wichtig ist:

- Möglichkeit der rückwirkenden Indexierung
- Anwendung auch als Identifier für herkömmliches Material (Gedrucktes) möglich

Verbindung Identifier-Objekt

verschiedene Möglichkeiten, z.B.:

- Einbettung in die Header verschiedener Dateitypen (HTML)
- im frei zugänglichen Bereich einer geschützten Datei
- als Feld einer Verwaltungsdatenbank

Beispiele: (nach [ix] und[x])

10.1007/s100529800702	(Springer, auf Basis des PII)
10.1002/(ISSN)0002-82321	(Wiley, Zeitschrift)
10.1002/(SICI)0002-82321(199601)47:1<>1.0.TX;2-J	(best.Ausgabe)
10.1002/(SICI)0002-82321(199601)47:1<23:TDOMII>2.0.TX;2-2	(Artikel)
10.1002/(SICI)0002-82321(199601)47:1<23:TDOMII>2.3.TX;2-U	(Abstract)

Springer: (Online first): DOI-Vergabe direkt nach der Fertigstellung und Annahme durch den Verlag, Zugang ist online lange vor Drucklegung möglich und erfolgt direkt auf die gewünschte Information (Artikel, Buch...)

Wiley: DOI auf Basis des ISSN für eine Zeitschrift bzw. des SICI (serial and item contribution identifier) für eine bestimmte Artikel, Vergabe des DOI ist deshalb an Drucklegung gebunden (SICI benötigt Seitenzahlenangaben); Zugang efolgt auf Angaben *über* die gesuchte Information

Beschreibung des DOI-Systems

Grundlage des DOI-Systems ist die Software des Handle System® des CNRI (URN-Netzwerk). Das System besteht aus einem Zentralverzeichnis, welches die Präfixe verwaltet und die Anfragen der Nutzer an die jeweiligen Präfixverwalter (Verlage usw.) weitergibt, und dortigen internen Verzeichnissen der mit DOIs versehenen Dokumente. Die Anfragen werden hier beantwortet; die Form, in der dies geschieht, ist dabei den Verlagen überlassen.

Die Aktivierung des DOI durch den User erfolgt durch die manuelle Eingabe des Codes (umständlich, Gefahr von Eingabefehlern)
oder durch Anklicken eines Buttons auf einer Internetseite.

Der DOI wird in eine Internet-Adresse folgender Form aufgelöst:

*http://dx.doi.org/*10.1007/s100529800702

Anfrage wird über einen Proxy-Server an den DOI-Verzeichnisserver (directory) unter der (derzeitigen) Adresse *http://dx.doi.org* weitergeleitet.

Der DOI-Verzeichnisserver (directory) speichet die Information über den aktuellen Ort des elektronischen Objektes (URL,unter der das Objekt zu finden ist); diese URL wird, via Proxy, als „HTTP-Redirect" zurück an den Nutzer gegeben, der sofort weiter an das Verlagsverzeichnis geleitet wird (Routing erfolgt im Hintergrund der Anwendung und ist für den Nutzer nicht sichtbar). Zu Beginn der Entwicklungsphase waren mehrere Verzeichnisse angedacht, inzwischen ist nur noch ein Verzeichnis geplant[xi].

Auflösung des obigen Beispiels in URL der Dokumentenquelle

http://dx.doi.org/10.1007/s100529800702

nach

http://www.springer.com/doi-d/100529800702

Die Verzeichnisse bei den Verlagen lösen den Suffix auf, machen dadurch das Dokument selbst zugänglich oder leiten die Anfrage auf Zugangsseiten mit weiteren Informationen (Bestellmöglichkeiten, Subscribe, Zugang zum Dokument über Eingabe von Passwords....) weiter;
aus DOI 100529800702 wird irgendein Artikel über ....

Weitere Aspekte der internationalen Diskussion:

– Umgang mit verschiedenen Versionen desselben Dokumentes
Vergabe eines ganz neuen DOI, Modifizierung des an das Urspungsdokument vergebenen DOI (Modifizierungen) oder unveränderten DOI nutzen? Welche Unterscheidungskriterien sollen hier gelten: gleicher Identifier = gleiches Produkt, dieselbe Version...

