Vorläufige Version zur Prüfung der textlichen Inhalte und Positionierung der Bilder. Vor Onlinestellung erhalten Sie ein gesetztes PDF zur endgültigen Freigabe.

Forschungspolitik

# PubMan – ein *Repository* für die MPG

Tschida, Ulla

Max Planck Digital Library, München

Korrespondierender Autor
tschida@mpdl.mpg.de

Zusammenfassung
PubMan (http://pubman.mpdl.mpg.de) ist eine Anwendung, die es Mitarbeitern von Forschungsorganisationen ermöglicht, Publikationsdaten nachhaltig zu speichern, zu verwalten und anzureichern sowie für andere Webdienste zur Verfügung zu stellen. Die Anwendung basiert auf der eSciDoc Infrastruktur, einem Gemeinschaftsprojekt der MPG und des FIZ Karlsruhe. Die Abteilung Forschung & Entwicklung in der MPDL entwickelt die Anwendung und betreut die Nutzer der Max-Planck-Institute. Die Anwendung wird seit Mai 2009 in der MPG eingeführt und das aktuelle *Repository* eDoc (http://edoc.mpg.de) sukzessive ablösen.

*Abstract*
*PubMan (http://pubman.mpdl.mpg.de) is an application which allows members of a research organisation to store, manage and enrich their publications and provide the data for re-use by other web-services. The solution is based on the eSciDoc infrastructure, a joint project by the MPS and the FIZ Karlsruhe. The Department of Research & Development at the MPDL develops the solution and supports the users at the Max Planck Institutes. PubMan is introduced in the MPS since May 2009 and will gradually replace the current repository eDoc (http://edoc.mpg.de).*

**Die Publikation im Mittelpunkt …**
PubMan ist eine web-basierte Anwendung, die allen aktuellen und ehemaligen Mitarbeitern der MPG als nachhaltiges Repository (digitales Archiv) für Publikationen und ergänzende Materialien (*Supplementary Material*) dient. Der Nachweis von Publikationen, die Bereitstellung der eigentlichen Inhalte sowie ausreichende semantische und formale Beschreibungen für entsprechende Suchanfragen sind im IT-gestützten wissenschaftlichen Betrieb eine nicht mehr wegzudenkende Größe. Je nach Wissenschaftsdisziplin werden "Publikationen" in unterschiedlichen Formen und Ausprägungen verfasst, von traditionellen Artikeln in Fachzeitschriften über Preprints und Arbeitspapieren bis hin zu Blog- und Wiki-Einträgen. Diese verschriftlichten Ergebnisse werden bereits während des Forschungsvorhabens unter thematischen und/oder interdisziplinären Gesichtspunkten zusammengestellt. Insofern sind auch die Bedürfnisse für die Nutzung der bereitgestellten Dokumente unterschiedlich: Autoren, wissenschaftliche Fachkollegen, Forschungskoordinatoren, Fördergremien, Institutsdirektoren, Bibliotheken, lokale Webmaster und die allgemeine Öffentlichkeit stellen als potentielle Nutzer jeweils unterschiedliche Anforderungen an die Beschreibung und mögliche Nachnutzung der Dokumente bzw. der Inhalte. Für das Design der Anwendung und die Ausgestaltung der darunter liegenden IT- Forschungsinfrastruktur sind die multiplen Nutzungsszenarien ausschlaggebend. Im Unterschied zu *Bibliographic Management Systems* wie z.B. EndNote oder

Vorläufige Version zur Prüfung der textlichen Inhalte und Positionierung der Bilder. Vor Onlinestellung erhalten Sie ein gesetztes PDF zur endgültigen Freigabe.

RefMan, die den Schwerpunkt auf die Verwaltung von bibliographischen Referenzen für einen Einzelnutzer legen, werden in dem *Repository* PubMan sämtliche Charakteristika der Entität "Publikation" ausgezeichnet (wie z.B. Autoren, Organisationen, Volltexte, Metadaten), um möglichst unterschiedliche Nachnutzungsszenarien bzw. Darstellungsweisen zu unterstützen. PubMan ist somit als Teil einer langfristigen eResearch Infrastruktur der MPG zu verstehen, die die nachhaltige Verfügbarkeit sowie den freien bzw. kontrollierten Zugang zu Publikationen gewährleistet.

**… und in Verbindung mit Forschungsdaten**

Die Bereitstellung der für eine Publikation verwandten Daten ist für die Wissenschaft attraktiv, da so Ergebnisse nachvollzogen oder neue Fragestellungen bearbeitet werden können.

