

Virtuelle Forschungsumgebungen und Forschungsdaten für Lehre und Forschung: Informationsinfrastrukturen für die (Natur-)Wissenschaften

Matthias Schulze, Stuttgart

Bei den Informationsinfrastrukturen für die Wissenschaft stehen in jüngster Zeit „Virtuelle Forschungsumgebungen“ und „Forschungsdaten“ ganz oben auf der Tagesordnung. Zweifellos stellt der Umgang mit diesen beiden Bereichen die Wissenschaft, aber auch die mit Infrastruktur befassten Einrichtungen wie Daten-/Rechenzentren, Wissenschaftsverwaltungen und Bibliotheken, vor ganz neue Herausforderungen, etwa bei Qualitätssicherung, (Metadaten-)Standards, Interoperabilität, Qualifizierung, Open Access und Langzeitarchivierung: Die Universitätsbibliothek Stuttgart ist seit Mitte 2009 an zwei komplementären Projekten beteiligt, in denen diese neuen Herausforderungen exemplarisch für Lehre und Forschung im Bereich der Naturwissenschaften angegangen werden. Zum einen handelt es sich um das von der KEG geförderte eContentplus-Projekt „LiLa – Library of Labs“, mit dem Ziel der Vernetzung von virtuellen Laboren für die akademische Ausbildung. Zum anderen geht es in dem vom Land Baden-Württemberg geförderten Projekt „BW-eLabs – Wissensmanagement in virtuellen und remote Laboren“ um die Erweiterung des Zugriffs auf heterogene experimentelle Ressourcen zur nachhaltigen Erschließung sowie Nutzung von Forschungsdaten und Experimenten für Forschungs- und Ausbildungszwecke. Der Beitrag stellt die beiden Projekte zur Diskussion und präsentiert erste Ergebnisse. Darüber hinaus werden die identifizierten Herausforderungen in einen größeren Zusammenhang eingeordnet.

Einleitung

„Forschungsdaten“ und „Virtuelle Forschungsumgebungen“ sind zwei der Schlagworte, die zur Zeit immer wieder auftauchen, wenn es um aktuelle Entwicklungen und Zukunftsaufgaben im Umfeld dessen geht, was als „Fachinformationsinfrastruktur in Deutschland“ bezeichnet wird.¹ Auf Kongressen, wissenschaftlichen Fachtagungen, Bibliothekartagen und anderen Veranstaltungen stößt man seit kurzem verstärkt und verlässlich auf diese Themen.² Und auch die Einrichtung von Arbeitsgruppen und das Publizieren von bzw. Arbeiten an Stellungnahmen zu diesen Themenfeldern kann als Indiz für ihre aktuelle und zukünftige Bedeutung aufgefasst werden.³ Die unterschiedlichsten Akteure wie Wissenschafts-

1 Dieser Beitrag wird sich allein auf die Entwicklungen in Deutschland beschränken. Auf internationale Vorbilder und Tendenzen wird nur im Einzelfall eingegangen. Alle URL dieses Beitrags wurden am 29. Juli 2010 überprüft.

2 Aktuell fanden im Frühjahr 2010 zwei Veranstaltungen statt: Zum einen eine Blockveranstaltung mit sechs Beiträgen zum Thema Forschungsdaten („Bibliotheken als Akteure im Forschungsdatenmanagement“) http://www.conwerk.net/contxt_bidkon10/programme/view_symp_detail_short_abstract.asp?referer=17_overview.asp&symposiumID=24 anlässlich des BID-Kongresses in Leipzig sowie eine Tagung in Potsdam zum Thema „eScience und Forschungsdatenmanagement“ http://iw.fh-potsdam.de/iw-tagungen_iscience.html.

3 Auch hier wieder nur einige Beispiele: Die Schwerpunktinitiative „Digitale Information“ der Allianz der Wissenschaftsorganisationen hat im Handlungsfeld Forschungsdaten eine „Policy“ zum Umgang mit Forschungsdaten entwickelt und unter dem Titel „Grundsätze zum Umgang mit Forschungsdaten“ (<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsätze/>) im Juli 2010 verabschiedet und beschlossen. Darüber hinaus hat die von der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) initiierte Kommission „Zukunft der Informationsinfrastruktur“ (KII), die mit der Aufgabe der Erstellung einer Gesamtkonzepts für die Informationsinfrastruktur in Deutschland betraut ist, insgesamt acht Arbeitsgruppen ins Leben gerufen, davon befasst sich eine mit Forschungsdaten und eine weitere mit Virtuellen Forschungsumgebungen. Im Herbst 2009 wurde zudem eine DINI AG „Virtuelle Forschungsumgebungen“ gegründet (vgl. <http://www.dini.de/ag/vforum/>).

organisationen, Forschungsförderer, Infrastruktureinrichtungen und Initiativen tummeln sich auf diesem Feld.⁴ Insofern ist es zweifellos an der Zeit, Erfahrungen in diesem Bereich zu sammeln, um die verschiedenen Stellungnahmen und ersten Positionspapiere entweder zu bestätigen oder korrigieren zu können. Dies geschieht üblicherweise durch Projekte in unterschiedlicher Ausprägung und in verschiedenen Wissenschaftsdisziplinen.⁵

Bei Forschungsdaten und Virtuellen Forschungsumgebungen gibt es durchaus bereits erste Erkenntnisse durch bibliothekarische Vorläufer-Projekte und -Initiativen: Beispielfhaft seien hier Aktivitäten der TIB Hannover und der SUB Göttingen genannt, die sich schon sehr früh und intensiv mit diesen Themen beschäftigt und in Pilotvorhaben erste Erfahrungen gesammelt haben.⁶ Inzwischen erscheint die Vielzahl der Projekte und Vorhaben in diesem Bereich, an denen Bibliotheken beteiligt sind bzw. eine Beteiligung unmittelbar bevorsteht (gerade hinsichtlich der aktuellen DFG-Ausschreibungen), schon kaum mehr überschaubar.

Die Universitätsbibliothek (UB) Stuttgart ist seit Mitte 2009 an zwei komplementären Projekten beteiligt, in denen sie sich ebenfalls den Herausforderungen von Virtuellen Forschungsumgebungen und Forschungsdaten in exemplarischer Form für die Natur- bzw. Ingenieurwissenschaften stellt. Spezielle Aufgaben für die Universitätsbibliothek sind neben der Bereitstellung und der Integration von digitalen Inhalten die Einbindung von OPUS als Hochschulschriftenserver bzw. Repository-Software wie auch das Metadatenmanagement. Im Folgenden werden diese Projekte vorgestellt, erste Ergebnisse skizziert und die Einzelbestrebungen in einen größeren Zusammenhang eingeordnet.