– Zugang zu Dokumenten
Wie der Zugang zu den indexierten Dokumenten erfolgen soll, liegt z.Zeit im Ermessen der Verlage. Es gibt die Möglichkeit, den DOI mit einer Einstiegsseite zu verknüpfen, die die verschiedenen Zugangsmöglichkeiten zum Dokument aufzeigt (Lieferung nach Zeichnung eines Abonnements mit Password, Bestellformular für herkömmliche Postlieferungen, „Schnupperzugang", Zugang lediglich auf den Gesamtkatalog...)[xii]. Welches Modell wird vom Nutzer bevorzugt, welches hat eine eher abschreckende Wirkung

– Kosten
Die Preisgestaltung für die Nutzung des DOI ist von Seiten der IDF noch nicht endgültig festgelegt. Zur Zeit sind für den Erwerb des Präfix 10.000 US-$ und ein jährlicher Beitrag, der von der Anzahl der pro Präfix eingetragenen einzelnen DOIs abhängig sein soll, im Gespräch. Für die Endnutzer sollen allerdings keine Kosten entstehen.

- Metadaten

  Welche Metadaten sind für die Nutzer bei der Suche nach elektronischen Dokumenten bzw. Objekten relevant? Es wird die Anbindung an den Metadatensatz des DublinCore diskutiert, wobei bei DublinCore wiederum verschiedene Teilaspekte in der Diskussion stehen (welche Felder gehören zum zentralen DC, wie sollen die Relationen ausgedrückt werden ..), die vor einer Einbindung in den DOI geklärt werden sollten.

Kritik am DOI (System)

beschränkt sich von Seiten der originär angesprochenen Nutzergruppe fast ausschließlich auf technische Details[xiii] und die noch nicht geklärten o.g. Aspekte. Die Bibliotheken befürchten Einschränkungen für die Nutzung und Verbreitung elektronischer Dokumente, wenn den Verlagen durch das DOI-System ein besseres Transaktionsmanagement zur Verfügung stehen wird. Für die Internet-Community stehen Fragen der Kontrolle (Internet als freie Kommunikationssphäre und Verbreitungsmedium für elektronische Dokumente) und der Sicherheit des DOI-Systems gegen Hacker im Vordergrund.

Vierter Teil
Copyright/Urheberechtsschutz

Mit dem Urheberechtwird dafür Sorge getragen, daß die Nutzungs- und Verwertungsrechte an einem Werk in den Händen des Herstellers bzw. Rechteinhabers (Urhebers) bleiben. So hat der Urheber i.d.R. das ausschließliche Recht,

- die vorübergehende oder dauerhafte *Vervielfältigung*
- *Umgestaltungen* in jeglicher Form
- Formen der (öffentlichen) *Verbreitung*
- die (öffentliche) *Wiedergabe*

  vorzunehmen oder zu erlauben. Eingeschränkt werden diese Rechte durch klar definierte Ausnahmen von den oben genannten „zustimmungsbedürftigen Handlungen", nämlich für die Vervielfältigung

- zum privaten Gebrauch
- zum eigenen wissenschaftlichen Gebrauch (Quellenangabe)
- zum Gebrauch für Unterrichtszwecke (Quellenangabe)
- für Verwaltungs- und Gerichtsverfahren und zum Zwecke der öffentlichen Sicherheit

Copyright im Internet

Eine Vielzahl von Seiten im Internet beschäftigt sich mit Rechtsfragen, naturgemäß auch mit Fragen, die mit der privaten und/oder kommerziellen Nutzung des Internets in Zusammenhang stehen (Themen z.B. Datensicherheit..., Urheberrecht). Online-Recherchen mit den Suchbegriffen Copyright/Urheberrecht ergeben deshalb sehr viele Treffer, davon beschäftigen sich ca. 60% mit dem Themenkomplex Copyright im Kontext des Internet. Der Urheberrechtsschutz in Bezug auf Datenbanken wird aus aktuellem Anlaß diskutiert (s.u.), zum Thema Copyright in bezug auf *wissenschaftliche* Daten im Internet gibt es hingegen keinerlei Veröffentlichungen im Netz.