Die verwendeten Daten können in verschiedensten Formen vorliegen (Simulationen, statistische Auswertungen, Tabellen, Grafiken, Algorithmen etc.) und liegen in den meisten Fällen auf verteilten Servern. Die ständig wachsende Anzahl an wissenschaftlich relevanten Daten bedeutet auch, dass in vielen Fällen aggregierte Sichten auf Daten nötig sind, z.B. Registraturen, die es den Wissenschaftlern erleichtert, herauszufinden, welche Informationen überhaupt verfügbar sind. Selbstverständlich spielen dabei Fragen der Authentifizierung und Authorisierung eine wichtige Rolle, um die Rechte der Autoren zu schützen. Eine IT-Forschungsinfrastruktur für die MPG muss insofern in der Lage sein, verschiedene Objekte zu verbinden und die Relationen sowie deren Bedeutung über standardisierte Schnittstellen auch in maschinen-lesbarer Form bereitzustellen. Die Anwendung PubMan ist somit ein Baustein der Forschungsinfrastruktur, der insbesondere der Beschreibung und Darstellung von Publikationen dient, aber darüber hinaus ebenso die Möglichkeit bietet, die archivierten Publikationen mit ergänzenden Materialien (*Supplementary Material*) in Formaten wie z.B. Excel, SPSS, Audio, Video etc. anzureichern. Überdies adressieren zwei weitere eSciDoc Anwendungen die Dissemination und die verteilte Bearbeitung von spezifischen Forschungsdaten, z.B. Bilder [1] oder Digitalisate [2].

**PubMan, eSciDoc und Open Access**

Im Zusammenhang mit der *Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen* 2003 [3] beschloss die MPG den Aufbau der ZIM Projektgruppe (*Heinz Nixdorf Center for Information Management* 2002 – 2006) bzw. der MPDL (*Max Planck Digital Library*, seit 2007), deren einzelne Projekte wie die Zeitschriften-Familie Living Reviews [4] oder das erste *Repository* der MPG eDoc [5] Open Access fördern und unterstützen.

Mit dem Großprojekt eSciDoc stellte die MPG die Weichen für eine Forschungsinfrastruktur, um der zunehmenden Bedeutung von IT-Dienstleistungen für exzellente Forschung Rechnung zu tragen. eSciDoc, ein gemeinsames Projekt der MPG und des Fachinformationszentrums Karlsruhe (FIZ Karlsruhe) [6], dient dem Aufbau einer nachnutzbaren Infrastruktur und einzelner Anwendungen zur Unterstützung der wissenschaftlichen Kommunikation und Publikation in Forschungsorganisationen. Konkret soll der langfristige und offene Zugang zu Forschungsresultaten und Forschungsmaterialien (Publikationen und Primärdaten) gewährleistet sowie innovative Formen der wissenschaftlichen und interdisziplinären Zusammenarbeit unterstützt werden. Das Projekt ist eingebettet in die langfristige Strategie der MPDL, exzellente Services für exzellente Forschung zu schaffen. Diese Anforderung adressiert das Projekt auf zwei Ebenen: Die Infrastruktur konzentriert sich auf grundlegende Standards und Technologien (Web-Dienste), die nötig sind für das Management und den langfristigen Erhalt von umfangreichen Daten, die wiederum aus verschiedensten Objekttypen (Datentypen) und deren Relationen bestehen können. Die einzelnen Anwendungen hingegen unterstützen spezifische wissenschaftliche Arbeitsabläufe, die die Wissenschaftler als Endnutzer während ihrer jeweiligen Forschungsfragen begleiten. Um dabei dem Anspruch gerecht zu werden, eine allgemein nachnutzbare Infrastruktur zu bauen, die gleichzeitig möglichst vielfältige disziplin-spezifische Anwendungen erlaubt, wurde ein service-orientierter Architekturansatz gewählt. Dieses Vorgehen ermöglicht das Zusammenspiel von verschiedenen modular aufgebauten Web-Diensten, die allgemein gültige

 © 2009-2010 Max-Planck-Gesellschaft

Vorläufige Version zur Prüfung der textlichen Inhalte und Positionierung der Bilder. Vor Onlinestellung erhalten Sie ein gesetztes PDF zur endgültigen Freigabe.

Funktionalitäten adressieren (z.B. Suche oder Export) und somit für verschiedene spezifische Anwendungen genutzt werden können.

