

eSciDoc – multimediale „Digital-Library-Systeme“ für wissenschaftliche Publikation und Forschungsdatenmanagement

Helga Bergmann

Im Projekt eSciDoc haben die MP-DL und FIZ Karlsruhe organisatorische und technische Grundlagen für multimediale, vernetzte Bibliotheks- und Forschungsanwendungen geschaffen. Seit Mitte 2008 stellen sie ihre Erkenntnisse, Konzepte und Software der internationalen Wissenschafts-Community als Open Source zur Verfügung – und haben damit internationale Weiterentwicklungen angestoßen. Bibliotheken und Anwender aus verschiedenen Wissenschaftsdisziplinen waren frühzeitig eingebunden: Einer der ersten Kooperationspartner war die VZG, die Verbundzentrale des Gemeinsamen Bibliotheksverbundes (GBV). Heute, gut anderthalb Jahre nach Auslaufen der Förderung von eSciDoc durch das Bundesministerium für Bildung und Forschung und der Projektöffnung für die Community unterstützen nicht nur die großen Wissenschaftsorganisationen, sondern auch nationale Initiativen vieler Länder sowie die Europäische Union weitere Entwicklungsprojekte zum Aufbau der künftigen Infrastruktur für das durchgängig digital gestützte wissenschaftliche Arbeiten: die Voraussetzung für eScience und eResearch.



Fachleute für eScience Infrastruktur: Matthias Razum, eSciDoc Projektleiter bei FIZ Karlsruhe, Dr. Leni Helmes, Leiterin der Entwicklung im FIZ und die Bibliothekarin Dr. Mikiko Tanifuji. Tanifuji baut auf Basis von eSciDoc Software die digitale Bibliothek der japanischen Materialwissenschaftler auf.

eScience (enhanced science) bezeichnet eine durchgängige digitale Kollaborations- und Informationsinfrastruktur, durch die Forschungsarbeit effizienter gestaltet werden kann. Forscherinnen und Forscher können von überall auf ihren Arbeitsplatz zugreifen.

eResearch hat sich als Fachbegriff für die neue Art der Forschung auf Basis digital verfügbarer Daten und Informationen etabliert: einer Forschung, die überhaupt erst durch die Digitalisierung der Informationsquellen und ihre Verbindung mit spezifischen Analysetools möglich wird. Zwei Beispiele zeigen eindrucksvoll, wie sich die wissenschaftlichen Arbeitsbedingungen dadurch verändern:

1) Wissenschaftlerinnen und Wissenschaftler müssen für Laborexperimente in Zukunft nicht mehr immer ins Labor gehen. Sie führen diese am Computer durch, entscheiden, welche Datensätze bereitgestellt werden

und wer sie einsehen darf. Sie haben von jedem Ort der Welt aus Zugriff auf die von ihnen selbst oder von anderen Forscherinnen und Forschern erarbeiteten Daten, die weitgehend automatisch von Forschungsdatenmanagement-Systemen dokumentiert und gespeichert wurden. Jeder Wissenschaftler, jede Wissenschaftlerin kann, wenn er möchte, (zumindest theoretisch) seine Ergebnisse selbst direkt publizieren, also auch ohne den Zwischenschritt über einen Verlag und/oder eine Bibliothek/Informationsabteilung.

2) Archäologen katalogisieren ihre Ausgrabungen direkt am Fundort mithilfe von tragbaren Minicomputern; Smartphones, Handhelds oder speziell für die jeweilige Disziplin entwickelten Endgeräten für mobile Erfassung und Informationsmanagement. Digitale Fotos, die Abmessungen des Fundstücks, geografische Angaben zur Fundstelle und die Anmerkungen und Erläuterungen

terungen der Archäologen werden vom Ausgrabungsort direkt in eine virtuelle Forschungsumgebung übertragen. Das netzbasierte Arbeiten verändert die wissenschaftliche Kommunikation, Zusammenarbeit und das Publizieren von Grund auf, und zwar in allen Disziplinen.

eScience und eResearch erfordern von der „Digital Library“ komplexe Problemlösungen

