

# **Aufbau und Betrieb digitaler Repositorien in Indien**

## **Eine empirische Studie am Beispiel Open Source basierter Repositorien**

*Achim Oßwald*

Cologne University of Applied Sciences  
Institut für Informationswissenschaft  
Claudiusstrasse 1, 50678 Köln  
achim.osswald@fh-koeln.de

### **Zusammenfassung**

Institutionelle oder themenorientierte digitale Repositorien bilden zunehmend einen wichtigen Beitrag zur Verbreitung wissenschaftlicher Erkenntnisse. Sie werden weltweit mehrheitlich mit Open Source Software betrieben. Auf der Grundlage einer empirischen Studie aus dem Jahr 2006 gibt der Beitrag einen Einblick in die Rahmenbedingungen und den aktuellen Stand beim Aufbau und Betrieb solcher Repositorien in Indien.

### **Abstract**

Institutional and subject repositories are getting growing importance for science related communication. Worldwide Open Source Software is used for setting them up. Based on an empirical study in 2006 the paper gives insights in the circumstances of setting up and maintaining such repositories in India.

## **1 Einführung**

Das Konzept Digitaler Bibliotheken ermöglicht die Bereitstellung von digitalen Ressourcen als thematische, institutions- bzw. organisations- oder dokumenttypbezogene Sammlungen digitaler Objekte. Neben den international renommierten, themenorientierten Repositorien (subject repositories) wie dem berühmten ArXiv.org<sup>1</sup>, wurden im deutschsprachigen Raum schon relativ frühzeitig (ca. ab 1998) institutionelle Repositorien (IR) zumeist als

---

1 „ArXiv.org (genannt "The Archive") ist ein Archiv für Preprints (auch e-prints oder eprints genannt) aus den Bereichen Physik, Mathematik, Informatik und Biologie ursprünglich am Los Alamos National Laboratory (LANL) gehostet, jetzt wird es an der Cornell University betrieben und weltweit gespiegelt.“ <http://de.wikipedia.org/wiki/ArXiv> (Stand dieser und der anderen Webadressen vom 16.1.2007)

sog. Hochschulschriftenserver eingerichtet. Für deren Betrieb wurden mittels öffentlicher Fördergelder deutschsprachige Open Source Programme entwickelt.<sup>2</sup> Ende 2006 haben von diesen Entwicklungen in Deutschland OPUS<sup>3</sup> mit derzeit 55 sowie MyCoRe<sup>4</sup> mit derzeit 14 Anwendungen die größte Bedeutung. Open Source Produkte des internationalen Angebots haben in Deutschland bislang nur wenig Bedeutung.

DINI, die Deutsche Initiative für Netzwerkinformation e. V. bemüht sich seit Jahren, um das funktionale Niveau der bestehenden Publikationsserver. Hierzu wurde im Jahre 2003 und in 2. Auflage im September 2006 das „DINI-Zertifikat Dokumenten- und Publikationsservice 2007“ [DINI 2007] erstellt und propagiert. Die dort aufgestellten qualitativen Beurteilungskriterien sind:

- Sichtbarkeit des Gesamtangebotes
- Leitlinien (Policy)
- Autorenbetreuung
- Rechtliche Aspekte
- Integrität
- Erschließung
- Zugriffsstatistik
- Langzeitverfügbarkeit

Im Jahr 2006 wurde das Zertifikat auch in Indien als qualitative Orientierung in die Diskussion gebracht [Obwald 2006]. Auch wenn es nach Ansicht indischer Experten<sup>5</sup> noch zu früh ist, die indischen Anwendungen mit den DINI-Kriterien zu messen, so kann die hier vorgelegte Analyse indischer Open Source-basierter Repositorien<sup>6</sup> doch als ein Beitrag in eine solche, qualitativ orientierte Diskussion dienen.

Im internationalen Kontext waren<sup>7</sup> und sind neben den schon genannten Produkten OPUS und MyCoRe, weiteren, national relevanten sowie den

---

2 Zur terminologischen Klärung und Konkretisierung des Begriffs „repository“ siehe [Heery / Andersson 2005], 2-4.

3 Eine Übersicht über die aktuellen OPUS-Anwendungen gibt [http://elib.uni-stuttgart.de/opus/gemeinsame\\_suche.php](http://elib.uni-stuttgart.de/opus/gemeinsame_suche.php). Weitere Informationen zu OPUS werden unter <http://elib.uni-stuttgart.de/opus/doku/dokumentation.php?la=de> bereitgestellt.

4 Basierend auf der Essener MILESS-Entwicklung. Weitere Informationen zu MyCoRe unter <http://www.mycore.de>.

5 So die Aussage von Devika Madalli bei einem DINI-Workshop am 21.2.2006 im Rahmen der International Conference on Semantic Web & Digital Libraries in Bangalore, Indien (<http://drtc.isibang.ac.in/icsd/index.html>).

6 Eine ausführlichere Darstellung der Analyse und ihrer Ergebnisse wird in einer gesonderten, englischsprachigen Publikation erscheinen.

7 Einen allerdings z.T. veralteten funktionalen Überblick gibt [OSI 2004].

mittlerweile auch im internationalen Kontext eingesetzten Entwicklungen Fedora und CDSWare die folgenden Open Source Programme für den Aufbau digitaler Repositorien von Relevanz:

- DSpace, ein im November 2001 von MIT Libraries und Hewlett-Packard vorgestelltes Produkt, für das im Januar 2007 weltweit 190 Installationen aufgelistet werden – davon 5 in Deutschland.<sup>8</sup>
- EPrints, ein an der University of Southampton, School of Electronics and Computer Science entwickeltes Programm, das 2000 erstmalig vorgestellt wurde.<sup>9</sup> Im Januar 2007 werden 218 sog. „Archive“ mit Eprints weltweit betrieben – davon 10 in Deutschland.
- Greenstone Digital Libraries (GSDL), eine von der University of Waikato im Rahmen des New Zealand Digital Library Projekt entwickelte Open Source Software, die von der UNESCO als Anwendungsprogramm zur Distribution von digitalen Materialien sowie zum Aufbau digitaler Bibliotheken empfohlen und genutzt wird.