1 Der Hintergrund: Die Zukunft der Wissenschaftskommunikation

Der Biologe und Wissenschaftsblogger Cameron Neylon, tätig als Senior Scientist am britischen „Science and Technology Facilities Council“,⁷ beschäftigt sich in seinem privaten Blog „Science in the Open“⁸ mit Fragen, die um technische, aber auch soziale Aspekte des Themenfeldes „Open Research/Science“ kreisen. In einem höchst instruktiven Beitrag vom 10. April 2010 befasst sich Neylon mit der Frage, wie es um die Wissenschaftskommunikation der Zukunft bestellt ist: „The future of research communication is aggregation.“⁹ Verkürzt gesagt tendiert Neylon dazu, einen wissenschaftlichen Aufsatz als Aggregation von Objekten zu sehen.¹⁰ Es geht ihm darum, zu betonen, dass diese Objekte ihre eigene

4 Einen recht guten und aktuellen Überblick bietet Heinz Pampel im ersten Teil seiner Präsentation (Folien 1 bis 13): „Umgang mit digitalen Forschungsdaten. Initiativen in Deutschland“. <http://www.slideshare.net/heinzpampel/2010-0622-huagepfinalverffentlich>.

5 Beispielfhaft seien hier die jüngsten Ausschreibungen der DFG aus den beiden letzten Jahren genannt: „Virtuelle Forschungsumgebungen. Infrastruktur und Demonstrationsprojekte“ (2009) sowie im Frühjahr 2010: „Informationsinfrastrukturen für Forschungsdaten“ (http://www.dfg.de/foerderung/info_wissenschaft/info_wissenschaft_10_02/index.html).

6 Ab 2004 bereits wurde ein Projekt „Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten“ (später dann ab 2006 fortgesetzt als „CODATA“) an der Technischen Informationsbibliothek (TIB) Hannover gestartet, um den Zugang zu wissenschaftlichen Forschungsdaten zu verbessern. Die TIB ist zudem seit 2005 zentrale DOI-Registrierungsagentur für Forschungsdaten. Inzwischen wurde mit „DataCite“ (<http://www.datacite.org/>) zu diesem Zwecke ein internationales Konsortium gegründet. Vgl. hierzu vor allem: <http://www.tib-hannover.de/de/die-tib/projekte/codata/> und <http://www.tib-hannover.de/fileadmin/datacite/index.html> sowie <http://www.tib-hannover.de/de/spezialsammlungen/forschungsdaten/>.

Die SUB Göttingen beschäftigt sich mit Forschungsdaten (unter anderem zum Aspekt der Langzeitarchivierung) und Virtuellen Forschungsumgebungen vor allem in verschiedenen Projekten im Rahmen der GRID-Initiative („TextGrid“, „WissGrid“). Vgl. hierzu: <http://www.wissgrid.de/>, <http://www.textgrid.de/>.

7 Vgl. <http://www.scitech.ac.uk/>.

8 <http://cameronneylon.net/category/blog/>.

9 <http://cameronneylon.net/blog/the-future-of-research-communication-is-aggregation/>.

10 Ebd.: “For me, the ‘paper’ of the future has to encompass much more than just the narrative descriptions of processed results we have today. It needs to support a much more diverse range of publication types, data, software, processes, protocols, and ideas, as well provide a rich and interactive means of diving into the detail where the user is interested and skimming over the surface where they are not. It needs to provide re-visualisation and streaming under the users control and crucially it needs to provide the ability to repackage the content for new purposes; education, public engagement, even main stream media reporting. I’ve got a lot of mileage recently out of thinking about how to organise data and records by ignoring the actual formats and thinking more about what the objects I’m dealing with are, what they represent, and what I want to do with them. So what do we get if we apply this thinking to the scholarly published article?”

„Identität“ als Daten bzw. Repräsentation von Daten haben und nicht – wie noch in der analogen Zeit beispielsweise – lediglich als Abbildungen in einem Beitrag (Aufsatz) verstanden und gesehen werden: „We want to use it to absorb the data it represents, and to do this we might want to zoom, pan, re-colour, or re-draw the data. But we want to know if we do this that we are using the same underlying data, so the data needs a home, an address somewhere on the web, perhaps with the journal, or perhaps somewhere else entirely, that we can refer to with confidence.“¹¹ Er macht seine Position auch anhand eines Schaubildes deutlich:

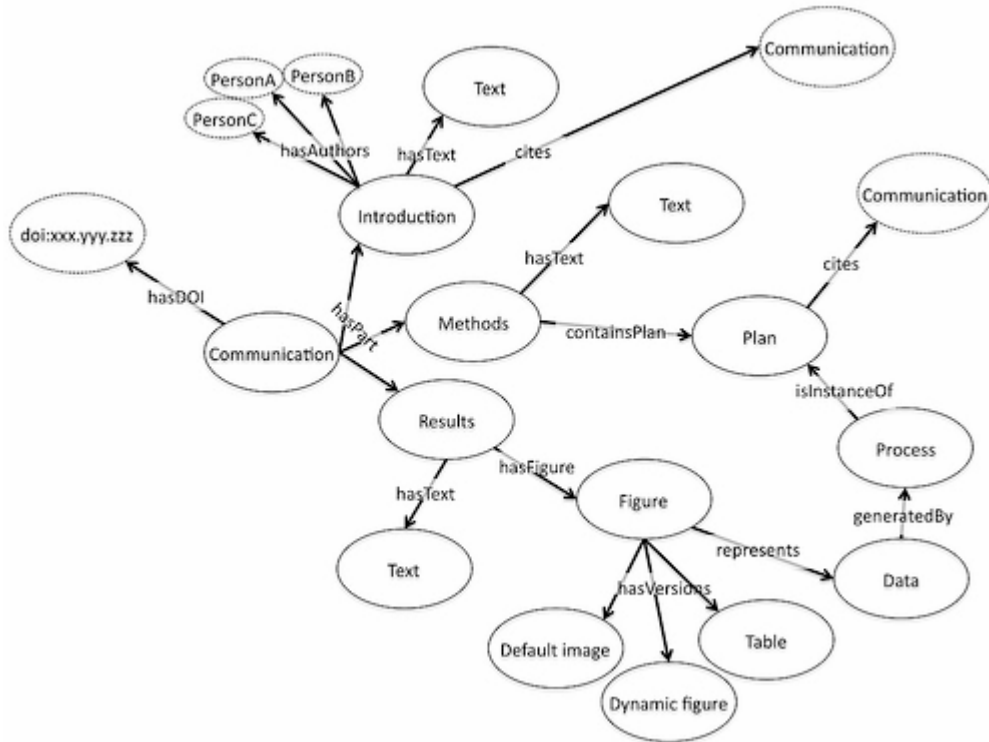


Abbildung 1: „The future of research communication is aggregation“ nach Cameron Neylon.¹²

Es geht Neylon also im Grunde um die Identität von Daten, ihre Referenzierbarkeit sowie Wiederverwendbarkeit und Nachnutzung in anderen Kontexten. Von diesem Verständnis her entwickelt er seine Sicht auf die Zukunft der Wissenschaftskommunikation und allgemein gesprochen auch auf das wissenschaftliche Publizieren an sich. Über den Umweg seines Verständnisses von Objekten und der den Wissenschaftlern naheliegenden Funktionalität des Zitierens kommt er zu dem Schluss: „If we take this view of objects and aggregates that cite each other, and we provide details of what the citations *mean* (this was used in that, this process created that output, this paper is cited as an input to that one) then we are building the semantic web as a *byproduct of what we want to do anyway*. Instead of scaring people with angle brackets we are using a paradigm that researchers understand and respect, citation, to build up meaningful links between packages of knowledge.“¹³ Neylons Überlegungen haben also implizit mit dem Thema „Semantic Web“ zu tun. Für ihn entsteht es demzufolge als logische Konsequenz – quasi nebenbei – aus dem heraus, was in Zukunft die Wissenschaftskommunikation und das wissenschaftliche Publizieren ausmacht.