An die neue Arbeitsumgebung ist auch die Erwartung geknüpft, dass sie die Forschungs- und Erkenntnisprozesse beschleunigt. Die dazu notwendige Neuorganisation des Managements wissenschaftlicher Daten, Informationen und Publikationen ist eine echte Herausforderung: Wenn Wissenschaftler weltweit digital publizieren, über Netze kommunizieren, kooperieren und von verschiedenen Orten aus gemeinsam ein Forschungsprojekt bearbeiten, brauchen sie dafür eine zuverlässige elektronische Arbeitsumgebung. Diese erfordert komplexe Technologien, um die verlässliche Organisation der Datenflüsse, die Dokumentation des erarbeiteten Wissens in Form von Primär- und Sekundärdaten und den Schutz des geistigen Eigentums gewährleisten zu können. Ein sicherer und zuverlässiger Daten- und Informationsaustausch mit kontrolliertem Zugang, die Nachvollziehbarkeit von Forschungsergebnissen, die weltweite Verknüpfung von Daten- und Informationsquellen, strukturierte Dokumentation und eine langfristige Verfügbarkeit der Daten, Werkzeuge und Ergebnisse sind für die digitale Wissenschaft und Forschung der Zukunft unerlässlich.

FIZ Karlsruhe und die MPDL haben mit dem gemeinsamen Projekt eSciDoc wichtige Voraussetzungen für eine innovative Arbeitsumgebung für die digital gestützte Forschung geschaffen. eSciDoc stellt modular aufgebaute Infrastrukturdienste zur Integration, Organisation und Verwaltung von verteilten wissenschaftlichen Ressourcen (eSciDoc-Services) und darauf aufsetzende Forschungsanwendungen für wissenschaftliches Arbeiten (eSciDoc-Applications) bereit.

eSciDoc – der Systembaukasten für netzbasiertes wissenschaftliches Arbeiten

Die eSciDoc-Infrastruktur besteht aus einer Reihe von unabhängigen, bei Bedarf untereinander kombinierbaren Softwaremodulen und bietet Dienste zur Speicherung von Objekten, für Suche und Indexierung, Statistik und Berichtswesen, dauerhaft sichere Identifikation (Persistent Identification), Arbeitsabläufe, Validierung und Transformation an. Die Projektpartner haben eine anwendungsunabhängige, auf Standards basierende Softwarearchitektur und verteilte Authentifizierung und Rechteverwaltung (Shibboleth konform) entwickelt, in deren Zentrum als zentrales Depot ein Fedora-basiertes Repository steht. eSciDoc unterstützt semantische Technologien und erlaubt die Integration eigener und/oder Drittanbieter-Softwarewerkzeuge über Web-Services (siehe auch <http://www.esdoc-project.de/JSPWiki/Wiki.jsp?page=Download>).

Das 2004 von FIZ Karlsruhe und Max-Planck-Gesellschaft ins Leben gerufene und vom Bundesministerium für Bildung und

Forschung bis Mitte 2009 geförderte Projekt wurde 2008 für die wissenschaftliche Community geöffnet, um eine nachhaltige Weiternutzung der Projektergebnisse zu erzielen. Heute, gut eineinhalb Jahre später, nutzt eine internationale eSciDoc-Community die Projektergebnisse für umfangreiche Entwicklungsarbeiten zum Aufbau von Strukturen für international vernetzte Wissenschaft und Forschung. Dabei werden die bestehenden eSciDoc-Projektergebnisse eingesetzt, weiterentwickelt und durch eigene Anwendungen ergänzt. Auch FIZ Karlsruhe, die MPDL und mit ihr zahlreiche Max-Planck-Institute erforschen und entwickeln weiterhin eigenfinanziert Anwendungen für verschiedene Disziplinen.

Bibliotheken auf dem Weg zu eResearch-Bibliotheken

Bibliotheken müssen sich an der neuen Forschungs- und Wissenschaftsinfrastruktur beteiligen, wenn sie weiterhin der Ort sein wollen, an dem Informationen gesammelt und erschlossen werden und wenn sie auch weiterhin den Zugang zu jedwedem Medium vermitteln sowie die Qualitätssicherung bei der Quellenauswahl leisten wollen. Deshalb gehörte die Verbundzentrale (VZG) des Gemeinsamen Bibliotheksverbunds (GBV) neben den Bibliotheken in der Max-Planck-Gesellschaft mit zu den ersten Projektpartnern von eSciDoc. Die VZG hat es sich zur Aufgabe gemacht, die Versorgung der deutschen Hochschulen, Forschungseinrichtungen und Bibliotheken mit elektronischen Fachinformationen zu verbessern. Dabei will sie speziell für Bibliotheken eine umfassende Lösung zur Verwaltung

Hands-On-Training: Auf den eSciDoc Days können interessierte Wissenschaftlerinnen und Wissenschaftler in Workshops den Umgang mit den eSciDoc Softwaremodulen lernen.