Im DOI-Kontext wird dieser Themenkomplex übrigens ebenfalls nicht diskutiert, da der Aspekt des Umgangs mit Datensätzen im Verlagswesen ein marginales Problem darstellt und deshalb im Entwicklungsprozess des DOI keine große Rolle spielt.

(Fußnote zu[xiv] „...we have repeatedly stumbled across a problem with the unique identification of aggregations of data...". „Inevitably, many very good questions must be left unanswered").

Zum Urheberrechtsschutz von Datenbanken

Datenbanken unterliegen als „Sammlung von Werken, Daten oder anderen unabhängigen Elementen, die systematisch oder methodisch angeordnet und einzeln mit elektronischen Mitteln oder auf andere Weise zugänglich sind"[xv] bzw. als „collection and assembling of preexisting materials or of data that are selected in such a way that the resulting work as a whole constitutes an original work of authorship"[xvi], dem Urheberrecht.

Die wesentlichen internationalen Vereinbarungen sind zum einen die „Datenbankrichtlinie der Europäischen Gemeinschaft", die am 11. März 1996 vom Europäischen Parlament verabschiedet wurde und bis 1.1.1998 in den Mitgliedsstaaten in nationales Recht umgesetzt werden sollte (die Umsetzung in Deutschland erfolgt im Rahmen des „Informations- und Kommunikationsdienste-Gesetzes"[xvii]) sowie die Bestimmungen des US-amerikanischen Copyright. Des weiteren ist der Urheberrechtsvertrag der „World Intellectual Property Organization" (WIPO) vom 20. Dezember 1996 zu nennen[xviii], der bisher allerdings in keinem der unterzeichnenden Staaten in nationales Recht umgesetzt wurde; als internationaler Vertzrag zwischen Staaten entfaltet sie keinerlei Rechtswirkung. Diese Ergänzungen zur Berner Konvention aus dem Jahre 19... stellen eine Zusammenstellung von Mindestanforderungen an den Urheberrechtsschutz dar; in Bezug auf das Copyright für Datenbanken bieten sie allerdings, genauso wie die US-amerikanische Gesetzgebung, einen wesentlich eingeschränkteren Rahmen als die entsprechenden Vereinbarungen der europäischen Datenbankrichtlinie

Die europäische Datenbankrichtlinie bietet im Vergleich zum amerikanischen Copyright einen viel weitergehenden Schutz für den Urheber allein schon durch die weiter gefaßte Begriffsdefinition „Datenbank". Es werden nicht nur Datenbanken, „die aufgrund ihrer Auswahl oder Anordnung des Stoffes eine eigene geistige Schöpfung ihres Urhebers darstellen"[xix], unter Schutz gestellt, sondern ein prinzipielles Schutzrecht (Schutzrecht *sui generis*) für den Urheber/Hersteller einer Datenbank ist auch dann vorgesehen, wenn „für die

Beschaffung, die Überprüfung oder die Darstellung ihres Inhaltes eine in qualitativer und quantitativer Hinsicht wesentliche Investition erforderlich ist"[xx]. Diese Regelung findet weder im amerikanischen Copyright noch in der WIPO-Vereinbarung eine Entsprechung. Ein gleichartiges Schutzrecht *sui generis* soll aber Inhalt von Ergänzungsvorschlägen sowohl in den USA als auch in der WIPO sein[xxi].
Der Copyrightschutz einer Datenbank an sich beinhaltet keinen Schutz der zugrunde liegenden Daten. Ideen und feststehende Fakten können grundsätzlich nicht geschützt werden. Für den Schutz anderer Daten muß auf andere Bestimmungen zurückgegriffen werden (z.B. Geschmacksmusterschutz, zivilrechtliche oder strafrechtliche Bestimmungen, o.ä.[xxii])

Kritik von Bibliotheken und Bibliotheksverbänden[xxiii]

