

Alexander Wagner

Veröffentlichungsdatenbank und Volltextrepositorium

Mit Veröffentlichungsdatenbank und Volltextrepositorium bietet die Zentralbibliothek zwei wesentliche Dienstleistungen im Rahmen des wissenschaftlichen Publizierens an. Derzeit primär auf die Belange der Außerdarstellung und Verwaltung ausgerichtet, sollen diese Systeme in eine gemeinsame Datenbank überführt und dabei gleichzeitig für die Nutzer attraktiver gestaltet werden. Aufgrund ihrer über zehnjährigen Historie enthält die Veröffentlichungsdatenbank des Forschungszentrums Jülich derzeit ca. 45.000 Einträge verschiedenster Dokumententypen. Obwohl überwiegend wissenschaftliche Veröffentlichungen in Fachzeit-

schriften nachgewiesen werden, sind auch Poster, Abstracts, Lehrmaterialien, Buchbeiträge etc. enthalten. Damit wird das gesamte Spektrum des wissenschaftlichen Publikationswesens abgebildet.

Die Veröffentlichungsdatenbank entstand 2000–2002 primär zur Darstellung des wissenschaftlichen Outputs des Forschungszentrums. Hierbei sind diverse Formate zu bedienen. Die Darstellung der eigenen wissenschaftlichen Tätigkeit im Rahmen der institutionellen Webauftritte ist hier eine besonders prominente, aber natürlich nicht die Einzige. Hierzu werden die jeweils relevanten Daten passend aggregiert und in normierter und in



formatierter Form an das lokale Content Management System geliefert.

Daneben dient die Veröffentlichungsdatenbank aber auch zahlreichen Berichten und Statistiken als Basis. So ist für den wissenschaftlichen Ergebnisbericht des Forschungszentrums Jülich jährlich eine entsprechende Bibliographie zu erstellen. Diese entspricht weitestgehend den an Hochschulen üblichen Hochschulbibliographien. Darüber hinaus werden in diesem Rahmen, sowie zur Meldung an die Helmholtz-Gemeinschaft Deutscher Forschungszentren e. V. im Rahmen der programmorientierten Förderung, aber auch zahlreiche statistische Kennzahlen ermittelt.

Diese dienen sowohl dem Leistungsnachweis der Einrichtung und der Institute als auch zur weiteren strategischen Planung. Daher enthält die Veröffentlichungsdatenbank zahlreiche Statistikschlüssel wie z. B. detaillierte Instituts-, Programm- oder Dokumentzuordnungen. Da im Rahmen der Evaluation u. a. auch bibliometrische Analysen eingesetzt¹, und hierzu Publikationen auf verschiedensten Ebenen (Institute, Programme ...) aggregiert werden, wurden ferner entsprechende Schlüssel eingefügt, die eine Verknüpfung der Veröffentlichungsdatenbank mit den gängigen Zitationsdatenbanken erlauben. Dies gestattet außerdem eine gezielte Analyse des Zeitschriftenportfolios auf Basis der zitierten Arbeiten und kann so im Rahmen des Bestandsaufbaus wertvolle Dienste leisten.

Über den lokalen Linkresolver besteht eine Verknüpfung der Veröffentlichungsdatenbank mit den Angeboten der Verlage, sodass die dort publizierte Version eines Artikels abgerufen werden kann. Gibt es ein Dokument auf dem Volltextserver JUWEL, wird auch diese entsprechend verknüpft.

Da zur Einführung der Veröffentlichungsdatenbank kein passendes Standardprodukt zur Verfügung stand, das all diese verschiedenen Anforderungen befriedigend erfüllen konnte, handelt es sich um eine Eigenentwicklung in Zusammenarbeit mit Logic-Works auf Basis von Standardsoftware, im wesentlichen Oracle als Datenbank, MS Access als Eingabeoberfläche in der Bibliothek und Webformulare. Hierdurch konnte auf zahlreiche lokale Besonderheiten eingegangen und einige Benutzerwünsche erfüllt werden. Allerdings ist im Laufe der Zeit durch immer wieder neue und steigende Anforderungen ein sehr komplexes Datenmodell gewachsen. Da die Helmholtz-Gemeinschaft Deutscher Forschungszentren e. V. zu den Erstunterzeichnern der