Von diesen drei Open Source Produkten ist GSDL am wenigsten Workflow-orientiert; stattdessen ist es primär auf die Bereitstellung ohnehin schon vorhandener digitaler Objekte oder Metadaten ausgerichtet.<sup>10</sup>

Alle drei Programme sind Unix-basiert. GSDL läuft allerdings auch auf Windows-Betriebssystemvarianten. Es ermöglicht zudem den Export und die Bereitstellung der Ressourcen der Repositorien z.B. auf eine CD-ROM und seine Anwendungsoberfläche ist in einer Vielzahl von Landessprachen verfügbar (vgl. <http://www.greenstone.org>).

Die Spezifika dieser Programme, insbesondere jedoch ihre Bereitstellung als Open Source Software haben dazu geführt, dass sie weltweit vorzugsweise im Hochschul- und Forschungskontext genutzt werden. Indische Einrichtungen haben nicht nur sehr frühzeitig diese Programme genutzt<sup>11</sup> und propagiert, sondern bilden – zumindest für DSpace und EPrints – mittlerweile die größte Nutzergemeinde außerhalb des anglo-amerikanischen sowie europäischen

---

8 Allgemein zu Dspace:

[http://wiki.dspace.org/index.php/EndUserFaq#About\\_DSpace](http://wiki.dspace.org/index.php/EndUserFaq#About_DSpace). Siehe auch <http://wiki.dspace.org/index.php/DspaceInstances>

9 Vgl. für weitere Informationen zu EPrints <http://www.eprints.org> sowie [Harnard / Tansley 2000].

10 Greenstone gibt auf seiner Homepage zahlreiche Beispielsanwendungen der Software. Die Anzahl der berichteten Downloads betrug nach Auskunft des Entwicklers bislang mehrere 10000. Die Anzahl der Implementierungen in Deutschland ist nicht bekannt, dürfte aber sehr gering sein.

11 So erfolgte z.B. die erste DSpace Anwendung außerhalb des MIT am Documentation Research & Training Centre (DRTC) des Indian Statistical Institute in Bangalore, Indien. Zudem fanden schon im Jahr 2002 Schulungsworkshops zu EPrints und DSpace in Indien statt.

Kontextes. Irritierenderweise wurde sie bislang von den einschlägigen Untersuchungen in diesem die internationale Diskussion dominierenden regionalen Kontext weitgehend ignoriert.<sup>12</sup>

Zu den drei genannten Open Source-Programmen<sup>13</sup> werden in den einschlägigen Verzeichnissen<sup>14</sup> allein für Indien folgende Zahlen öffentlich zugänglicher digitaler Repositorien / Bibliotheken<sup>15</sup> genannt:

- *DSpace*: 22 digitale Repositorien (vgl. Anlage 1)
- *EPrints*: 7 digitale Repositorien (vgl. Anlage 2)
- *Greenstone Digital Library*: 6 Webadressen digitaler Bibliotheken mit mehreren Subbibliotheken (vgl. Anlage 3).

Insbesondere hinsichtlich GSDL kann darüber hinaus von einer erheblichen Anzahl von lokalen, nicht über das WWW erreichbaren Installationen ausgegangen werden.

Ein nicht unerheblicher Teil dieser auf internationalen und nationalen Listen geführten Anwendungen war allerdings während einer vierwöchigen Untersuchungsperiode (Mitte November – Mitte Dezember 2006) nicht erreichbar. Dies korrespondiert mit Berichten in aktuellen Veröffentlichungen zur Entwicklung digitaler Bibliotheken in Indien [z. B. Ghosh/Das 2006]. Insgesamt wird dabei erkennbar, dass Fragen der Qualitätssicherung und Erreichbarkeit der Anwendungen von entscheidender Bedeutung sind:

- „The uniform resource locator (URL) has been changed in due course of time, but other web pages link to the earlier URL (i.e., dead link).
- The web server is not operational on 24/7 timeframe that makes it inaccessible to the scholars across the world. ...
- The service has been discontinued but the web pages that are linked to this IR are not updated.
- The web server is down for the time being.” [Das/Sen/Dutta 2005, 96]

Darüber hinaus wurde festgestellt (ebd. 94), dass nach anfänglicher Euphorie der Zugang nachgewiesener Dokumente z.T. deutlich zurückging oder gar

---

12 Vgl. z.B. die Untersuchung von [van Westrienen / Lynch 2005].

13 Die anderen genannten Open Source-Entwicklungen – incl. Fedora – sind nach Auskunft indischer Experten bislang nicht in erwähnenswerter Weise zur Anwendung gekommen.

14 Dies waren für die hier genannten Zahlen das “Registry of Open Access Repositories” (ROAR; <http://roar.eprints.org>); “Sites Powered by EPrints” (<http://www.eprints.org/software/archives/>), “DSpace installations world wide” (<http://wiki.dspace.org/index.php//DspaceInstances>); “Digital Repositories using DSpace in India“ <http://drtc.isibang.ac.in/DRTC/dspace-india.html>) und Greenstone Support Organization for India (GSOI) <http://greenstonesupport.iimk.ac.in/indiandls.htm>.

15 Auffallend (und noch näher zu untersuchen) in der indischen Szene ist der Umstand, dass sehr gerne der Terminus „Digitale Bibliothek“ synonym zu digitale Repositorien benutzt wird. Diese Uneindeutigkeit spiegelt sich hier wieder.

ganz nachließ. Dass solche Probleme nicht nur Indien-spezifisch sind, zeigen jüngste Veröffentlichungen in der internationalen Fachliteratur (vgl.[Gierveld 2006]; [Sale 2006]).

Vor diesem Hintergrund ergibt sich mindestens die folgende Frage:

- Was sind die speziellen Rahmenbedingungen und Probleme, unter denen diese Repositorien in Indien eingerichtet und betrieben werden?

Deren Beantwortung kann interessante Hinweise auf die Entwicklung digitaler Bibliotheken in einem der Länder Südostasiens mit hoher wirtschaftlicher und wissenschaftlicher Dynamik geben. Darüber hinaus können Überlegungen abgeleitet werden, in welcher Weise und unter welchen Voraussetzungen dort die von DINI formulierten Qualitätsanforderungen zum Tragen kommen können.