Umlagert: Ulla Tschida von der MPDL zeigt jungen Wissenschaftlern von der MPG entwickelte eResearch-Anwendungen, die auf eSciDoc Infrastrukturmодulen aufsetzen.



von digitalem Content in Bibliotheksumgebungen bereitstellen. Um in eine gemeinsame, homogene Bibliothekssystem-Infrastruktur bereits vorhandene Digitalisate aus Katalog-, Publikations- oder Museumsdaten zu integrieren, nutzt die VZG die eSciDoc-Infrastruktur mit dem Repository-System Fedora. Die VZG hat sich wegen der flexiblen Systemarchitektur mit umfassender Funktionalität der Vielzahl fertiger Datenmodelle und Schnittstellen für die eSciDoc-Infrastruktur entschieden.

eSciDoc im Einsatz bei Bibliotheken und Universitäten

eSciDoc basierte Bibliotheks- und Forschungsdatenmanagement-Systeme sind international im Entstehen und innerhalb der Max-Planck-Gesellschaft sogar schon im produktiven Einsatz. Die größte spanische Universität, Universidad Nacional de Educación (UNED), hat sich beispielsweise zur Information über ihren Bibliotheksbestand und zur Verwaltung verfügbarer Publikationen für die eSciDoc-Anwendung PubMan entschieden. Diese Lösung unterstützt neben strukturierter Erfassung, Versionsverwaltung, Verbreitung und Retrieval von wissenschaftlichen Publikationen auch deren Rechtemanagement und Archivierung. (<http://www.escidoc-project.de/JSPWiki/en/PublicationManagement>). Die spanischen Wissenschaftlerinnen und Wissenschaftler haben für ihre eSciDoc-Anwendung eine spanische Oberfläche entwickelt. In Deutschland haben Goportis, der Leibniz-Bibliotheksverbund Forschungsinformation, und FIZ Karlsruhe als Leibniz-Institut für Informationsinfrastruktur ein Koopera-

tionsabkommen zu Entwicklung, Angebot und Betrieb von neuen eScience-Dienstleistungen geschlossen. Goportis ist eine Kooperation der drei deutschen zentralen Fachbibliotheken TIB (Technische Informationsbibliothek, Hannover), ZB MED (Deutsche Zentralbibliothek für Medizin, Köln/Bonn) und ZBW (Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft, Kiel/Hamburg). Die ersten gemeinsamen Aktivitäten liegen auf dem Gebiet des Hostings von digitalen Verlagspublikationen. E-Journals, e-Books und Datenbanken werden in einer nachhaltigen Infrastruktur gespeichert, verwaltet und der Wissenschaft zugänglich gemacht. Weiterhin geplant ist die Zusammenarbeit in den Bereichen Langzeitarchivierung und Volltextversorgung.

An der Humboldt-Universität zu Berlin wollen Forscherinnen und Forscher disziplinübergreifend Bilddaten mit eSciDoc Infrastrukturmodulen speichern und nachhaltig bereitstellen. Die Entscheidung, eSciDoc zugrunde zu legen, begründen die Verantwortlichen vor allem mit den Funktionen, die eine sehr gute allgemeine Handhabung von Metadaten über Disziplinengrenzen hinweg erlauben sowie der einfachen Verwaltung von selektiven Zugriffsrechten. Ein besonderer Pluspunkt ist für sie auch die flexible eSciDoc-Infrastruktur, die es erlaubt, bereits vorhandene Werkzeuge zur Prozessierung von (Bild)daten weiterzuverwenden.

„Samurai“ fördert die Bekanntheit japanischer Wissenschaftler

Das japanische National Institute for Materials Science (NIMS) arbeitet konsequent

daran, auf Basis von eSciDoc-Anwendungen alle Informationsquellen der materialwissenschaftlichen Forschung – Publikationen, Bilder, Grafiken, Primärdaten sowie persönliche Daten und Informationen der Wissenschaftlerinnen und Wissenschaftler am Institut – in der „NIMS eSciDoc Digital Library“ zusammenzuführen. Die japanische eScience-Library soll nicht nur eine neuartige, forschungsfördernde Arbeitsumgebung für die eigene Forschung werden, sondern auch die Forschungsleistungen des Instituts und die dahinterstehenden Forscher weltweit bekannt machen – einer der wichtigsten Gründe, warum Universitäten, Forschungsinstitute und ihre Bibliotheken



immer stärker an einer funktionierenden eScience-Infrastruktur interessiert sind. NIMS entwickelt für die weltweite Präsentation seiner Forscher eine in die eSciDoc-Library integrierte Expertendatenbank (NIMS Researcher Database) namens Samurai, in der die wissenschaftlich relevanten Daten und Informationen zu den Forscherinnen und Forschern am Institut gespeichert werden. Über Personal Identifier (PID) ist jeder