Berliner Erklärung² gehört, wurde im Februar 2005 neben der Veröffentlichungsdatenbank ein institutionelles Repositorium konzeptioniert und am 11. Juli 2006 freigeschaltet. Im Rahmen des Jülicher OpenAccess Modells³ handelt es sich hierbei um ein reines Volltextrepositorium zur Aufnahme von Zweitveröffentlichungen. JUWEL (Jülicher Wissenschaftliche Elektronische Literatur) enthält derzeit über 4.300 Dokumente, die weltweit frei zugänglich und über die einschlägigen Suchmaschinen wie BASE, aber auch Google et al findbar sind. Für die Realisation von JUWEL wurde ein pragmatischer Ansatz gewählt, der sich an im OpenAccess-Bereich üblichen internationalen Standards (DublinCore und OAI-PMH^{4,5}) orientiert und auf die freiwillige Mitwirkung der Autoren vertraut. Nach umfangreichen Tests verschiedener vorhandener Softwarelösungen wurde das Repositorium schließlich auf Basis von DSpace⁶ realisiert^{7, 8}.

Im Gegensatz zur Veröffentlichungsdatenbank können sich für JUWEL alle Mitarbeiter des Forschungszentrums zur Dateneingabe anmelden und ihre Dokumente in das Repositorium übertragen. Der allerdings meist genutzte Weg zur Erfassung ist die Übersendung des Dokuments an die Zentralbibliothek, die nach Überprüfung der urheberrechtlichen Bestimmungen die Katalogisierung und Freischaltung übernimmt. Da insbesondere die Eingabe von Daten in die Veröffentlichungsdatenbank besonderer Sorgfalt bedarf, sind hierfür in den Instituten sog. Institutssachbearbeiter zuständig. Sie sind auch für die Bibliothek zentrale Ansprechpartner, die ggf. nötige Korrekturen vornehmen, wenn bei der Überprüfung durch die Bibliothek, z. B. unvollständige Angaben, auffallen. Anfällig ist das System für Dubletten, da hier nur eine einfache manuelle Kontrolle möglich ist. Eine Unterstützung durch die Datenbank steht hier seit Längerem auf der Desideratenliste. Wird ferner neben der rein bibliographischen Verzeichnung einer Veröffentlichung auch ein Volltext angeboten, so bedingt die Trennung von Veröffentlichungsdatenbank und

¹ Rafael Ball und Dirk Tunger. Bibliometrische Analysen – ein neues Geschäftsfeld für Bibliotheken? B.I.T.online, 7 (4): 272–278, 2004.

² Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. October 2003.

³ Jülicher OpenAccess Modell. http://www2.fz-juelich.de/zb/oa_model.

⁴ Dublin Core Metadata Initiative. <http://dublincore.org/>.

⁵ The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

⁶ DSpace. <http://www.dspace.org/>.

⁷ Ulrike Burkard und Bernhard Mittermaier. Konzeption, Implementierung und Betrieb eines Volltextservers am Forschungszentrum Jülich. Vortrag bei der ASpB-Tagung 2007 an der TU Berlin, 25.-28.09.2007, 2007.

⁸ Ulrike Burkard and Waldemar Hinz. JUWEL – Open Access Server des FZ Jülich. JUWEL – Open Access Server des FZ Jülich, 2007.

JUWEL derzeit eine doppelte Datenhaltung in zwei verschiedenen Internformaten mit der Notwendigkeit einer Synchronisation, um im Web auch einen direkten Volltextlink anzeigen zu können. Hier ist z.T. auch die manuelle Neuerfassung von Daten, die für die Veröffentlichungsdatenbank bereits katalogisiert wurden, notwendig. Da JUWEL auf Dublin Core beschränkt ist, ist eine Übergabe von Daten an die Veröffentlichungsdatenbank nicht möglich.

Weiterhin sind beide Datenbanken derzeit nur sehr begrenzt in den wissenschaftlichen Arbeitsprozess eingebunden, da die Im- und Exportfunktionalitäten beider Systeme aufgrund ihrer Geschichte beschränkt sind. Insbesondere der Wunsch, die Veröffentlichungsdatenbank enger mit dem eigenen Publikationsprozess verbinden zu können, ist im Hause seit längerem ein Desiderat und wird nun auch vermehrt von den Wissenschaftlern an die Bibliothek herangetragen. Mindestens ist eine enge Verzahnung mit den