Bibliotheken sehen neben der Sammlung, Katalogisierung und Archivierung von Information ihre zentrale Aufgabe darin, verschiedenste Informationsträger für jedermann zur uneingeschränkten Nutzung zur Verfügung zu stellen und damit das Grundrecht auf Informationsfreiheit zu sichern. Die Bestimmungen der Europäischen Datenbankrichtlinie stellen eine wesentliche Nutzungeinschränkung für elektronische Datenbanken und anderer Software in Bibliotheken dar, da trotz einer nichtkommerziellen Nutzung der Abschluß von Lizenzverträgen (gekoppelt mit Gebühren) notwendig wird. Dies wird gewertet als eine „Kommerzialisierung des Urheberechtes", in der „die wirtschaftlichen Interessen der Urheber (...) als höherwertiger angesehen (werden) als die demokratische Informationsfreiheit".

Der Autor sieht das Urheberrecht originär als ein „Schutzinstrument für geistiges Kulturgut", welches sich langsam, aber sicher „zu einem reinen Wirtschaftsrecht" entwickelt hat, das „die Entstehung neuer Ungleichheiten zwischen denen, die über Informationen verfügen, und denjenigen, die nicht darüber verfügen, begünstigt". In Kontrast zu diesen negativen Entwicklungstendenzen stilisiert er die Bibliotheken als Wahrer des Rechtes auf freien Zugang zu Information hoch und fordert dazu auf, „den kommerziellen Spieß einfach um(zu)drehen (und) offensiv und durchaus auch unkonventionell gegen die drohende Ausgrenzung durch das digitale Urheberecht vorzugehen".

Fünfter Teil

Vorschläge für das weitere Vorgehen des ICDP-Clearinghouse bzgl. DOI

Die nachfolgenden Thesen sollen zum Ende dieses Berichtes ein paar Punkte aufzeigen, die vor einem möglichen Einsatz des DOI im Internet im Rahmen des Clearinghouse intern und auch im Rahmen der IDF diskutiert werden müssen. Es hat wenig Sinn, ein Werkzeug wie den DOI ohne pragmatische Ansätze zu Antworten auf die genannten methodischen Fragen und zu Fragen der Akzeptanz eines solchen Systems durch die Zielgruppe einzusetzen.

Definitionsprobleme

Für Datensätze aus dem wissenschaftlichen Kontext muß eine klare Definition des Begriffs „digitales Objekt" gefunden werden, dann lassen sich diese Datenpakete im Internet-Kontext problemlos als „Werke" im Sinne des Urheberrechts behandeln und mit einem Identifier wie dem DOI versehen. Mögliche Kriterien hierbei: Umfang, Format, ...?
(Dies gilt auch für Grafiken, Karten usw.).

Interne Organisation

Weiterhin muß VOR einer Anmeldung bei der IDF überlegt werden, wie das interne Verzeichnis für die Dokument-DOIs organisiert werden muß, ob eventuell mehrere DOI-Präfixe notwendig sein werden....

Kosten

Preis-Leistungsverhältnis: lohnt sich eine Applikation von DOIs allein vom Kostenaspekt für das GFZ-DRZ? Es ist zum einen notwendig, eine halbwegs realistische Abschätzung über die Anzahl auszugebender DOIs pro Jahr abzugeben. Zum anderen muß überlegt werden, welche Organisation als DOI-Provider in die IDF eintritt (die einmalige Anmeldungsgebühr beträgt derzeit 10.000$) Es wäre deshalb sinnvoller, wenn eine größere Organisation als das DataCenter, wie beispielsweise die Hermann-Helmholtz-Gesellschaft als Ganzes (oder andere???), diese Funktion übernehmen könnte.
Ein/Anbindung in ein Projekt des DFN? Andere Töpfe?

Zusammenarbeit mit Verlagen

Verknüpfung Datensätze - veröffentlichte Artikel: in welcher Form soll sie gestaltet werden? Auf Datensatz verweisender DOI in Referenzen,...?