¹¹ Ebd.

¹² Ebd. Das von Neylon erstellte Schaubild findet sich hier: <http://www.flickr.com/photos/24801682@N08/4506964677/sizes/l/http://cameronneylon.net/blog/the-future-of-research-communication-is-aggregation/>.

¹³ <http://cameronneylon.net/blog/the-future-of-research-communication-is-aggregation/> (Hervorhebungen so im Original).

Unter der Prämisse des Ansatzes von Neylon¹⁴ – ganz abgesehen davon, ob und wie sich Bibliotheken mit dem Thema Semantic Web auseinandersetzen (was sie sicherlich tun müssen) –, ist es ein logischer Schritt, wenn sich Bibliotheken zukünftig nicht mehr nur um herkömmliche Text- und Publikationsarten wie Monographien, Zeitschriften und Aufsätze – ob analog oder digital – kümmern, sondern ihren Blick auf das ausweiten, was am Ende die „Objekte“ (des Semantic Webs) ausmacht.¹⁵ Beispielhaft soll das im Folgenden an zwei Projekten demonstriert werden, deren Themen sich im weitesten Sinne im obigen Neylon-schen Schaubild an den drei Objekttypen „Process“, „Data“ und „Methods“ festmachen lassen. Es wird also um Virtuelle Forschungsumgebungen („Process“, „Methods“) und um Forschungsdaten („Data“) gehen.

2 Die konkrete bibliothekarische Perspektive der Universitätsbibliothek Stuttgart

Die Motivation der UB Stuttgart, sich mit Themen wie Forschungsdaten und Virtuellen Forschungsumgebungen zu beschäftigen, basiert zum einen natürlich auf der oben in der Einleitung bereits genannten Erkenntnis, sich als Infrastruktureinrichtung für die Wissenschaft auch den Zukunftsthemen zu stellen und möglichst zu Beginn einer Entwicklung an diesen zukunftssträchtigen Aufgaben mitzuarbeiten. Zum anderen ist es aber auch die ureigene Aufgabe der Universitätsbibliothek, den wissenschaftlichen Output und die publizierten Erkenntnisse der Wissenschaftler der Universität Stuttgart zu bündeln und (an einem Ort) nachzuweisen.

Gerade auf dem Feld des elektronischen Publizierens hat sich die UB Stuttgart recht früh engagiert.¹⁶ Ausgangspunkt für diese Aktivitäten – und eine weitere (wenn nicht die zentrale) Motivation für die Beteiligung an den beiden hier vorgestellten Projekten – ist die Open-Source-Software OPUS, die gemeinsam mit dem Rechenzentrum der Universität Stuttgart Ende der 1990er Jahre entwickelt und veröffentlicht wurde.¹⁷ OPUS ist in der Binnensicht der Universität Stuttgart das Repository der Universität Stuttgart (OPUS = Online Publikationsverbund Universität Stuttgart), zugleich steht OPUS aber auch für die Repository-Software an sich. Sie ist inzwischen in Deutschland die am weitesten verbreitete im bibliothekarischen Kontext: Über einhundert produktive Installationen werden zurzeit mit der OPUS-Software betrieben: Vorrangig von Bibliotheksverbänden, von Bibliotheken für Universitäten und Fachhochschulen, aber auch von Forschungseinrichtungen, Instituten und anderen wissenschaftlichen Institutionen.¹⁸ So steht bei der Mitarbeit der Universitätsbibliothek Stuttgart in den beiden im Folgenden beschriebenen Vorhaben nicht nur die Weiterentwicklung des eigenen universitären Repository im Fokus, sondern auch die OPUS-

14 Cameron Neylon hat seine Gedanken in einem ergänzenden Blog-Beitrag noch weiter ausgeführt und sich dabei speziell auch den technisch-editorischen Aspekten der Thematik zugewandt: „Implementing the ‚Publication as Aggregation““ (Blogbeitrag vom 19. Mai 2010: <http://cameronneylon.net/blog/implementing-the-„publication-as-aggregation“>). An dieser Stelle kann auf die weiterführenden Überlegungen leider nicht eingegangen werden.

15 Seien dies Daten, Protokolle, Versuchsanordnungen, Software, Labor-Setups oder andere „supporting information“ jeglicher Art.

16 Dies geschah unter anderem in der Form von Projektbeteiligungen und Mitgliedschaften in überregionalen Initiativen und Einrichtungen, so etwa bei DINI, der DINI-AG Elektronisches Publizieren usw. Zu den Projekten der UB Stuttgart vgl. <http://www.ub.uni-stuttgart.de/wirueberuns/projekte/>.

17 Ausführliche Informationen zu OPUS an der Universität Stuttgart finden sich hier: <http://elib.uni-stuttgart.de/opus/>. Speziell zur Geschichte, Motivation und Hintergründen: <http://elib.uni-stuttgart.de/opus/doku/about.php>, Veröffentlichungen über OPUS: <http://elib.uni-stuttgart.de/opus/doku/veroeff.php>. Die Geschäftsstelle für OPUS ist inzwischen am Bibliotheksservice-Zentrum (BSZ) in Konstanz angesiedelt: <http://samos.bsz-bw.de/>.

Zur Zeit befasst sich die OPUS-Kernentwicklergruppe damit, die OPUS-Version 4 fertigzustellen. Das Release soll im Herbst 2010 erfolgen.

18 Vgl. hierzu die „Liste der produktiven OPUS-Installationen“:

http://samos.bsz-bw.de/index.php?id=85&no_cache=1 (Am Stichtag 27. Juli 2010 sind es 104 Installationen).

Community insgesamt wird von den in den Projekten entwickelten Fortschritten und Features profitieren.