**Was kann PubMan?**

Für die Dateneingabe stehen verschiedene Möglichkeiten zur Verfügung: Metadaten können in kurze oder detaillierte genre-spezifische Masken eingegeben werden. Zusätzlich können Metadaten aus lokalen oder externen Anwendungen (EndNote, BibTex, RIS, WoS) importiert oder von anderen Archiven (arXiv, BioMed Central, Spires, PubMed Central) zusammen mit Volltexten geladen werden.

Publikationsdaten können von registrierten Nutzern – mit persistenten Identifikatoren versehen – weltweit freigeschaltet werden. Die Volltexte können dabei unterschiedliche Zugriffsrechte erhalten. Bei Bedarf können auch zusätzliche Stufen der Qualitätssicherung integriert werden, um erst nach einer Prüfung auf Vollständigkeit sowie Korrektheit die Freischaltung vorzunehmen.
Besonderes Augenmerk wurde in der Konzeption und Planung auf Aspekte der Nachnutzung von Daten, d.h. der Metadaten und Komponenten wie z.B. Volltexte oder *Supplementary Material*, gelegt. Die Implementierung einer Standard-Schnittstelle für Suche und Export von Daten [7] ermöglicht es Menschen und Maschinen, Publikationen nach definierten Abfragen in bestimmten Formaten, Zitierstilen und Sortierungen auszugeben. Diese Funktionalität unterstützt die einfache Einbindung von Publikationslisten in Webseiten oder die individuelle Generierung von Reports nach bestimmten Abfragen.

Durch die Einbindung des separaten Webdienstes CoNE zur Verwaltung von kontrollierten Entitäten wie z.B. Personen [8] werden individuelle *Researcher* Portfolios angeboten, die – neben optionalen persönlichen Angaben zum Curriculum Vitae und Forschungsfeldern – eine automatisch generierte Publikationsliste beinhalten. Durch die Implementierung eines Plug-in für die Blog-Software WordPress können Publikationsdaten aus PubMan auch dynamisch in persönliche oder projekt-spezifische Blogs integriert [9].

Zusätzlich unterstützt eine aus CoNE automatisch generierte Vorschlagsliste den Nutzer bei der Eingabe von Personennamen oder Zeitschriftentiteln.

Außerdem stellt das *Repository* selbst entsprechende Einstiegspunkte in die Publikationsaktivitäten der einzelnen Institute bzw. deren Abteilungen und Projekte bereit, die je nach Institutsbedürfnissen organisiert sind. Einzelne Datensätze oder Listen können in verschiedenen Formaten und Layouts ausgegeben werden. Zur Zeit werden die Formate EndNote, BibTeX, PDF, RTF, HTML, XML und ODT sowie die Zitationsstile APA und AJP unterstützt.
Grundlegende Funktionen eines *Repositories*, wie die Vergabe von persistenten Identifikatoren (PIDs), automatische Versionierung und standardisierte Beschreibung der Daten für Dissemination und Langzeit-Archivierung sind durch die Anbindung an die eSciDoc Infrastruktur gewährleistet.

Ein detaillierter Überblick der momentan implementierten PubMan Funktionalitäten ist im MPDL MediaWiki CoLab zu finden [10].

**Wo steht PubMan heute?**

Im Mai 2009 wurde eine fachlich "runde" Version den Bibliotheken vorgestellt, die in enger Abstimmung mit fünf *Early Adopter* Instituten fertig gestellt werden konnte. Seitdem wird der Dienst schrittweise weiterentwickelt und parallel an einzelnen Instituten eingeführt. Die Einführung des Systems bedeutet in den meisten Fällen auch die parallele Migration der Publikationsdaten vom bestehenden *Repository* eDoc. Bereits seit April 2009 arbeitet das MPI für Psycholinguistik ausschließlich mit PubMan. Das

Vorläufige Version zur Prüfung der textlichen Inhalte und Positionierung der Bilder. Vor Onlinestellung erhalten Sie ein gesetztes PDF zur endgültigen Freigabe.

Institut entwickelte einen neuen Web-Auftritt und entschied sich für den Einsatz von PubMan als einzige Quelle für Publikationen. Wissenschaftler und Bibliothek verwalten gemeinsam die Publikationsdaten und Volltexte in PubMan und profitieren von der automatischen Integration der Daten in lokale Webseiten. Durch den Anreiz einer immer aktuellen Homepage stieg sowohl das Bewusstsein für die Datenqualität als auch für die Möglichkeiten der Bereitstellung der Volltexte als Open Access. Das MPI für Gravitationsphysik (AEI) nutzt PubMan seit Herbst 2008 und koordinierte die schrittweise Einführung am Campus Golm. Die Migration der Daten des AEI aus eDoc wurde im Dezember 2009 abgeschlossen. Das MPI für Chemische Ökologie beeinflusste die fachliche Weiterentwicklung des Webservice CoNE zur Verwaltung von Personen und deren Zugehörigkeit zu bestimmten Gruppen und Abteilungen innerhalb eines Institutes.