Bei der Verknüpfung mit Artikeln, die nicht nur online, sondern auch in Druck veröffentlicht werden (bei denen also nicht jederzeit ein Update erfolgen kann), müssen Vereinbarungen mit den Verlagen schon im Vorfeld, spätestens parallel zum Einreichen der Artikel zur Veröffentlichung erfolgen und liegen zu diesem Zeitpunkt in der Verantwortung des Veröffentlichenden, meist des Autoren/Wissenschaftlers (wobei hier erschwerend die Frage der Akzeptanz des DOI-Systems in der wiss. Gemeinde hinzukommt; nähere Überlegungen hierzu siehe unten).

Güte der angebotenen Datensätze[xxiv]

(„Traue keinen Daten, die Du nicht selbst gefälscht hast.")

Welcher Wissenschaftler traut schon fremden Daten, von denen er nicht weiß, auf welche Art und Weise sie zustande gekommen sind? Es ist deshalb notwendig, den Prozeß der Datenerzeugung zu dokumentieren, d.h. den Datensätzen müssen umfangreiche Metadaten, beispielsweise über die Art und Weise der Datenerhebung, zur Seite gestellt werden.

Charakterisierung von Datensätzen

Für die Charakterisierung verschiedener Datensätze werden jeweils unterschiedliche Metadaten benötigt. Welche sind absolut notwendig, welche obligatorisch? Ein Ansatzpunkt zumindest für die Gestaltung dieser Metadaten bietet die Diskussion um den „Dublin Core"[xxv].

Zertifizierung von Datensätzen

Es müssen Werkzeuge/Verfahren gefunden werden, die garantieren, daß die ins Internet gestellten Daten „sauber" erhoben wurden. Deshalb muß vor einer möglichen „Veröffentlichung" im Internet ein Zertifizierungsprozess, der eine bestimmte Qualität der Daten garantiert, durchlaufen werden. Wie dieser Zertifizierungsprozess im Einzelnen auszugestalten ist, wird in weiteren Diskussionen zu klären sein.

Akzeptanz des DOI-Systems für Datenliferanten

Mindestens genauso wichtig ist es, dem einzelnen Wissenschaftler die Angst davor zu nehmen, daß Konkurrenten ihre Daten zur Widerlegung ihrer eigenen Ergebnisse benutzen könnten.

Aktualität der angebotenen Daten

Wie aktuell sollen und können die über ein DOI-Clearinghouse-Modell angebotenen Daten sein?

Vergleiche mit den Erfahrungen des World Data Center u.ä. Projekten zeigen:

WDC und ähnliche Datensammlungen entwickelten sich aus inhaltlich klar definierten Kampagnen (hier: Internationales Geologisches Jahr) und haben sich eine gewisse Akzeptanz für die von ihnen verwalteten (Meta-)Datensätze aus diesen Zusammenhängen erworben. Dabei sind die verwalteten Daten allerdings nicht tagesaktuell, da sie von den datenerhebenden Forschungseinrichtungen erst nach einer gewissen Wartefrist von den Erhebenden zur allgemeinen Nutzung freigegeben werden (im vorliegenden Beispiel nach zwei Jahren). Inhaltlich sind die genannten Projekte immer auf bestimmte Fachgebiete beschränkt und verweisen fast immer auf die datenerhebenden Institute, mit denen der datensuchende Nutzer dann selbst in Verbindung treten muß, um an die Daten zu gelangen.

Anmerkungen und Literaturverzeichnis

[i] Literatur zu diesem Thema über die verschiedenen Interessenverbände im Internet
Brian Green, Mark Bide: Unique Identifiers: a brief introduction (März '97)
(www.bic.org.uk/bic/uniquid.html)
Digital Information Objects and the STM Publisher; reproduced from STM Annual Report 1997
(www.elsevier.co.jp/inca/about/diginfo/Menu.shtml)

[ii] A Guide to Using Digital Object Identifiers for Creators, Publishers and Information Providers, p.4;
last update 10.Oktober 1997
(www.doi.org/guidebook/guidebook.html) 8/98