Großes internationales Interesse: Matthias Razum, eSciDoc Projektleiter bei FIZ Karlsruhe, erläuterte dem Auditorium die mehrschichtige Architektur der eSciDoc Software



VRE-Testbett: An der Ausgrabungsstätte Silchester werden mobile Technologien erprobt, mit denen Archäologen ihre Funde vor Ort katalogisieren können. Das Projekt wurde auf den eSciDoc Days 2009 in Karlsruhe präsentiert.



Eine Auswahl von eScience-Projekten in Europa

In Deutschland und Europa laufen neben den im Bericht erwähnten Aktivitäten viele weitere spannende eScience- und eSciDoc-Nachnutzungsprojekte. FIZ Karlsruhe und/oder die MPDL sind an fast allen Entwicklungsvorhaben als Projekt- oder Kooperationspartner beteiligt. Zahlreiche Universitäten mit verschiedenen Lehrstühlen und wissenschaftliche Serviceinstitutionen sowie nicht weniger als 14 verschiedene Max-Planck-Institute (MPIs) engagieren sich für den Aufbau einer durchgängigen eScience-/eResearch-Infrastruktur. So arbeitet beispielsweise das **MPI für evolutionäre Anthropologie**, Leipzig, an einer **Open Access Digital Library zur Beschreibung des Sprachgutes** (Language Description Heritage). Das **MPI für chemische Ökologie**, Jena, setzt **eSciDoc PubMan zum Management seiner Publikationen** ein und entwickelt eigene, fachspezifische Klassifikations- und Dokumentationsformulare für die Verlinkung der Publikationslisteneinträge mit Autoren- und weiteren Informationen. Das **MPI für Wissenschaftsgeschichte**, Berlin, entwickelt **natürliche Sprachtechnologien und XML-Werkzeuge für eSciDoc-Anwendungen**.

Weitere wichtige Projekte mit verschiedenen Beteiligten

SCAPE: Das von der EU geförderte Projekt „Scalable systems and services for preserving digital content“ (SCAPE) entwickelt eine geeignete Infrastruktur zur langfristigen Bewahrung (preservation) von digitalen Inhalten. Beteiligt sind das Austrian Institute of Technology (AIT), die Nationalbibliotheken der Niederlande, Dänemarks, Großbritanniens und Österreichs, European Archive Foundation, Open Planets Foundation, Science and Technology Facilities Council, die Universitäten Wien (TU), Manchester, Paris, und Berlin (TU), Exlibris, Microsoft Research, Keeps SA und FIZ Karlsruhe.

BW-eSci(T): Das Projekt zum Aufbau einer virtuellen Forschungsumgebung (VRE) an der Universität Tübingen wird

vom baden-württembergischen Ministerium für Wissenschaft, Forschung und Kunst gefördert. Die VRE soll am Beispiel der Disziplin Computerlinguistik eine prototypische Lösung darstellen, die auch auf andere Disziplinen übertragbar ist. Am Projekt beteiligt sind neben FIZ Karlsruhe die Universität Tübingen mit der Disziplin Computerlinguistik und dem Zentrum für Information, Kommunikation und Medien (IKM) mit Universitätsbibliothek und dem Zentrum für Datenverarbeitung.

Mit der Entwicklung von VREs beschäftigen sich auch das Silchester VRE der Universität Reading, GB das eine virtuelle Forschungsumgebung für Archäologie aufbaut, sowie DARIAH (Digital Research Infrastructure for the Arts and Humanities) des King's College, GB, das auch von der EU gefördert wird.

BW-eLabs: Im Projekt BW-eLabs wird eine digitale Infrastruktur für virtuelle und fernsteuerbare Labore auf der Basis der eSciDoc-Software entwickelt. Forscherinnen und Forscher sollen damit einen besseren Zugriff auf kostenintensive Laborausstattungen für nanotechnologische Experimente erhalten. Projektpartner sind die Universität Stuttgart, die Universität Freiburg, die Hochschule der Medien in Stuttgart und FIZ Karlsruhe. Das Projekt wird ebenfalls vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert.

e-Kinematix: Die Entstehung einer integrierten Informationsinfrastruktur für den Maschinenbau ist Gegenstand des DFG geförderten Projekts. eSciDoc dient als technische Infrastruktur. Projektpartner: RWTH Aachen mit dem Institut für Getriebetechnik und Maschinendynamik, TU Ilmenau mit dem Institut für Maschinen- und Gerätekonstruktion und FIZ Karlsruhe.