Gleichzeitig arbeitet die MPDL an disziplinspezifischen Erweiterungen für PubMan: Zusammen mit dem MPI für Evolutionäre Anthropologie werden spezifische Funktionalitäten für linguistische Textsammlungen entwickelt. Um die Besonderheiten von juristischen Publikationsformen zu berücksichtigen, erweitert die MPDL in Kooperation mit den rechtwissenschaftlichen Instituten die Möglichkeiten für die Datenein- und ausgabe.

Die Software der Anwendung sowie die zugehörigen Web-Services stehen Entwicklern und/oder Forschungsorganisationen zur gemeinsamen Weiterentwicklung als Open Source zur Verfügung.

Zusammen mit dem Partner FIZ Karlsruhe, das die Basisdienste der Infrastruktur entwickelt, wurden auf zahlreichen Veranstaltungen und Konferenzen interessierte Nachnutzer der Infrastruktur und der Anwendung PubMan gefunden. Mit Jahresende 2009 evaluieren ca. 10 Institutionen aus dem In- und Ausland die Nachnutzung und verteilte Weiterentwicklung der Anwendung. Über eine eigene Mailingliste wird externen Entwicklern Beratung und Hilfestellungen bei Installation und Set-up der eSciDoc Infrastruktur sowie der Anwendung PubMan gegeben.
Nach einer sehr erfolgreichen Evaluierung der Anwendung PubMan durch den Fachbeirat der MPDL (Juli 2009) wurde das technische Design der Anwendung und Infrastruktur zusätzlich durch drei Institutsdirektoren für sehr gut befunden (August 2009). Im Besonderen wurde das Potential der Infrastruktur für den weiteren Ausbau hervorgehoben.

**Kontakt, Links**

Die Abteilung Forschung & Entwicklung in der MPDL unter der Leitung von Malte Dreyer entwickelt die Anwendung und betreut die Nutzer der Max-Planck-Institute. Die Produktivversion ist unter http://pubman.mpdl.mpg.de/pubman/ einsehbar, das Suchen und Blättern ist auch ohne Login möglich. Für Demonstrationszwecke und erste Gehversuche mit PubMan können Sie die Testversion unter http://test-pubman.mpdl.mpg.de/pubman/ (Nutzername: demo, Passwort: demo) besuchen.
Sollten Sie detaillierte Informationen benötigen, hilft man Ihnen unter pubman-support@gwdg.de gerne weiter.

**Drittmittelfinanzierung**

Bundesministerium für Bildung und Forschung (BMBF), 2004-2009

**Referenzen und weiterführende Links**

[ 1 ] FACES Collection of the MPI for Human Development: http://faces.mpdl.mpg.de/faces/

 © 2009-2010 Max-Planck-Gesellschaft

Vorläufige Version zur Prüfung der textlichen Inhalte und Positionierung der Bilder. Vor Onlinestellung erhalten Sie ein gesetztes PDF zur endgültigen Freigabe.

[ 2 ] "Virtueller Raum Reichsrecht" Collection of the Max Planck Institute for European History of Law: http://virr.mpdl.mpg.de/

[ 3 ] Conference on Open Access to Knowledge in the Sciences and Humanities 20 – 22 Oct 2003, Berlin: http://oa.mpg.de/openaccess-berlin/berlindeclaration.html

[ 4 ] Living Reviews. Online-only refereed review journals. An open access service: http://livingreviews.org

[ 5 ] Max Planck Society eDoc Server: http://edoc.mpg.de/

[ 6 ] Neben der Unterstützung durch Eigenmittel der beiden Projektpartner wurde das Projekt über einen Zeitraum von 2004 – 2009 vom Bundesministerium für Bildung und Forschung (BMBF) gefördert.

[ 7 ] eSciDoc Search&Export Service (REST): http://pubman.mpdl.mpg.de/search/SearchAndExport_info.jsp

[ 8 ] eSciDoc CoNE – Control of Named Entities: http://pubman.mpdl.mpg.de/cone/

[ 9 ] Ein weiteres Beispiel für ein themen-spezifisches Blog findet man hier: Reinhold von Sengbusch: http://sengbusch.blogs.mpdl.mpg.de/

[ 10 ] PubMan Functionalities: http://colab.mpdl.mpg.de/mediawiki/PubMan_Functionalities