## **2 Zielsetzungen und Methodik**

Ziel dieser Untersuchung ist die Ermittlung von Stand und Rahmenbedingungen des Aufbaus und Betriebs digitaler, öffentlich über das Internet zugänglicher, mittels Open Source Software betriebener Repositorien in Indien. Wie dargelegt, sollen auf dieser Grundlage nicht nur Erkenntnisse über und ggf. Empfehlungen zu den landesspezifischen Rahmenbedingungen des Betriebs solcher Anwendungen abgeleitet werden, sondern gleichzeitig auch allgemeine Hinweise, die für die Betreiber digitaler Bibliotheken – im Sinne institutioneller und fachlicher Repositorien – von Interesse sein könnten.

Die Studie ist weitgehend empirisch-analytisch angelegt, indem

- die ermittelten Anwendungen hinsichtlich ihres Angebots nach einfachen Kennzahlen sowie ergänzenden Angaben analysiert werden;
- auf der Grundlage einer Fragebogen-basierten Umfrage Selbstauskünfte zu den Entstehungs- und Bereitstellungsbedingungen der Repositorien ermittelt und in das Gesamtbild einbezogen werden.<sup>16</sup>

Aus fachlichen wie auch politischen Gründen wurde hierbei bewusst darauf verzichtet, in diesem Stadium der Entwicklung die in Deutschland entwickelten DINI-Kriterien an die indischen Repositorien anzulegen.

---

16 Die Untersuchung erfolgte Mitte November bis Mitte Januar 2007 und fand im Rahmen eines Forschungsaufenthaltes des Autors in Indien statt. Der Fragebogen wird vom Autor auf Nachfrage gerne zur Verfügung gestellt. Logistische und technische Unterstützung erfolgte durch das Documentation Research & Training Centre (DRTC), Indian Statistical Institute, Bangalore Centre (<http://drtc.isibang.ac.in/DRTC/index.html>).

## 3 Öffentlich zugängliche digitale Repositorien in Indien: Analyse und Befragung

### 3.1 Analyse der Repositorien

#### 3.1.1 Organisations- und softwarebezogene Ergebnisse

**Verfügbarkeit.** Von den insgesamt 35 (D 22 / E 7 / G 6)<sup>17</sup> mit den drei genannten Open Source-Programmen betriebenen digitalen Repositorien waren während des Untersuchungszeitraums 10 (D 8 / E 1 / G 1), also knapp 30% permanent nicht erreichbar. Wesentliche Gründe hierfür dürften Netzwerk- und Serverprobleme sein.<sup>18</sup>

**Hauptsächliche Anwender.** Die Hauptgruppe der Anwender der öffentlich zugänglichen digitalen Repositorien sind bislang wissenschaftliche Forschungseinrichtungen von nationaler Bedeutung wie Indian Institute of {Science / Technology / Management ...}, National Institute of {...} etc. sowie einige wenige Universitäten oder andere wissenschaftliche Einrichtungen. Sie sind mit den internationalen Entwicklungen im Bereich digitaler Bibliotheken vertraut und verfügen darüber hinaus zumeist über interne Strukturen (IT-Know-how; Netzanbindung; 24/7-Unterstützung im technischen Bereich), die für die Einrichtung von digitalen Repositorien notwendige Voraussetzung sind.

**Genutzte Open Source Software.** Auch wenn bedacht werden muss, dass die öffentlich verfügbaren Listen kein vollständiges Bild der Softwarenutzung für digitale Repositorien geben, so ist dennoch offensichtlich, dass bislang mehrheitlich DSpace als Anwendungsprogramm genutzt wird.

Fachveröffentlichungen zur Entwicklung der indischen Bibliotheksszene (z.B. [Jain/Babbar 2006]) machen außerdem deutlich, dass Open Source-Software in diesem Segment dominiert und proprietäre, d.h. herstellereigene Software bislang keine Marktrelevanz hat.

**Funktionale Unterschiede zwischen den genutzten Open Source-Programmen.** Zielsetzung und Funktionsweise der beiden Programme DSpace und EPrints unterscheiden sich nachgewiesenermaßen nur geringfügig (vgl. z.B. Madalli 2003, Borghoff et al. 2005 sowie Madalli/Seth

---

17 D steht für DSpace-, E für EPrints- und G für GSDL-Anwendungen.

18 Darüber hinaus deutet diese Diskrepanz darauf hin, dass entweder die aktualisierende Kommunikation zwischen den Anwendern und den Anbietern der Übersichtslisten nicht funktioniert oder die Werbefunktion dieser Listen dominiert.

2005). Beide sind originär auf die Archivierung von Materialien durch Wissenschaftler bzw. Mitglieder der entsprechenden Zielgruppe, den damit verbundenen Workflow sowie die Recherche und Bereitstellung der archivierten Ressourcen ausgerichtet. Auch wenn mit ihnen ebenso die Bereitstellung von retrospektiv digitalisierten Ressourcen möglich ist, so war dies nicht ihr originärer Entwicklungszweck.

Im Gegensatz dazu wurde Greenstone Digital Library primär dafür entwickelt, schon digital vorhandene Materialien bzw. Metadaten recherchefähig zusammen- und bereitzustellen. Die Ausgabe und Distribution auf digitalen Datenträgern ist insbesondere für datentechnisch noch unzulänglich vernetzte Anwendungszusammenhänge von wesentlicher Bedeutung.<sup>19</sup> Workflow-bezogene Funktionen fehlen. Hingegen sind die Funktionen der Oberflächen-gestaltung incl. der Bereitstellung in einer Vielzahl von Landessprachen<sup>20</sup> sowie die anwendungsspezifische Rechercheaufbereitung weit entwickelt. Greenstone zielt insofern mit seiner anderen Funktionalität auf primär bereitstellungs- und distributionsorientierte Vorhaben. Der Input wird bei Greenstone normalerweise von Informationsspezialisten vorgenommen. Die andersartige, faktisch breiter gefächerte Anwendergemeinschaft von GSDL ergibt sich auch durch die Verfügbarkeit auf Windows-Betriebssystemen, womit ggf. auch weniger IT-erfahrene Anwendergruppen erreicht werden.