Astronomer's Workbench: Das im Januar 2011 gestartete Projekt beschäftigt sich mit dem Aufbau einer Digital Library für die Astronomieforschung. Unter Federführung des MPI für Astrophysik arbeiten das MPI für extraterrestri-

sche Physik, das MPI für Astrophysik und das MPI für Gravitationsphysik zusammen, um Publikationen, Forschungsdaten und Werkzeuge in einer Umgebung zu verwalten.

DILPS: Das Projekt Distributed Image Library Processing System (DILPS) wird von der DFG gefördert. Es ist ein netzbasiertes Bilddatenbanksystem, das eine virtuelle Umgebung für die Archivierung, Verwaltung und Recherche von Bild- und Metadaten bietet. Das bestehende System soll erweitert und dabei auf die eSciDoc-Infrastruktur umgestellt werden. Projektpartner sind das Karlsruher Institut für Technologie (KIT) sowie die Staatliche Hochschule für Gestaltung der Universität Karlsruhe. FIZ Karlsruhe ist unterstützend tätig.

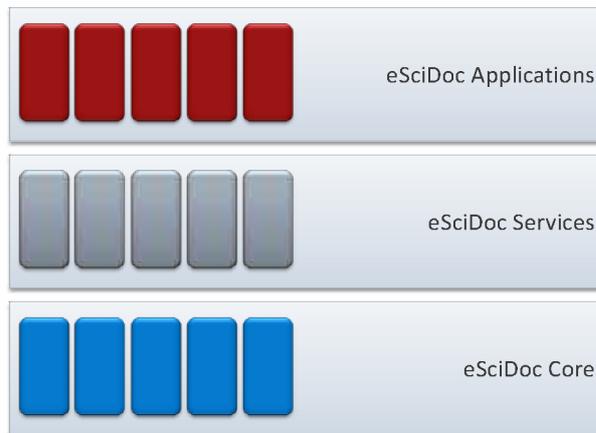
CLARIN: Mit dem EU-Projekt Common Language Resources and Technology Infrastructure (CLARIN) soll auf Basis von eSciDoc europaweit eine digitale Infrastruktur für die Geisteswissenschaften entwickelt werden. In Dänemark wurde dazu von der Royal Library of Denmark, Kopenhagen, das Projekt Clarin.dk gestartet.

Lifecycle Logger: Das MPI für Biochemie der Universität Göttingen untersucht den Lebenszyklus digitaler Objekte und entwickelt Software für die Verfolgung eines Objektes von den ersten Forschungsdaten (Rohdaten) bis zur Publikation.

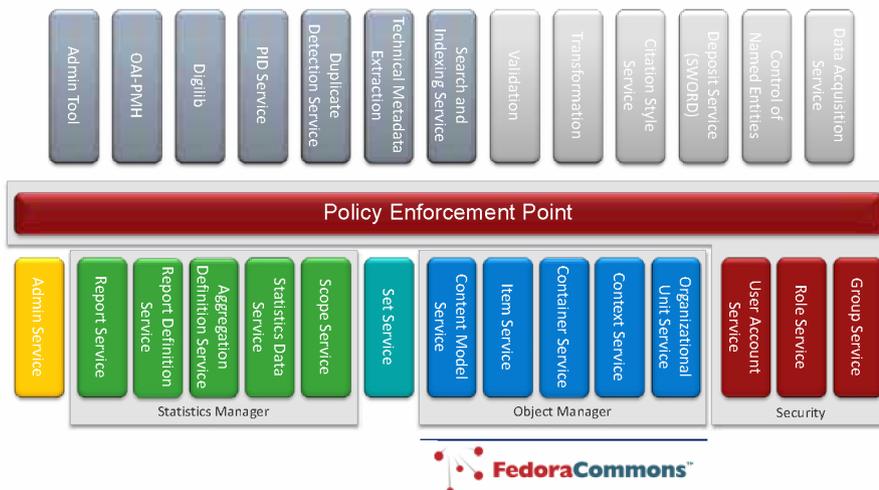
Jus CMS: Die eSciDoc Solution PubMan wird für Jus CSM genutzt, um ein durchgängiges Content-Management-System für die Rechtswissenschaften aufzubauen. Beteiligt sind daran mehrere MPI-Institute für rechtswissenschaftliche Forschung (MPI für ausländisches und internationales Privatrecht, MPI für Steuerrecht und Öffentliche Finanzen, MPI für ausländisches und internationales Strafrecht, MPI für ausländisches öffentliches Recht und Völkerrecht, Max-Planck-Institut für ausländisches und internationales Sozialrecht).