So spielen natürlich die konkreten hier genannten Motive und Perspektiven die Hauptrolle beim Engagement der UB Stuttgart. Allerdings sind ebenso auch mittel- bis langfristig von Bibliotheken die Themen, die Neylon im Umfeld der Weiterentwicklung der Wissenschaftskommunikation aufgeworfen hat, umfassender in den Blick zu nehmen: Wie entwickelt sich diese überhaupt, wie sieht das wissenschaftliche Publikationsverhalten zukünftig aus? Als Startpunkt jedoch erscheint die „Weitung des Blicks“ auf die relativ konkreten Themen (Forschungsdaten – Virtuelle Forschungsumgebungen) und Anforderungen eine sinnvolle und pragmatische Herangehensweise.

3 Wissensmanagement in virtuellen und remote Laboren: BW-eLabs und LiLa¹⁹

3.1 Experimente in virtuellen und remote Laboren

Für die natur- und ingenieurwissenschaftliche Forschung stellen Experimente einen (wenn nicht den) zentralen Teil der Forschungsmethodik dar. Ihre Durchführung obliegt jedoch häufig zahlreichen Einschränkungen, die vor allem in der finanziellen Ausstattung der Einrichtungen, in der räumlichen und in der Betreuungskapazität begründet liegen. Durch den Einsatz neuer Medien können diese Schwierigkeiten verringert werden. Zwei Konzepte hierbei sind Virtuelle Labore und Remote-Experimente.

Virtuelle Labore sind Softwareumgebungen, die ein Rahmenwerk zum Durchführen von Experimenten am Computer bilden. Virtuelle Labore können – durch geeignete Instanzenbildung – von beliebig vielen Gruppen gleichzeitig genutzt werden. Da die Experimente vollständig im „virtuellen Raum“ stattfinden, sind hier auch Versuche möglich, die im realen Leben – etwa aus Sicherheitsgründen – niemals realisierbar wären.

Remote-Experimente sind reale, ferngesteuerte Experimente, die von einem Standort außerhalb eines realen Labors kontrolliert werden. Sie bestehen aus zwei zentralen Komponenten: Einerseits dem eigentlichen realen Versuchsaufbau, und andererseits der Technologie, die den Zugriff von außen ermöglicht. Remote-Experimente erlauben, unabhängig von Zeit und Ort, reale natur- und ingenieurwissenschaftliche Versuche und Versuchsaufbauten zu nutzen.

Obwohl inzwischen zahlreiche virtuelle Labore existieren und viele große experimentelle Setups mit Remote-Zugriff ausgestattet (bzw. ausstattbar) sind und auch eingesetzt werden, stehen jedoch kaum „Portale“ als organisatorische „Klammern“ solcher experimentellen Ressourcen zur Verfügung. Gleichzeitig fehlt auch die Anbindung virtueller/remote Labore an andere relevante wissenschaftliche Infrastrukturen: Etwa zu bibliothekarischen Repositorien oder zu wissenschaftlichen Dokumentenmanagementsystemen. Dadurch bleiben die experimentellen Setups der Labor- und Forschungsumgebungen sowie die in ihnen durchgeführten Experimente und die ermittelten Ergebnisse (z. B. Forschungsdaten oder Publikationen) voneinander getrennt. Zum Beispiel kann eine Zuordnung und Verknüpfung von Resultaten zu den Aufbauten sowie die Transparenz und Reproduzierbarkeit von Ergebnissen somit kaum hergestellt werden.

¹⁹ Im Folgenden werden zwar beide Projekte vorgestellt, aber der Schwerpunkt wird deutlich auf dem Projekt „BW-eLabs“ liegen. „LiLa“ wird im Zusammenhang mit einigen in beiden Vorhaben parallel laufenden Entwicklungen thematisiert werden. Darüber hinaus werden am Ende kurz die Besonderheiten in diesem internationalen Vorhaben skizziert. In dem hier vorgegebenen Rahmen kann es zudem nicht darum gehen, alle oder viele Teilaspekte der jeweiligen Vorhaben zu präsentieren. Vielmehr soll die allgemeine Darstellung der Projekte nur soweit erfolgen, als sie zum Verständnis notwendig ist. Im Mittelpunkt werden die Arbeitspakete und Aspekte stehen, die von der UB Stuttgart bearbeitet werden und damit vorrangig im bibliothekarischen Interesse sind.

3.2 Vielerlei Daten: Metadaten, Forschungsdaten, „statische und dynamische“ Daten

Die erwähnte unzureichende Verfügbarkeit von Experimentalkapazitäten ist nicht auf einzelne Disziplinen beschränkt, sie wirkt sich jedoch speziell dort aus, wo experimentelle Ausstattung sehr kostenintensiv ist. Dies trifft in besonderer Weise auf das Gebiet der Nanotechnologie zu, das einen außerordentlich großen Einfluss auf die aktuelle technologische Entwicklung hat und dem die Rolle einer zentralen Schlüsseltechnologie des 21. Jahrhunderts zugeschrieben wird.²⁰ Aus diesem Grunde wurde für das Projekt BW-eLabs²¹ neben dem Institut für Technische Optik der Universität Stuttgart (dort beschäftigt man sich u.a. mit digitaler Holografie)²² das Freiburger Materialforschungszentrum (FMF), das sich fakultätsübergreifend und interdisziplinär mit anwendungsnaher Grundlagenforschung u.a. im Bereich der Nanotechnologie befasst,²³ als zweiter wissenschaftlicher Projektpartner, „Content Provider“ und Betreiber von remote/virtuellen Laboren gewonnen.

Ziel des Projekts BW-eLabs ist der Aufbau eines Portals, um die oben angesprochenen fachspezifischen Herausforderungen experimenteller Natur- und Ingenieurwissenschaften unter anderem am Beispiel der Nanotechnologie anzugehen: Während herkömmliche Repositorien statische und langlebige Daten wie z. B. Publikationen hosten, ist das Ziel dieses Projektes eine Erweiterung bereits bestehender Repositorien um 1) Forschungsdaten aus aktuellen Experimenten, um 2) weitere Forschungsdaten aus informellen Dokumenten und Dokumentationen (Protokolle, Notizen), die entlang der wissenschaftlichen Arbeitsabläufe im Labor entstehen und 3) um die Experimente selbst. Ziel der Gesamtarchitektur des Vorhabens ist, einerseits experimentelle Ergebnisse und die hierzu notwendigen Prozessabläufe zur Nachnutzung etwa in Forschung und Lehre verfügbar zu halten und andererseits durch die Verfügbarkeit von Laborkomponenten über das Internet Kooperationen zwischen geographisch verteilten Forschergruppen zu ermöglichen. Zielgruppen des BW-eLabs-Projektes sind vornehmlich Wissenschaftlerinnen und Wissenschaftler, aber auch der wissenschaftliche Nachwuchs – etwa graduierte Studierende verschiedener Fachgebiete.²⁴

Ein Ziel von BW-eLabs ist es zudem, Ergebnisse nicht nur den Forschern innerhalb eines Projekts (also z. B. dem lokal arbeitenden Wissenschaftler und seiner entfernt arbeitenden Kollegin) verfügbar zu machen, sondern ausgewählte Datenobjekte mit einem Persistent Identifier zu versehen, um diese anschließend referenzierbar publizieren und damit zitierbar machen zu können. In diesem Punkt ist das Vorhaben damit schon sehr nah an der

20 Bedingt durch die „Winzigkeit“ der zu untersuchenden Objekte und Effekte ist die Anforderung an die Präzision der Messapparaturen enorm, der finanzielle Aufwand für die Bereitstellung des notwendigen experimentellen Equipments (Reinräume, Elektronenmikroskope, hochpräzise Spektrometer etc.) außerordentlich hoch. Als Resultat steht hochprofessionelles Equipment derzeit in der Regel nur einer kleinen Gruppe gut ausgestatteter Institutionen zur Verfügung. Damit bietet sich das Gebiet der Nanotechnologie als Prototyp und Pilotprojekt für BW-eLabs in besonderer Weise an.