**Unterstützung in Indien.** Für alle drei Anwendungsprogramme gibt es generelle und speziell auf Indien ausgerichtete Unterstützungsangebote. Noch am wenigstens ausgeprägt ist dies für **EPrints**, für dessen Interessenten zuletzt 2004 in Indien mehrere Workshops durchgeführt wurden, deren Teilnehmer wohl die Multiplikatorenfunktion noch nicht im erwarteten Maße realisiert haben. Für die Anwender besteht die Möglichkeit, an der weltweiten email- und Forumskommunikation von EPrints-Anwendern teilzuhaben.

Weitaus stärker, personell präsenter und aktiver ist die Unterstützung für **DSpace** und seine Anwender durch das DRTC des Indian Statistical Institute, das auch einen Vertreter in den DSpace-Beirat entsendet. Mittels bislang 12 Schulungworkshops in Indien, der Bereitstellung von die Installation vereinfachenden Skripten sowie einer Mailingliste wurden und werden Anwender unterstützt.

Für **GSDL** wurde im Jahre 2006 eine Unterstützergemeinschaft für den Raum Süd-

---

19 Bezeichnenderweise schlägt sich dies nicht in einer höheren Anzahl von öffentlich verfügbar gemachten Anwendungen nieder, da GSDL vorzugsweise als Einzelplatzversion oder im Intranet genutzt wird.

20 GSDL ist in Englisch, Französisch, Spanisch und Russisch verfügbar; für die Anwendungsoberfläche wird eine Vielzahl von weiteren Sprachen - darunter mehrere indische Sprachen - angeboten.

ostasien gegründet<sup>21</sup>, die seitens der Hauptentwickler von GSDL in Neuseeland unterstützt wurde und wird. Die weit über die sechs aufgelisteten Anwendungen hinausgehende Nutzergruppe wird primär mittels einer Mailingliste beraten. Durch die UNESCO-Kooperation ist GSDL insbesondere in bibliotheks- und informationsorientierten Fachkreisen bekannt und wird dort angewendet.

In allen drei Fällen sind die Unterstützungsaktivitäten wesentlich initiiert und getragen von internationalen, persönlichen Kontakte indischer Informationswissenschaftler, die - finanziert durch nationale oder internationale Förderprogramme – Gelegenheit zu fachlichem Austausch in einem jener Länder hatten, die die fachwissenschaftliche repository-Diskussion dominieren (vgl. FN 12).

### *3.1.2 Statistische Analyse*

Der Großteil der Repositorien / digitalen Bibliotheken wurde in den Jahren 2004-2006 eingerichtet. U.a. vor diesem Hintergrund ist erklärbar, dass die meisten Sammlungen bislang nur aus mehreren Hundert Dokumenten bestehen. In diesem Punkt folgen sie offensichtlich sehr ähnlichen Entwicklungsstadien wie die Repositorien in anderen Ländern.<sup>22</sup>

Für die erreichbar gewesenen Datenbestände ergibt sich danach folgendes Bild: Tabelle 1 zeigt die Anzahl der bereitgestellten Dokumente in DSpace-Anwendungen, Tabelle 2 die in Eprints-Anwendungen und Tabelle 3 die in den öffentlich zugänglichen GSDL-Bibliotheken in Indien. Die Darstellung in Tabelle 3 ist abweichend, um die Daten für die GSDL-charakteristischen „Bibliotheken“ (Teilkollektionen von Dokumenten u./o. Metadaten) wiedergeben zu können. Die dort z.T. sehr hohen Zahlen ergeben sich durch den Nachweis von (importierten) bibliografischen Metadaten (MD). Die Datensätze sind ggf. mit dem Volltext oder einer digitalen Kopie eines Dokumentes verlinkt. Berechnungen auf der Grundlage der in den Repositorien gemachten Jahreszuordnungen bei der Eingabe der Datensätze haben für die meisten der untersuchten Repositorien eine jährliche Inputrate von unter 100 Dokumenten ergeben.

Bei den gegebenen Zahlen zeigt die bei DSpace wie auch EPrints verfügbare Option der Differenzierung des Dokumentbestandes in sog. „communities“ und „collections“ (DSpace-Terminologie für Ober- und Untergruppen von Dokumenten) noch wenig erkennbare Vorteile - auch wenn sie für die Wissenschaftler zur Eingabe ihrer Dokumente in der für sie jeweils fachlich relevant Rubrik vorbereitet wurden.

---

21 Für weitere Details siehe <http://greenstonesupport.iimk.ac.in/>.

22 Vgl. die Angaben bei [van Westrienen / Lynch 2005] sowie die Zahlen bei [Heery / Andersson 2005] S. 21f, die für das Jahr 2005 auf gleichem Niveau liegen.

Daten zu DSpace-basierten Repositorien	D 1	D 2	D 3	D 5	D 6	D 7	D 8	D 11	D 12	D 14	D 15	D 16	D 17	D 20
Anzahl von Dokumenten im Repository	202	84	120	1197	175	193	1867	199	257	357	477	324	1415	146
Anzahl der Autoren (z.T. geschätzt)	312	75	154	788	82	192	2305	192	199	451	673	256	590	39
Anzahl der „collections / communities“	5	1	4	8	19	6	4	4	1	1	9	1	8	8

Tabelle 1: DSpace-basierte Repositorien: Basisdaten (November 2006)<sup>23</sup>

Daten zu EPrints-basierten Repositorien	E 1	E 2	E 4	E 5	E 6	E 7
Anzahl von Quellen im Repository	153	170	5836	2038	1376	75
Anzahl der Autoren (z.T. geschätzt)	200	85	300	960	*	100
Anzahl der „collections / communities“	9	18	9	11	*	10

Tabelle 2: EPrints-basierte Repositorien: Basisdaten (November 2006); \* keine Angaben

Nur wenige der Anwendungen weichen von der Basiskonfiguration (sog. „out of the box“-Konfiguration) der Programme ab, so dass kollektionsbezogene Funktionserweiterungen sowie institutionsbezogene Designvarianten zumeist nicht zum Tragen kommen. Entsprechend spröde ist das Erscheinungsbild der meisten Anwendungen. Problematischer im Hinblick auf die Integration der nachgewiesenen Dokumentbestände ist jedoch der Umstand, dass die Repositorien in den meisten Fällen

- nicht durch klar gekennzeichnete Links, in vielen Fällen sogar überhaupt nicht über die Homepage der jeweiligen Bibliothek / Informationseinrichtung der Institution aufrufbar sind;
- funktionale stand-alone-Anwendungen geblieben sind, so dass

<sup>23</sup> Die Nummern der Tabelleneinträge (z.B. D5) korrespondieren mit der Nummerierung der in den Anlagen aufgelisteten Repositorien.

normalerweise keine bestandsübergreifende Recherchen in gedruckten und digital bereitgestellten Ressourcen möglich ist. Auch die Integration in die sonstigen bibliothekarischen Anwendungsprogramme fehlt.