Schematic View of eSciDoc



Services of the eSciDoc Infrastructure



4

eSciDoc Infrastructure - An Update

eSciDoc Days 2010, København

19.11.2010

Struktur und Baukasten, zwei Folien aus dem Beitrag von Matthias Razum

Wissenschaftler mit seinen Publikationen im Archiv und von dort weiter mit den zugrundeliegenden Forschungsmaterialien und -datensätzen verbunden. Außerdem wird die Bilddatenbank „NIMS Research Image Data Archive“ auf die neue eSciDoc-Anwendung Imeji umgestellt. Imeji ist bereits eine Weiterentwicklung der eSciDoc-Anwendungen FACES und Diamonds. FACES, eine Fotosammlung des Max-Planck-Instituts für Bildungsforschung (MPIB), besteht aus mehr als 2000 Bildern für das Studium der Gesichtsmimik und ergänzt sie um zahlreiche beschreibende Attribute und Metadaten. Diamonds, die Fotokollektion von NIMS, beinhaltet über 150 Aufnahmen von Kristallstrukturen, ergänzt um beschreibende Attribute und Metadaten.

Die erweiterte Imeji-Software ermöglicht nun die Verknüpfung mit dem Publikationsarchiv und erlaubt die Vergabe von weiter differenzierten Zugriffsrechten. So können Nutzer mit dem Datenbestand oder aber auch nur mit bestimmten Teilmengen der Sammlungen arbeiten.

Zahlreiche internationale Initiativen erarbeiten Infrastrukturkonzepte für eScience

Mittlerweile beteiligen sich nicht nur viele Universitätsbibliotheken, Hochschulen und Forschungsinstitute an der Nachnutzung und Weiterentwicklung von eSciDoc, auch viele Länder haben auf nationaler Ebene erkannt, dass die immer stärker datengetrie-

bene Forschung anspruchsvolle, gut durchdachte Informations- und Forschungsdatenmanagementsysteme braucht. In Deutschland erarbeitet die Kommission für Informationsinfrastruktur (KII), deren Vorsitzende die Geschäftsführerin von FIZ Karlsruhe, Sabine Brünger-Weilandt, ist derzeit ein solches Konzept, das im späten Frühjahr 2011 vorliegen soll. In Dänemark will man einen Fahrplan für den Aufbau einer Forschungsinfrastruktur, die eScience beinhaltet, bis Sommer 2011 fertig gestellt haben. Japan startet Anfang 2011 mit seinem neuen 5-Jahresplan, in dem die Forschungsschwerpunkte einschließlich eScience festgelegt sind. In den Projekten hat sich gezeigt, dass eSciDoc ein guter Ansatz ist, um eScience und eResearch zu fördern und die digitale Wissenschaftsinformation in kontrollierte Bahnen zu lenken.

Doch während Bibliothekare und Information Professionals gemeinsam mit den Fachwissenschaftlern eScience Libraries aufbauen, wirft das Semantische Web mit seinen linked data schon wieder neue Fragen auf: Wie geht man mit Autoren- und Urheber-schaft, Bearbeitungs- und Publikations-befugnissen um, wenn die Informationen rund um die Welt vernetzt von menschlichen Nutzern, aber auch von Maschinen immer stärker fragmentiert und/oder zu neuen Verbindungen zusammengesetzt genutzt werden? Hier wird eine Instanz benötigt, die die Sicherung des zukünftigen Zugangs zur Information übernimmt und dabei die korrekte Nutzung und Interpretation von Information unterstützt, insbesondere über Disziplinengrenzen hinweg, so die Worte der stellvertretenden Direktorin der dänischen Royal Library, Birte Christensen-Dalsgaard, auf den eSciDoc Days 2010 im November in Kopenhagen, Dänemark.

AUTORIN

HELGA BERGMANN-OSTERMANN

Medizinjournalistin, Dipl.-Übersetzerin
 Am Stockacker 9
 94508 Schöllnach
 Tel.: 09903-941730
 Fax: 09903-9510100
 Mobil: 0175-36 136 38
 h.bergmann-ostermann@t-online.de