21 Eine allgemeine Projekt-Homepage gibt es leider noch nicht. Erste Informationen finden sich auf der Projekte-Seite der UB Stuttgart: <http://www.ub.uni-stuttgart.de/wirueberuns/projekte/bw-elabs/>.

Da das Vorhaben im Herbst 2009 (mit einer Laufzeit von 30 Monaten) begonnen wurde, können hier – nach etwa einem Drittel der Projektlaufzeit – nur (Teil-)Ergebnisse vorgestellt werden, die „work in progress“ darstellen. Projektpartner bei BW-eLabs sind neben der Universität Stuttgart (beteiligt mit dem Rechenzentrum, der Universitätsbibliothek und dem Institut für Technische Optik), die Hochschule der Medien Stuttgart, die Universität Freiburg (Freiburger Materialforschungszentrum FMF und Rechenzentrum) und das Fachinformationszentrum (FIZ) Karlsruhe. Finanziert wird das Vorhaben vom Land Baden-Württemberg, durch das Ministerium für Wissenschaft, Forschung und Kunst. In einem parallel geförderten Schwesterprojekt geht es unter dem Kürzel „BW-eSci(T)“ um die „Entwicklung einer prototypischen e-Science-Forschungsumgebung an der Universität Tübingen“ vorrangig um linguistische Fragestellungen. Vgl.: <http://www.zdv.uni-tuebingen.de/projekte/bw-escit.html>.

22 Vgl.: <http://www.uni-stuttgart.de/ito/>.

23 Vgl.: <http://www.fmf.uni-freiburg.de/>.

24 Teile der entwickelten Forschungsumgebung lassen sich darüber hinaus durchaus auch in der akademischen Ausbildung einsetzen – insbesondere dort, wo Studierenden ein umfassender Zugang zu professionellem experimentellen Equipment ermöglicht werden soll.

Sichtweise Cameron Neylons wie zukünftig Publikationstätigkeit und Wissenschaftskommunikation aussehen wird: Forschungsdaten als Objekte. Zentrales Element der BW-eLabs Gesamtarchitektur ist die eSciDoc-Software-Infrastruktur, die vom FIZ Karlsruhe²⁵ zur Verfügung gestellt wird.²⁶ Sie dient speziell zur Speicherung, Verwaltung und Beschreibung der anfallenden Forschungsdaten (im Projektkontext als „dynamischer Content“²⁷ bezeichnet) sowie in der automatischen Speicherung dieser Daten während der ablaufenden virtuellen bzw. Remote-Experimenten. Die systematische Erfassung, Erschließung und Verwaltung der anfallenden Forschungsdaten ermöglicht ein optimiertes Forschungsdatenmanagement.²⁸

Als Grundlage für die Erstellung eines Metadatenkonzepts für dynamische und statische Forschungsdaten erfolgte eine Anforderungsanalyse sowohl beim Freiburger Materialforschungsinstitut als auch beim Institut für Technische Optik der Universität Stuttgart. Hierbei werden unter „dynamischen Daten“ die von den Geräten (Kamera, Mikrowelle, Photolumineszenz- und Absorptionsspektrometer) während des Experiments ausgegebenen Mess- und Konfigurationsdaten verstanden. Im Falle des FMF (Freiburger Materialforschungsinstitut) sind dies überwiegend Spektren, beim ITO (Institut für Technische Optik an der Universität Stuttgart) Hologramme. „Statische Daten“ umfassen die daraus abgeleiteten Publikationen sowie damit verknüpfte publizierte Datensätze (oft auch als „supporting information“ bezeichnet). Die Anforderungsanalyse beschreibt den Status quo bezüglich der Arbeitsprozesse und der relevanten Datenobjekte.²⁹

Metadaten sind ein entscheidender Faktor für die Archivierung, Publikation und Nachnutzung von Forschungsdaten. Nur hinreichend beschriebene Daten können durch Wissenschaftler wahrgenommen und korrekt eingeordnet werden. Erst derartig beschriebene Daten können sinnvoll veröffentlicht und zitiert werden. Zusätzlich erlauben geeignete Metadaten das Browsen und Suchen in den Forschungsdaten. Deshalb ist die Definition über die Verwendung geeigneter Metadaten-Modelle, die die Daten auf dem richtigen Level beschreiben, gerade auch aus bibliothekarischer Sicht, selbstverständlich von außerordentlicher Bedeutung.

Neben dynamischen Daten arbeitet das Projekt BW-eLabs auch mit „statischen Daten“, unter denen vor allem Daten zu Publikationen verstanden werden, auch hierfür sind Metadaten vorzusehen. Publikationen enthalten bereits heute vielfach Verweise auf Datenobjekte

25 Das FIZ Karlsruhe, Leibniz-Institut für Informationsinfrastruktur, fördert mit seinen Informations- und Serviceangeboten (u.a. Datenbanken, Fachportale, Recherchedienste etc.) den Wissenstransfer in den Bereichen Wissenschaft und Technik. Vgl. <http://www.fiz-karlsruhe.de/home.html>.

26 http://www.fiz-karlsruhe.de/escidoc_project.html.

27 Unter „dynamischen Daten“ werden im Projekt BW-eLabs die im Rahmen von (virtuellen bzw. remote durchgeführten) Experimenten anfallenden Forschungsdaten verstanden, meist in Form von Messwerten.

28 Vgl. hierzu auch Razum et al. (2009).

29 Um die Komplexität der entstehenden Daten insgesamt anzudeuten, hier eine Auflistung der beim FMF in remote Experimenten entstehenden Daten: 1) Primärdaten: Absorptionsspektren (Galaxy SPC), Dunkelmessung Absorptionsspektrometer (Galaxy SPC), Leermessung Absorptionsspektrometer (Galaxy SPC), Photolumineszenzspektren (Galaxy SPC), Konfiguration der Mikrowelle (Soll-Werte, unbekanntes Format, eventuell, MS-Access), tatsächliches Verhalten der Mikrowelle (Ist-Werte, unbekanntes Format, Übersicht verfügbar als PDF); 2) Abgeleitete Daten: Messwerte (Spektren) als numerische Daten (Full - Metadata Format), Origin - Berechnungen und Visualisierungen (Formate noch unklar); 3) Strukturdaten: Projekte, Experimente, Versuchsreihen; 4) Publikationen: Referenzen auf Primärdaten („supporting information“), Referenzen auf Publikationen („Motivation“ für Experimente).