Daten zu Greenstone-basierten Repositorien	G 1	G 2	G 4	G 5	G 6
Anzahl von Dokumenten / Metadatensätzen im Repository bzw. in seinen sog. Bibliotheken (z.T. geschätzt)	~ 200 ~ 200 ~ 100 ~ 40 ~ 130 ~ 340 55 4 ~390 10	72 >17000 MD 28 7 58 7 57 ~ 100 ~ 1000 13	>3000  keine Angaben	12 36 7 15 20 7 5 7 210 10 15 6 9 4 13 1	40000 MD
Anzahl der Autoren	Angaben kaum ermittelbar				
Anzahl der „Libraries“	10	10	2	16	1

Tabelle 3: GSDL-basierte Bibliotheken: Anzahl der in den jeweiligen „Bibliotheken“ bereitgestellten Dokumente bzw. Metadatensätze (November 2006)

Größtenteils sind die Repositorien als institutionelle Sammlungen angelegt. In einigen Fällen ist dies auch in einer einfachen Sammelleitlinie formuliert. Detaillierter und elaborierter sind größtenteils Aufgabendarstellungen und Sammelleitlinien, die für die themenorientierten Repositorien angelegt wurden. Es sind dies vorzugsweise jene, die auch bei OAIster (<http://oaister.umdl.umich.edu/o/oaister/>) für den Austausch von Metadaten registriert sind bzw. einen OAI-basierten Datenaustausch realisieren (z.B. LDL, Librarians' Digital Library, <https://drtc.isibang.ac.in/>).

## 3.2 Fragebogenbasierte Umfrage zum Betrieb der digitalen Repositorien

### 3.2.1 Inhaltliche Zielsetzung, Verteilung und Rücklauf

Parallel zur direkten Analyse der Websites und Repositorien erfolgte im November / Dezember 2006 eine Befragung der Ansprechpartner / Betreiber der ermittelten Repositorien mittels Fragebogen.

Der Fragebogen umfasste 25 Fragen (Nr. jeweils in Klammern genannt) zu

- Grundlegenden Kontaktdaten (1-2)
- Typ und Inhalt des digitalen Repositoriums (3-5)
- Ablauf und Voraussetzungen für die Einrichtung des digitalen Repositoriums (6-11)
- den Umständen, unter denen das DR betrieben wird (12-19)
- der genutzten Open Source Software (20-21)
- Verfügbarkeit und Nutzung des digitalen Repositoriums (22-24)
- Zukünftigen Erwartungen bezüglich des digitalen Repositoriums (25)

Nach einem Pretest im Rahmen direkter Befragung wurde der Fragebogen am 23.11.2006 an 30 Einrichtungen (D 19 / E 5 / G 6)<sup>24</sup> mittels e-mail verschickt. Innerhalb 16 Tagen wurde ein Rücklauf von 13 beantworteten Fragebögen verzeichnet (D 9 / E 2 / G 2). Mit der für derartige Umfragen überdurchschnittlichen Rücklaufquote von 43,3% können die nachfolgenden Aussagen auf einer vergleichsweise stabilen Grundlage erfolgen.

### 3.2.2 Ergebnisse der Umfrage

Die Auswertung der Antworten auf die Fragebögen zeigen die nachfolgend zusammengefassten Trends und Ergebnisse:

1. Für die untersuchten Repositorien lässt sich ein klarer Trend zu institutionellen Repositorien erkennen.
2. Die meisten der Repositorien bieten bislang textbasierte Materialien an.
3. Mangels Input durch Autoren aus den Institutionen überwiegen bislang digitalisierte Dokumente in den Repositorien. Dennoch haben originär digitale Materialien schon einen hohen Anteil an den Beiträgen.
4. Die meisten der Repositorien wurden im Jahr 2005 begonnen, nur einige wenige früher in den Jahren 2001/2002 bzw. später im Jahr 2006.
5. Sie wurden zumeist auf Initiative einer institutionellen Bibliothek bzw. eines/r Bibliothekars/in, in einigen Fällen auch nach Initiative externer Aktivisten, kaum aber von den Wissenschaftlern oder der Verwaltung

---

<sup>24</sup> Mangels Verfügbarkeit der Website bzw. der Angabe eines Ansprechpartners konnte nur an soviel Repositorienverantwortliche ein Fragebogen verschickt werden.