mittlerweile zahlreich eingesetzten Literaturverwaltungsprogrammen gewünscht. Aber auch ein integrierter Publikationsworkflow steht hoch auf der Wunschliste der Jülicher Wissenschaftler. Mitte 2009 wurde daher ein Angebot eingeholt, um einige dieser Funktionalitäten in der jetzigen Veröffentlichungsdatenbank nachzurüsten. Bei der Ausarbeitung dieses Angebots stellte sich heraus, dass umfangreiche Änderungen nötig würden, die nicht unerhebliche Kosten nach sich zögen. Ausgehend von der bewährten Kooperation mit dem Deutschen Elektronensynchrotron (DESY) schon beim Betrieb der jetzigen Veröffentlichungsdatenbank, wurde als Alternative eine komplette Umstellung von Veröffentlichungsdatenbank und JUWEL auf ein neues, noch zu findendes System in Erwägung gezogen. Ziel konnte hier nur sein, die vorhandene proprietäre Lösung durch eine verfügbare Standardlösung zu ersetzen und ein möglichst weit verbreitetes flexibles Internformat einzuführen, um den nötigen Wartungs- und Einarbeitungsaufwand zu minimieren. Auch eine Konsolidierung der Softwarebasis wurde angestrebt, da dies auf Nutzerseite ein einfacheres Handling ermöglichen und auch der lokalen EDV Entlastung bringen sollte. Um die üblichen Nachweisinstrumente (z. B. BASE) zu bedienen, muss es weiterhin möglich bleiben, die Volltexte von JUWEL getrennt via OAI-PMH zu exponieren.

Da ferner im Rahmen von EU-Projekten entstandene Arbeiten ebenfalls getrennt exponiert wer-

den müssen (OpenAIRE), ist hierfür eine weitere OAI-PMH-Quelle notwendig. Der Aufbau einer zusätzlichen Datenbank mit erneuter Duplizierung der Daten sollte aber von vornherein vermieden werden.

Zur Realisation dieser komplexen Anforderungen wurden zwei Szenarien untersucht: Einerseits der Ausbau des bereits bestehenden DSpace-Systems um die nötigen Funktionalitäten, insbesondere Statistikschlüssel und Normsatzfunktionalitäten, andererseits der Aufbau eines komplett neuen Systems auf Basis des vom CERN entwickelten Invenio⁹, welches am DESY für die dortige Inspire-Instanz¹⁰ bereits eingesetzt wird. Nach eingehender Prüfung beider Lösungen sowohl von bibliothekarischer Seite als auch auf der technischen

Ebene durch die beteiligten EDV-Abteilungen, fiel die Entscheidung zugunsten von Invenio. Kurz vor Projektbeginn stieß mit der Deutschen Gesellschaft für Schwerionen-

forschung (GSI) noch ein weiterer Partner hinzu. Dort ist, neben den in Jülich und Hamburg bereits definierten Funktionalitäten, auch der lokale Bibliothekskatalog abzubilden, eine Anforderung, die auf Basis von DSpace nicht erfüllbar wäre. Im Laufe des Jahres 2011 kristallisierte sich außerdem eine Zusammenarbeit mit der Hochschulbibliothek der RWTH Aachen heraus, zunächst um die gemeinsame JARA-Publikationsdatenbank¹¹ einzurichten, mittelfristig aber auch die dortige Hochschulbibliographie sowie den Hochschulschriftenserver der RWTH Aachen auf Invenio umzustellen.

Im Gegensatz zu DSpace, oder anderen im Bereich Schriftenserver eingesetzten Lösungen, arbeitet Invenio intern auf Marc 21-Datensätzen, und damit auf einem weit verbreiteten Format, das auch zur Beschreibung komplexer Dokumente geeignet ist¹². Die Realisation von Inspire demonstriert ferner die Tauglichkeit der Software für die Verwaltung großer Datenbestände und auch die korrekte Abbildung der im Nutzungskontext notwendigen zahlreichen Dokumenttypen und Formate. Die Verwendung von Marc 21 eröffnet die Möglichkeit, einfach Daten auszutauschen und Fremddaten zu übernehmen. Auch für die nötigen Erweiterungen jenseits der rein bibliographischen Daten, wie Statistikschlüssel etc., bietet Marc

9 Invenio. <http://invenio-software.org/>.

10 Inspire. <http://www.projecthepinpire.net/>.

11 Jülich Aachen Research Alliance (JARA). <http://www.jara.org/>.

12 Marc Standards. <http://www.loc.gov/marc/>.



21 einfache Lösungen über zahlreiche freie Feldbereiche (z. B. Kategorie 9xx).