Am ITO entstehen in virtuellen und remote Experimenten folgende Daten: 1) Primärdaten: Digitale Hologramme (Bitmaps, TIFF), Beschreibende Daten, Experimentalaufbau (Freitext, eventuell mit Foto), zusätzlich als numerische Werte (z. T. mit Einheit) die Distanz zwischen CCD, Kamera und Objekt, Position des Objektes und Orientierung des Objektes, Lichtquelle (Leistung, Wellenlänge), Kameraeigenschaften (Auflösung in Pixel und Pixelabstand); 2) Abgeleitete Daten: Rekonstruktionen (Bitmaps, TIFF), Komprimierte digitale Hologramme (Bitmaps, GIF bzw. JPEG); 3) Strukturdaten: Projekte, Experimente, Versuchsreihen; 4) Publikationen: Referenzen auf Primärdaten ("supporting information").

bzw. Messwerte. Momentan oftmals noch als „supporting information“ von Verlagen an die Publikation angehängt, ist es ein Ziel von BW-eLabs, die dabei auftretenden Beschränkungen hinsichtlich Umfang und Nachnutzbarkeit zu umgehen und die Daten nur noch über Verweise (z. B. anhand von persistenten Identifikatoren) aus der Publikation heraus zu verknüpfen. Damit werden die Datenobjekte eigenständige Entitäten, die einen eigenen Publikationsverlauf nehmen.³⁰

Die UB Stuttgart hat – wie oben bereits dargestellt – als zentrale Einrichtung der Universität Stuttgart die Aufgabe der Bereitstellung und Verwaltung eines klassischen Dokumentenservers (OPUS) mit digitalen Dokumenten und deren „statischen“ Metadaten. Auf dieser Basis wird im Projekt BW-eLabs die direkte Einbindung digitaler wissenschaftlicher Dokumente, die in virtuellen Laboren entstanden sind, ermöglicht. Zudem werden so auch wichtige Forschungsdaten des zum Dokument gehörigen Forschungsprozesses dauerhaft öffentlich verfügbar und über bereits bestehende Bibliotheksportale und Suchmaschinen einem breiten (Fach-)Publikum zugänglich gemacht. Um die Vernetzung mit den virtuellen und remote Laboren der Projektpartner (FMF/ITO) sicherzustellen, wurde ein Konzept für den Datenaustausch und die dafür benötigten Schnittstellen zur Einbettung der Daten und Dokumente in den OPUS-Dokumentenserver der Universität Stuttgart definiert und entwickelt.

Zunächst ging es darum – gemeinsam vor allem mit dem Projektpartner FIZ Karlsruhe (wie oben bereits beschrieben, zuständig für die Metadaten des „dynamischen“ Contents) – zu eruieren, welche Metadaten-Standards und -Profile bereits existieren, ob und wie diese erweitert werden können, um eine möglichst breite Vernetzung und Zukunfts- und Ausbaufähigkeit zu gewährleisten.³¹ Das derzeit parallel zu BW-eLabs im EU-Projekt „Library of Labs (LiLa)“ – hierzu unten mehr – entwickelte Metadatenprofil „Library of Labs Application Profile (LiLa AP)“ eignet sich für die Datenbeschreibung ebenfalls nur eingeschränkt. Es konnten jedoch die dort gesammelten Erfahrungen im Umgang mit neuen Metadatenfeldern direkt in BW-eLabs genutzt werden, wobei die bereits in OPUS vorhandenen Metadatenfelder zur Beschreibung von Dokumenten als Ausgangspunkt für die Integration der Forschungsdaten verwendet wurden. Das so entwickelte Metadatenprofil für BW-eLabs wird nun Schritt für Schritt entsprechend den Anforderungen der Forscher erweitert, um letztendlich die Forschungsdaten über OPUS auch in den Bibliothekssystemen nachweisen zu können.

Als Ergebnis der Zusammenarbeit von UB Stuttgart und FIZ Karlsruhe wurde ein Metadatentransferschema (vgl. Abb. 2) entwickelt, das als Basis für den Austausch von Daten zwischen eSciDoc und OPUS dient und damit die Integration der Forschungsdaten in OPUS ermöglicht.

Diese Darstellung verdeutlicht insbesondere, dass für den Nachweis in OPUS und den nötigen Publikationsprozess von Forschungsdaten die Zuweisung eines persistenten Identifikators (hier eines DOI)³² erforderlich und die zentrale Komponente ist. Das Konzept sieht zudem vor, dass die Forschungsdaten für die Integration in OPUS als neue (eigenständige) Publikationsform aufgefasst werden.

30 Auch dies ist konzeptionell wieder sehr nah an dem, was Cameron Neylon formuliert hat.

31 Die gemeinsam mit dem FIZ betrachteten Metadatenkonzepte ISO 2146 (http://www.iso.org/iso/catalogue_detail.htm?csnumber=44936) und das vom Science and Technology Facilities Council (STFC) entwickelte Core Scientific Metadata Model (<http://www.scitech.ac.uk/>) erwiesen sich hierbei nur als sehr bedingt geeignet für BW-eLabs.

32 DOI: Digital Object Identifier. Vgl.: <http://www.doi.org/>. Die für BW-eLabs angestrebte Lösung wird sich nicht auf ein einziges Persistent Identifier (PI)-System beschränken. Neben DOI werden auch andere PIs möglich sein, wie etwa die in der bibliothekarischen Welt sehr verbreiteten Uniform Resource Names (URNs).