- der wissenschaftlichen Einrichtungen realisiert.
6. Kenntnisse über die Einrichtung eines digitalen Repositoriums kamen zumeist von Dritten wie z.B. den Initiatoren und Referenten der Schulungsworkshops. Sie brachten ihr Fachwissen bei der eigentlichen Einrichtung des Servers mit ein. Alternativ dazu erfolgte dies durch die Spezialisten aus den IT-Abteilungen.
  7. In ungefähr der Hälfte der Fälle gibt es weder detaillierte Sammelleitlinien noch sog. institutional policies zum Aufbau der Repositorien. Nur in Einzelfällen (z.B. National Institute of Technology, Rourkela) wird eine institutionelle Selbstverpflichtung formuliert und im Kontext des Repositoriums bekannt gemacht. Allerdings stellen die meisten der Trägerorganisationen die notwendige Infrastruktur in Form von Hardware, Netzwerkanbindung und Arbeitszeit von Fachpersonal zur Verfügung.
  8. Die wöchentliche Arbeitszeit für den Betrieb des Repositoriums wird mit 3 bis maximal 10 Stunden angegeben.
  9. Die folgenden technischen Probleme werden genannt, die Einfluss auf den Betrieb nehmen:
    - Serverprobleme
    - „übergelaufene“ Logfiles
    - Unmöglichkeit, den gesamten Inhalt des Repositoriums zu exportieren
    - Netzwerkprobleme
    - Formatprobleme bei den angelieferten Dokumenten.<sup>25</sup>
  10. Die folgenden organisatorischen Probleme wurden genannt:
    - fehlende Bereitschaft der Autoren ihre Veröffentlichungen selbst ein zuliefern bzw. der Sammlung zur Verfügung zu stellen
    - fehlende Erlaubnis der Verleger, Dokumente auf diesem Wege bereit zu stellen.
  11. Bibliothekare überprüfen die von den Autoren eingegebenen Metadaten (immanentes Workflow-Feature bei DSpace und Eprints) – sofern sie nicht ohnehin selbst diese Daten eingeben. Der damit verbundene arbeitszeitliche Aufwand variiert in Abhängigkeit von der Anzahl der aufgenommenen Dokumente von einer Arbeitswoche für eine Person pro Monat bis hin zu zwei Vollzeitmitarbeitern.
  12. Die meisten der Anwender scheinen mit der von ihnen genutzten Software zufrieden zu sein<sup>26</sup>, auch wenn einige deutlich machten, dass sie noch erhebliche Erfahrungsdefizite bei der Nutzung sehen. Einige

---

25 Die genannten Probleme könnten auch auf eine noch nicht hinreichende Kompetenz im Umgang mit der Software hindeuten.

26 Aufgrund der statistischen Gegebenheiten bei der Beantwortung des Fragebogens bezieht sich diese Aussage vorrangig auf die DSpace-Anwender.

Anwender haben mit allen 3 Produkten Erfahrungen gesammelt, um die für ihre Bedürfnisse geeignetste Software zu ermitteln.<sup>27</sup>

13. Die monatliche Zugriffsrage ist mit 20 bis zu 6-7000 Zugriffen sehr unterschiedlich und spiegelt die sehr heterogene Anwendungslandschaft wider. Hierzu passt auch die Aussage, dass einige Betreiber diese Zahl nicht wissen und bislang auch nicht zu ermitteln versuchten.
14. Die öffentliche Bereitstellung von statistischen Daten wird bislang nur von ganz wenigen Betreibern praktiziert.
15. Nur einige wenige der Sammlungen sind nicht nach dem Konzept 24/7 verfügbar. Diesbezüglich muss aber einbezogen werden, dass die Antwortenden ohnehin eine positive Auswahl darstellen.
16. Die Erwartungen an die zukünftige Entwicklung der Sammlung zielen insbesondere darauf, die Bereitschaft der Wissenschaftler zur Weitergabe ihrer Veröffentlichungen zu steigern und / oder die Anzahl der bereitgestellten Dokumente – und damit auch die Attraktivität der Sammlungen – durch retrospektive Digitalisierung zu erhöhen.

### **3.3 Zusammenfassung der statistischen Analyse und der Umfrageergebnisse**

Die Ergebnisse der Angebotsanalyse wie auch der Umfrage geben ein komplementäres Bild von der aktuellen Situation der mittels den drei Open Source Software-Produkten bereitgestellten, öffentlich zugänglichen digitalen Repositorien in Indien. Danach wurden – abgesehen von einer kleinen Anzahl – die meisten der Sammlungen erst in den vergangenen 2 Jahren angelegt und beinhalten demzufolge noch vergleichsweise wenige Dokumente (s.o. FN 22). Als besonders erfolgreich erweisen sich zwei Typen von Repositorien:

- Jene institutionellen Repositorien, die klar formulierte und offensichtlich auch organisationsintern kommunizierte Sammelrichtlinien entwickelt haben und hierfür ggf. auch mehr als nur infrastrukturelle Unterstützung seitens der Organisation(sleitung) erfahren.
- Jene, die als themenorientierte Repositorien einen klar definierten fachlichen Sammel- bzw. Bereitstellungsauftrag formuliert haben.

---

27 Daraus haben sich dann auch einige wenige nicht weitergeführte Anwendungen ergeben, die das Gesamtbild verfälschen. Das produktspezifische Know-how scheint häufig an Einzelpersonen gebunden zu sein und sich allein schon deshalb einem produktbezogenen Vergleich zu entziehen. Selten wurden bislang wirkliche Produktvergleiche wie bei [Madalli / Seth 2005] auf der Grundlage von Tests veröffentlicht – überraschenderweise auch nicht bei den Asian Digital Libraries-Konferenzen.

## 4 Zusammenfassung und Schlussfolgerungen

Die Studie hat gezeigt, dass viele international orientierte indische Wissenschaftseinrichtungen sehr aktiv bei der Einrichtung von digitalen Repositorien sind. Insbesondere dort sind auch in den Bibliotheks- und Informationsabteilungen Spezialisten mit internationalen Kontakten aktiv. Dass dabei zu den in der internationalen Open Source- bzw. Open Access-Szene gängigen Produkten gegriffen wurde und wird, ist vermutlich auf eine Mischung aus Sprach- und Kulturorientierung sowie finanzielle bzw. logistische Unterstützungen zurückzuführen. Zudem funktioniert diese Szene sehr ausgeprägt über persönliche Beziehungen, so dass entweder persönliches bzw. institutionelles Engagement alleine oder in Verbindung mit strukturpolitischer Einflussnahme zur Multiplikationseffekten führt(e), als deren Ergebnis zahlenmäßig wachsende Anwendergemeinden entstanden. Angesichts der Einwohnerzahl und der Entwicklungsdynamik des Schwellenlandes Indien ist diese Anzahl im internationalen Vergleich zwar schon hoch, liegt aber noch weit unter dem tatsächlichen Potenzial und Bedarf des Landes.

Allerdings haben nur einige dieser Anwendungen einen professionellen Stand erreicht, der erwarten lässt, dass sie auch längerfristig erfolgreich existieren werden.