Diese sind bei Bedarf flexibel erweiterbar, ohne dass Änderungen am Systemkern erforderlich sind. Lediglich die entsprechenden Auswerteroutinen müssen ggf. angepasst werden, was aber unabhängig von der Datenerfassung und auch jetzt bereits der Fall ist. Insgesamt sind also fast alle Anpassungen von Invenio an die lokalen Bedürfnisse auf einer relativ hohen Ebene angesiedelt, es sind keine eigenen Entwicklungen erforderlich. Durch wesentlich feinere Felderungen in Marc 21 entfallen die schmerzlichen Einschränkungen des in DSpace verwendeten Dublin Cores wie z. B. bei Zeitschriftenzitat. Diese werden derzeit in JUWEL in einem vorgegebenen Format innerhalb eines Textfeldes erfasst und müssten bei der anstehenden Konversion entsprechend aufgespalten werden. Dieses Vorgehen ist, trotz bibliothekarischer Exaktheit, vergleichsweise fehleranfällig. Hilfreich ist hier bei der Migration die deutlich bessere Datenqualität der Veröffentlichungsdatenbank sowie die zusätzlich erfassten Schlüssel, die ggf. eine Ergänzung fehlender oder Korrektur falscher Daten erlauben.

Auch Projektzuordnungen, Autoren- und Institutsidentifikationen etc. stellen für die Statistiken aufgrund von Institutsneuausrichtungen und der Zunahme virtueller Institute eine besondere Herausforderung dar, können aber in Marc Authority¹³ gut abgebildet werden, da hier sowohl vertikale als auch horizontale Verknüpfungen darstellbar sind. Eine entsprechende Erweiterung von Invenio findet unter Beteiligung der Projektpartner derzeit am CERN statt¹⁴ und ist in der ersten Phase bereits abgeschlossen. Invenio ist ferner bereits in Hinblick auf Ein- und Ausgabeschnittstellen, Suche, graphische Oberfläche, soziale Funktionalitäten wie Kommentare und Tagging, Personalisierung, lokale Ablagen etc. sehr gut ausgestattet, sodass mit Freischaltung des Systems für die Nutzer ein echter Mehrwert entsteht und nicht nur eine neue, andere Oberfläche. Auf der Eingabeseite konnte durch Schnittstellen zu CrossRef eine Übernahme des weitgehenden Zitats auf Basis der DOI realisiert werden. Eine, insbesondere für das DESY relevante Anbindung von arXiv.org ist bereits im Prototyp entwickelt, die Datenübernahme von anderen freien Quellen ist ebenfalls reali-

siert oder prototypisch verfügbar. Dadurch ist eine Reduzierung der Fehler bei der Dateneingabe zu erhoffen. Da Invenio ferner über einen feldbasierten und algorithmischen Dublettencheck verfügt, ist auch bei der Dublettenproblematik eine Verbesserung möglich. Die grundlegend andere Struktur von Marc 21 im Vergleich zu den bisher verwendeten Formaten, macht allerdings die Konversion der Altdaten vergleichsweise aufwendig, zumal hier im laufenden Betrieb gearbeitet werden muss. Hierbei werden jedoch gleichzeitig bestehende Desiderate eingearbeitet, fehlende Daten ergänzt und es findet eine Dublettenbereinigung statt, sodass hier die Datenqualität insgesamt angehoben wird. Im Projektkontext wurden die Zeitschriftendatensätze, die für die Evaluationen besonders wichtig sind, überarbeitet und Mechanismen entwickelt, um hier relevante Informationen wie Fachzuordnungen oder die Indexierung in Fachdatenbanken mit deutlich niedrigerem Aufwand zu erfassen und zu pflegen als dies derzeit der Fall ist. Die Zusammenlegung von JUWEL und Veröffentlichungsdatenbank in einem neuen gemeinsamen System und die zahlreichen funktionalen Erweiterungen lassen für die Nutzer eine deutliche Verbesserung sowohl bei der Erfassung, der Suche als auch der Nachnutzung der Daten erwarten. Es steht zu hoffen, dass mit diesen Verbesserungen, die in zahlreichen Gesprächen mit den Wissenschaftlern erörtert wurden, die Veröffentlichungsdatenbank vom Verwaltungswerkzeug zu einem integralen Arbeitsbestandteil werden kann.

AUTOR

DR. ALEXANDER WAGNER

Fachinformationsmanager
a.wagner@fz-juelich.de

¹³ Marc 21 Format for Authority Data. <http://www.loc.gov/marc/authority/>.

¹⁴ Christopher Michael Dickinson, Jean-Yves Le Meur, Abou Khaled Omar und Pascal Felber. Authority Control for INVENIO. PhDthesis, Fribourg, EIA, 2011. Presented 07. Sep. 2011.