[iii] Priscilla Caplan: You Call it Corn, We Call it Syntax-Independent Metadata for Document-like objects.
The Public Access Computer Systems Review6, no.4 (1995): 19-23
(info.lib.uh.edu/pacsrev.html) 14.9.1998

[iv] nach: Norman Paskin: Information Identifiers; originally published in LEARNED PUBLISHING, Vol. 10, No.2, pp135-156 (April 1997)
(www.elsevier.nl/inca/homepage/about/infoident/)
Digital Information Objects and the STM Publisher, STM annual Report 1997
(www.elsevier.co.jp/inca/homepage/about/diginfo/menu.shtml

[v] http://www.doi.org

[vi] National Information Standards Organization (USA)

[vii] International Group for Electronic Commerce in the Book and Serials Sectors

[viii] hierzu u.a. in Norman Paskin: Information Identifiers, Learned publishing
(www.elsevier.nl/inca/homepage/about/infoident)

[ix] Infoseite des Springer-Verlages über die Anwendung des DOI in LINK
(http://link.springer.de/doi/doi.htm) 8/98

[x] „Wiley DOI Progress Report" vom DOI Workshop am 5.8.1998 in New York
über http://www.doi.org/....

[xi] Priscilla Caplan, DOI or Don't We?; The Public-Access Computer Systems Review9, no.1 (1998)
(http://info.lib.uh.edu/pacsrev.html) 16.9.98

[xii] Beispiele hierfür unter http://doi.wileynpt.com
http://link.springer.de/doi

[xiii] Clifford Lynch: Identifiers and Their Role in Networked Information Applications (1997)
(www.arl.org/newsltr/194/identifier.html) 8/98

[xiv] Mark Bide: In Search of the Unicorn, BNBFR Report 89, February 1998, p.7
(www.bic.org.uk/bic/unicorn2.pdf)

[xv] Richtlinie 96/9/EG des Europäischen Parlamentes und des Rates vom 11.März 1996 über den rechtlichen Schutz von Datenbanken (Datenbankrichtlinie), Kap.I, Art.1(1)
(http://www2.echo.lu/legal/de/ipr/datenbank/datenban.html) 6.9.98

[xvi] 17.U.S.C. §101, nach: Tysver, Daniel A., Database Legal Protection, BitLaw
(http://www.bitlaw.com) 8/98

[xvii] Gesetz zur Regelung der Rahmenbedingungen für Informations- und Kommunikationsdienste (Informations- und Kommunikationsdienste-Gesetz - IuKDG) vom 22.Juli 1997, BGBl.I, Jahrgang 1997, Nr.52);
hier besonders Artikel 7: Änderung des Urheberrechtes
nach (http://datenschutz-berlin.de/recht/de/rv/tk_med/iukdg.htm) 7.9.98

[xviii] WIPO Copyright Treaty (adopted by the Diplomatic Conference on Dec 20, 1996; Geneva)
(http://www.wipo.org/eng/main.htm) 7.9.98

[xix] Datenbankrichtlinie, Kap.II, Art.3(1)

[xx] Datenbankrichtlinie, Kap.III, Art.7(1)

[xxi] nach: Tysver, Daniel; Database Legal Protection, BitLaw
(http://www.bitlaw.com) 8/98

[xxii] Datenbankrichtlinie, Kap.IV, Art.13

[xxiii] Harald Müller, Digitales Urheberrecht - eine Gefahr für die Bibliotheken? Bibliotheksdienst Heft 8/98

[xxiv] hierzu auch: DFG-Denkschrift „Vorschläge zur Sicherung guter wissenschaftlicher Praxis", Wiley-VCH 1998
zum Umgang mit Primärdaten: Empfehlung 7 (S. 12)

[xxv] letzte Stand nach: Diann Rusch-Feja, Entwicklung der Dublin Core-Metadaten (Bericht über den 5. Dublin Core Metadata Workshop in Helsinki und einige Bemerkungen über DC-Metadaten in Deutschland), Bibliotheksdienst 2/98
(http://www.dbi-berlin/dbi_pub/bd_art/98_02_00.htm) 9.9.98