Vorzugsweise dort wo der anwendungstechnische Enthusiasmus Einzelner oder unterstützende Infrastruktur aus einer IT-Abteilung besteht, wurden die Anwendungen bislang im Erscheinungsbild und funktional über die Basisinstallation hinaus entwickelt. Die tatsächlichen technischen Kenntnisse zur Betreuung, Einbindung in andere Anwendungen und zur funktionalen Weiterentwicklung der konkreten Software-Installation sind in den meisten Fällen auch aus Anwenderperspektive noch nicht so ausgeprägt, dass die verfügbaren Anwendungen mehrheitlich als Vorzeigebispiele angesehen werden können. Hier scheint noch erheblicher Schulungs- und weiterführender Betreuungsaufwand notwendig zu sein, um ein konsistenteres Gesamtbild in diesem Bereich zu erzeugen. Schulungen allein für die Initialisierung der Einrichtung eines Repositoriums sind sicher ein wichtiger Schritt, ihnen müssen jedoch weiterführende, anwendungsbegleitende Unterstützungsmaßnahmen folgen, die über die Beantwortung von Fragen via e-mail-Liste hinausgehen.

In jenen Fällen, in denen Anwendungen erfolgreich realisiert wurden und auch auf einen zielgruppenspezifischen Bedarf stoßen, erfahren sie eine eindrucksvoll hohe Nachfrage und reziprok dann auch um so mehr innerorganisatorische Unterstützung. Ein sehr gutes Beispiel hierfür ist das für den Nachweis und die Bereitstellung von Forschungspublikationen angelegte

institutionelle Repository ePrints@IISc (<http://eprints.iisc.ernet.in/>) des Indian Institute of Science in Bangalore.

Insofern lässt sich die eingangs formulierte Frage nach den speziellen Rahmenbedingungen und Problemen, unter denen Repositorien in Indien eingerichtet und betrieben werden, weitgehend mit den auch international geltenden Erkenntnissen (vgl. z.B. [Gierveld 2006]) beantworten – sieht man einmal von regionalen kommunikationstechnischen Einschränkungen ab.

Als zentrale Erfolgsfaktoren sind daher zu nennen:

- Technische Kompetenz bei den verantwortlichen Betreuenden und/oder entsprechende Unterstützung innerhalb der Organisation bzw. von außen
- Verankerung der Sammel- und Bereitstellungsidee bei der relevanten Zielgruppe sowie den Führungskräften der Organisation bzw. den peers des Faches
- Sammelrichtlinien und -zielsetzungen, die in Form von mission oder policy statements formuliert und propagiert sowie von der Organisationsleitung unterstützt werden.
- Personelle und organisatorische Infrastrukturunterstützung seitens der zentralen Informationseinrichtung, im Normalfall also der Bibliothek oder einer entsprechenden fachlichen Informationabteilung.

Den initialisierenden Qualifizierungsangeboten zur Einrichtung der Repositorien müssen deshalb weiterführende, betreuende Angebote folgen, um die Nachhaltigkeit und Qualität der Sammelaktivitäten abzusichern. Dem muss durch entsprechende Basisqualifizierung in der Ausbildung von Bibliothekaren / Informationsfachleuten sicher Rechnung getragen werden, es besteht aber darüber hinaus erheblicher Weiterbildungs- und Betreuungsbedarf, der vermutlich durch die Aktivisten der Open Source-Szene nicht geleistet werden kann. Alternativ zu der von den Anwendern von Open Source-Produkten häufig erwarteten entgeltfreien Beratung aus dem Kollektiv der software- und anwendungsbezogenen Know-how-Träger könnte sich insofern zumindest für kostendeckende, ggf. sogar gewinnorientierte Beratungs- und Betreuungsdienstleistungen ein Marktsegment entwickeln. Deren Rahmenbedingungen und Chancen sind allerdings nicht nur für Indien ein noch weitgehend unerforschtes Thema. Solche Dienstleistungen könnten für die Ausbildung von Qualitätsanforderungen orientiert am DINI-Zertifizierungskonzept sicher von Vorteil sein.

Die für Indien identifizierten Qualitätsanforderungen korrespondieren bezeichnenderweise zumindest mit den ersten drei der von DINI aufgestellten Zertifizierungsanforderungen (vgl. Einführung). Sie sind eindeutig produktunspezifisch und dürften auch für die Anwendungen in Deutschland zutreffen.

Schließlich sind auch in Deutschland bislang erst 19<sup>28</sup> dieser mindestens 84 Publikationsserver von DINI zertifiziert.

## 5 Literatur

- Borghoff et al. (2005). Vergleich bestehender Archivierungssysteme, München 2005. Hrsg. von nestor, Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland (Nestor materialien ; 3) URN: urn:nbn:de:0008-20050117016
- Das, Anup Kumar; Sen, B K, Dutta, Chaitali (2005). Collection Development in Digital Information Repositories in India. – Vishwabharat@TDIL; 2005; 17; 91-96, <http://hdl.handle.net/1849/203>
- DINI-Arbeitsgruppe „Elektronisches Publizieren“ (2007): DINI-Zertifikat Dokumenten- und Publikationsservice 2007. (DINI Schriften 3-de) [Version 2.0, September 2006]; <http://edoc.hu-berlin.de/series/dini-schriften/2006-3/PDF/3.pdf>
- Harnard, Stevan; Tansley, Robert (2000). Eprints.org Software for Creating Institutional and Individual Open Archives. D-Lib Magazine, 6 (2000) 10, <http://www.dlib.org/dlib/october00/10inbrief.html#HARNAD>
- Heery, Rachel; Andersson, Sheila (2005). Digital Repositories Review. Report for JISC, 2005; [http://www.jisc.ac.uk/uploaded\\_documents/digital-repositories-review-2005.pdf](http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf)
- Ghosh, S: B.; Das, Anup Kumar (2006). Open access and institutional repositories – a developing country perspective: a case study of India. World Library and Information Congress: 72nd IFLA General Conference and Council “Libraries: Dynamic Engines for the Knowledge and Information Society” 20-24 August 2006, Seoul, Korea, [http://www.ifla.org/IV/ifla72/papers/157-Ghosh\\_Das-en.pdf](http://www.ifla.org/IV/ifla72/papers/157-Ghosh_Das-en.pdf)
- Gierveld, Heleen (2006). Considering a Marketing and Communications Approach for an Institutional Repository. – In: Ariadne 49, October 2006 <http://www.ariadne.ac.uk/issue49/gierveld/>
- Jain, P.K., Babbar, Parveen (2006). Digital libraries initiatives in India. In: The International Information & Library Review 38 (2006), 161-169
- Madalli, Devika (2003). A Digital Library of Library and Information Science using Dspace. – In: DRTC Workshop on Semantic Web, 8th-10th December, 2003, DRTC Bangalore, 7. Paper G.
- Madalli, Devika; Seth, Renu (2005). Open Source Software for Building Digital Repositories. - In: Kaul, H. K., Sen, Gayathri (Ed.): Library and Information Networking, Proceeding of the National Convention on Library and Information Networking (NACLIN 2005) held at PES Institute of Technolog, Bangalore, August 22-25, 2005 , 394-405
- Open Society Institute (2004). Open Society Institute: A Guide to Institutional Repository Software, 3rd Edition, New York, August 2004; [http://www.soros.org/openaccess/pdf/OSI\\_Guide\\_to\\_IR\\_Software\\_v3.pdf](http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf)
- Oßwald, Achim (2006). Deutsch-Indisches E-Learning-Projekt zum Thema Digitale