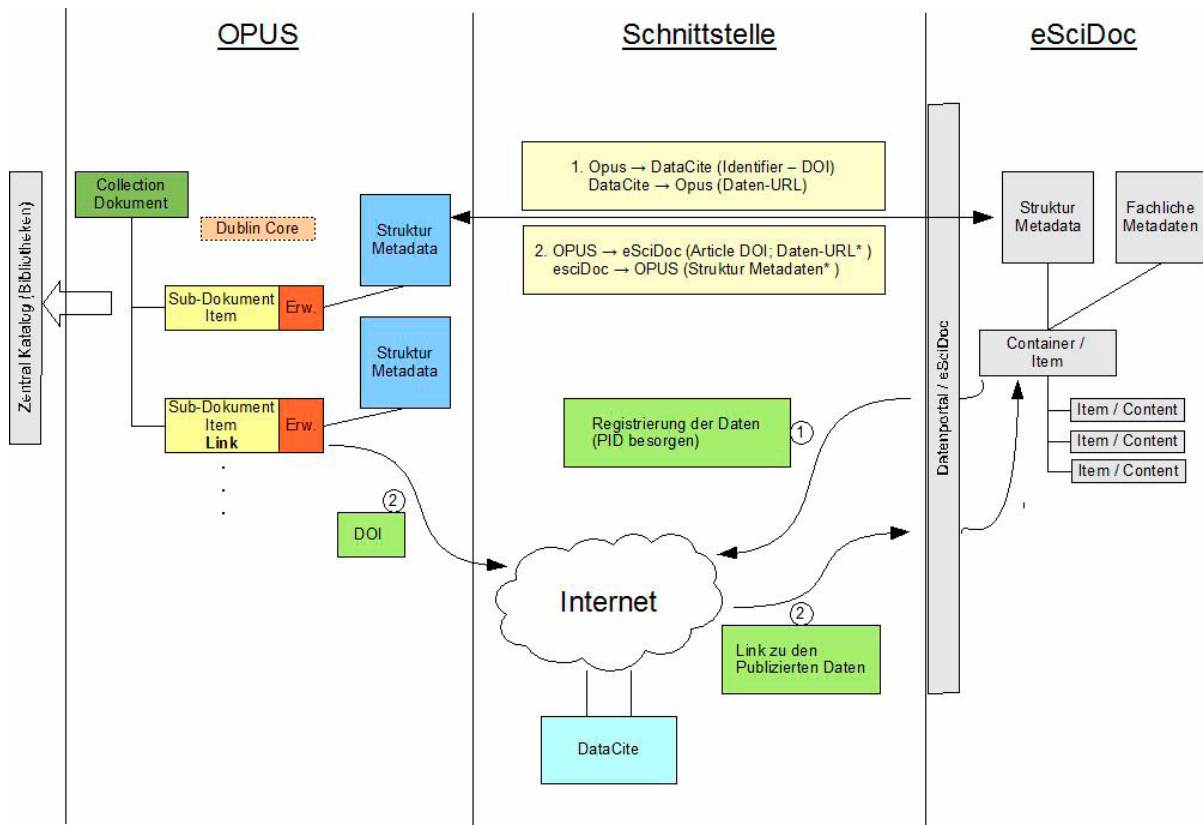


Abbildung 2: Metadatentransferschema eSciDoc – OPUS (Entwurf).

Das Metadatentransferschema und der Entwurf des Metadatenprofils bilden somit den Kern der bisherigen Projektaktivitäten der UB Stuttgart bei BW-eLabs. Im weiteren Verlauf des Projekts wird es insbesondere darum gehen, die geplante Verknüpfung der Forschungsdaten mit den Publikationen umzusetzen. Darüber hinaus steht die Einbindung der Forschungsdatensätze über die erhobenen Metadaten und die OPUS-Lösung hinaus in die Bibliothekskataloge auf dem Arbeitsplan. Dies wird in Zusammenarbeit mit dem Bibliotheksservice-Zentrum Baden-Württemberg (BSZ)³³ für den Online-Katalog des Südwestdeutschen Bibliotheksverbundes Baden-Württemberg, Saarland, Sachsen (SWB)³⁴ durchgeführt. Auf diesem Wege werden die Forschungsdaten dann auch in den lokalen Bibliothekskatalogen (wie etwa dem Katalog der UB Stuttgart), aber auch in Suchmaschinen recherchierbar sein.

3.3 Fokus Lehre: Das Projekt „LiLa – Library of Labs“

Im EU-Projekt „LiLa - Library of Labs“ ermöglichen elf Partner (acht Universitäten und drei Unternehmen) aus sieben Ländern über ein zentrales Portal den wechselseitigen Zugang zu Laboren mit ferngesteuerten und virtuellen Experimenten (Simulationen) der jeweiligen Hochschulen.³⁵ Studierende der Ingenieur- und Naturwissenschaften können so in einer virtuellen Umgebung Experimente durchführen, für welche die Kapazitäten realer Labore

33 Vgl. <http://www2.bsz-bw.de/>.

34 Vgl. <http://swb2.bsz-bw.de/DB=2.1/>.

35 <http://www.lila-project.org/> Das Projekt startete Mitte Mai 2009 für die Dauer von zwei Jahren. Beteiligte Partner sind: Universität Stuttgart (Rechenzentrum und UB), die TU Berlin, Sun Microsystems (Oracle), Aristotle University Thessaloniki, TU Delft, Linköpings Universitet, MathCore Engineering AB, Universität Basel, Universidad Politécnica Madrid, University of Cambridge und Computational Modelling Cambridge Ltd.

vor Ort nicht ausreichen.³⁶ Das Rechenzentrum der Universität Stuttgart koordiniert dieses Projekt, finanziert wird es im Rahmen des EU-Programms *eContentplus*.

Während BW-eLabs vorrangig – wie oben beschrieben – auf die Forschung (Forschungsdaten) abzielt, ist der Schwerpunkt bei LiLa ein anderer: Es geht hier in erster Linie um den Bereich Lehre (E-Learning).³⁷ Ziel des Vorhabens ist die Erstellung des zentralen Portals zum Zugriff auf die Ressourcen. Spezielle Komponenten des LiLa-Portals werden sein: Ein Buchungssystem, ein Tutorensystem, die Anbindung an Bibliotheksressourcen sowie die 3D-Umgebung für die Online-Kollaboration.

Die UB Stuttgart ist im Rahmen der Arbeitspakete der Universität Stuttgart in LiLa eingebunden und konkret im Bereich des Arbeitspaketes „Integration of Content“ vor allem für die Erstellung des zentralen Metadatenprofils („Library of Labs Application Profile –LiLa AP) und die An- und Einbindung der Experimente und Labore in das LiLa-Portal und die Bibliothekswelt zuständig. Dies dient primär der Suche nach Inhalten im Portal wie darüber hinaus. Die technische Architektur allgemein wie auch die Verortung der Aufgaben der UB Stuttgart werden in der folgenden Abbildung deutlich:

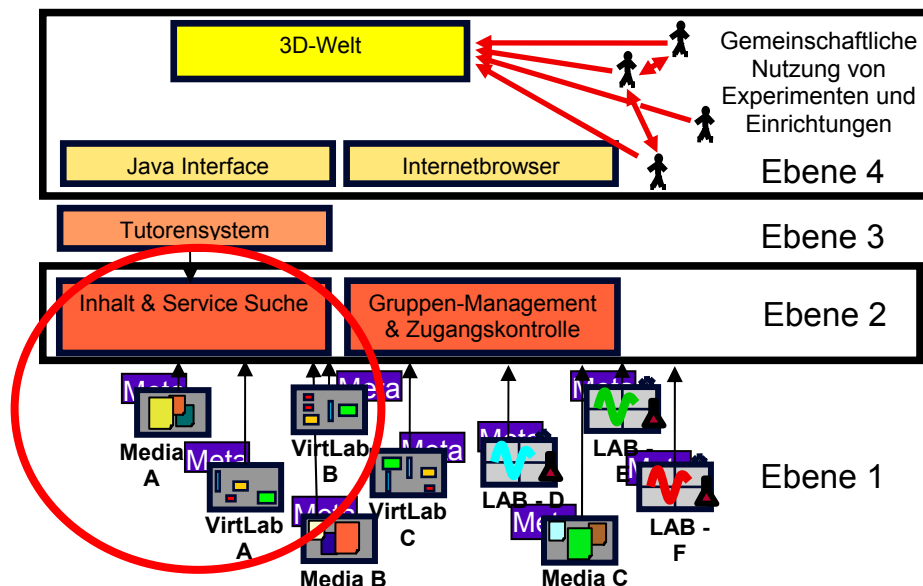


Abbildung 3: LiLa - Komponenten der technischen Architektur.