Bibliotheken. In: Bibliotheksdienst 40 (2006), 7, 903-905  
Sale, Arthur (2006). The Patchwork Mandate. 2006, <http://eprints.utas.edu.au/410/>  
van Westrienen; Lynch, Clifford A. (2005) Academic Institutional Repositories:  
Deployment Status in 13 Nations as of Mid 2005, D-Lib Magazine, 11(2005) 9;  
<http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>

## 6 Anlagen

### Anlage 1: Digitale Repositorien mit DSpace

1. Central Plantation Crops Research Institute, Kasargod  
<http://210.212.229.11:8080/dspace/>
2. GB Pant University of Agriculture & Technology, Pant Nagar  
<http://202.141.116.205/dspace/>
3. Indira Gandhi Institute for Development Research, Mumbai  
<http://oii.igidr.ac.in:8888/dspace>
4. INFLIBNET, Ahemdabad (nicht erreichbar)  
<http://dspace.inflibnet.ac.in/>
5. Indian Institute of Astrophysics Repository  
<http://prints.iiap.res.in/browse-title>
6. Indian Institute of Management, Kozhikode  
<http://dspace.iimk.ac.in/>
7. Indian Institute of Science, Bangalore  
Electronic Theses and Dissertations of Indian Institute of Science  
<http://etd.ncsi.iisc.ernet.in/>
8. Indian Institute of Technology, New Delhi  
<http://eprint.iitd.ac.in/dspace/>
9. Indian Institute of Technology, Kharagpur (nur Intranet Zugang)  
<http://10.17.32.208/dspace>
10. Indian National Science Academy, New Delhi (nicht erreichbar)  
<http://drtc.isibang.ac.in/insa>
11. Indian Statistical Institute, Bangalore  
<http://library.isibang.ac.in:8080/dspace/>
12. LDL: Librarians' Digital Library, DRTC  
<https://drtc.isibang.ac.in/>
13. National Centre for Radio Astrophysics, Pune (nicht erreichbar)  
<http://ncralib.ncra.tifr.res.in/dspace>
14. DSpace at National Chemical Laboratory, Pune, India  
<http://dspace.ncl.res.in/dspace/index.jsp>
15. National Institute of Oceanography, Goa  
<http://drs.nio.org/>
16. National Institute of Technology, Rourkela  
<http://dspace.nitrkl.ac.in/dspace/>

17. Raman Research Institute, Bangalore  
<http://dspace.rrri.res.in/>
18. Sri Venkateswara University, Tirupati (nicht erreichbar)  
<http://202.41.85.207:8080/dspace>
19. University of Hyderabad, Hyderabad (nicht erreichbar)  
<http://202.41.85.207:8080/dspace>
20. ICFAI BUSINESS SCHOOL (IBS), Ahmedabad  
<http://202.131.96.59:8080/dspace/>
21. Bangalore Management Academy, India (nicht erreichbar)  
<http://59.92.116.53:8080/dspace>
22. Vidyanidhi Digital Library & E-Scholarship Portal, University of Hyderabad (India); <http://dspace.vidyanidhi.org:8180/dspace/> (nicht erreichbar)

### **Anlage 2: Digitale Repositorien mit EPrint**

1. DU Eprint Archive (156); <http://eprints.du.ac.in/>
2. ePrints@IIMK: Indian Institute of Management Kozhikode Scholarship Repository (?); <http://eprints.iimk.ac.in/>
3. Indian Institute of Information Technology (?); <http://eprints.iiita.ac.in/>; (nicht erreichbar)
4. Indian Institute of Science, Bangalore, India (5751); <http://eprints.iisc.ernet.in/>
5. NAL-IR (2182); <http://nal-ir.nal.res.in/>
6. OpenMED@NIC (1293); <http://openmed.nic.in/>
7. OneWorld South Asia, Open Archive Initiative  
<http://open.ekduniya.net/>

### **Anlage 3: Digitale Bibliotheken mit Greenstone Digital Library**

1. Archives of Indian Labour  
<http://www.indialabourarchives.org/>
2. Digital Library Collections Built at NCSI  
<http://144.16.72.147/gsd/cgi-bin/library>
3. Developmental Informatics Laboratory  
<http://www.mlasia.iitb.ac.in/gsd/cgi-bin/library> (nicht erreichbar)
4. Electronic Theses and Dissertations IIT Bombay  
<http://www.library.iitb.ac.in/%7Emnj/gsd/cgi-bin/library>
5. Indian Institute of Management Kozhikode  
<http://www.iimk.ac.in/gsd/cgi-bin/library>
6. Indian Institute of Science Publications Database Bangalore  
PRABHAVI: Indian Institute of Science Publications Database  
<http://vidya-mapak.ncsi.iisc.ernet.in/cgi-bin/library?a=p&p=home>