Um die Qualität und die Interoperabilität des Metadatenprofils zu garantieren, hat sich die UB Stuttgart entschlossen, eine Zertifizierung des Profils anzustreben. Dies geschieht vor allem auch, da die Absicht besteht, LiLa auszuweiten und weitere Partner und Experimente einzubinden. Hierzu wurden erste Kontakte nach Australien (University of Technology, Sydney) und in die USA (MIT) und nach Österreich („Lab2go“)³⁸ geknüpft, ebenso gibt es einen regen Austausch mit dem sich im Aufbau befindlichen „Global Online Laboratory Consortium (GOLC)“.³⁹ Vor diesem Hintergrund ist die (internationale) Interoperabilität des

36 Die Inhalte der hier eingebrachten Experimente sind höchst vielfältig und stammen u.a. aus den Bereichen der Physik (Mechanik, Ferromagnetismus, Thermodynamik), der Nanotechnologie, den Ingenieurwissenschaften usw. Eine ausführliche Darstellung zu den LiLa-Experimenten, Laboren und Setups: <http://www.lila-project.org/content/index.html>.

37 Vgl. hierzu die beiden Veröffentlichungen Richter u.a.: LiLa sowie Richter u.a.: Library of Labs (LiLa) .

38 Vgl. <http://www.lab2go.net/>.

39 Vgl. <http://www.online-lab.org/>.

Metadatenprofils umso dringlicher. Das Kompetenzzentrum Interoperable Metadaten (KIM),⁴⁰ hervorgegangen aus einem Projekt mit Partnern in Deutschland, Österreich und der Schweiz, ist hier Ansprechpartner und bringt die nötige Expertise ein.⁴¹

Nach der Erstellung der ersten Fassung des LiLa Application Profiles kam es zu einem regen Informationsaustausch mit zahlreichen Anregungen mit KIM.⁴² Diese Hinweise werden bei der Bearbeitung aufgegriffen und nach Einarbeitung der inzwischen im LiLa-Projekt gemachten Praxis-Erfahrungen wird die neue Fassung zur Zertifizierung voraussichtlich bis Jahresende 2010 eingereicht werden.

Ausblick und Zusammenfassung

Die hier vorgestellten Vorhaben im Bereich „Forschungsdaten“ und „Virtuelle Forschungs-umgebungen“ sind für die Universitätsbibliothek Stuttgart Neuland. Und auch wenn – wie oben beschrieben – zur Zeit sehr viele Projekte mit bibliothekarischer Beteiligung im Umfeld dieser Themen begonnen werden, so handelt es sich ohne Zweifel insgesamt noch um innovative und zukunftssträchtige Aufgaben, die keineswegs unmittelbar und in nächster Zeit die Informationsinfrastruktur-Landschaft allgemein prägen werden.

Dass aber mit diesen und anderen Vorhaben Themen aufgegriffen werden, die sowohl der Wissenschaft wie auch den sie unterstützenden Einrichtungen auf den Nägeln brennen, das zeigt sich nicht nur an solchen Beiträgen wie dem oben angeführten Text des Wissenschaftsbloggers Cameron Neylon, sondern auch an der zunehmenden Bedeutung, die der Thematik „Semantic Web“ überhaupt zugesprochen wird.⁴³ Insofern können über die hier vorgestellten Ergebnisse – die naturgemäß nur Zwischenergebnisse sein konnten, da die präsentierten Projekte noch laufen bzw. sich teilweise sogar noch in ihrem ersten Laufzeit-Drittel befinden –, deren Allgemeingültigkeit und Folgen noch keine Aussagen getroffen werden. Allerdings kann so viel durchaus schon gesagt werden: Die Bibliotheken als wichtiger Bestandteil der wissenschaftlichen Informationsinfrastruktur in Deutschland versuchen, sich den neuen Herausforderungen zu stellen und die Zukunft – ein Stück weit wenigstens – mitzugestalten.

40 <http://www.kim-forum.org/>.

41 Ebd. heißt es: „Das Kompetenzzentrum Interoperable Metadaten (KIM) ist eine Informations- und Kommunikationsplattform für Metadatenanwender und -entwickler zu den Themen: Anwendungsübergreifende Integration von Metadaten, Nationale und internationale Metadatenstandards, Anschluss der deutschsprachigen Community an die internationale Metadaten-Community. KIM fördert und unterstützt die Erarbeitung von Metadatenstandards, die interoperable Gestaltung von Formaten und damit die optimale Nutzung von Metadaten in digitalen Informationsumgebungen mittels Lehrmaterialien, Schulungen, Beratungen.“

42 Auf dem Bibliothekskongress in Leipzig im März 2010 wurde in einem Workshop mit dem Titel „Vom Datensilo ins Semantic Web - die Interoperabilität von Metadaten“ auch die bibliothekarische Öffentlichkeit erstmals über die Zusammenarbeit zwischen KIM und LiLa informiert: Vgl.: http://www.opus-bayern.de/bib-info/volltexte/2010/921/pdf/UBS_LiL_a-AP_BID2010_Spiecker.pdf. Hierzu finden sich auch auf der KIM-Webseite umfangreiche Materialien: <http://www.kim-forum.org/material/vortraege/bidkongress.htm>.

43 Hierzu sei u.a. verwiesen auf eine im November 2009 in Köln abgehaltene Veranstaltung unter dem Titel: „SWIB09 - Fachtagung für innovative Bibliothekare“: <http://www.swib09.de/index.html>. Diese Tagung findet – auch dies bezeichnend, denn nun ist das Ganze erfreulicherweise keineswegs mehr vorrangig auf „Bibliothekare“ fokussiert, sondern viel stärker vernetzt angelegt – ihre Fortsetzung im September 2010 unter dem Titel: „Digitale Wissenschaft 2010“: <http://digitalewissenschaft.de/>.

Literatur

- Razum, Matthias, u.a.: *Research Data Management in the Lab*.
<http://or2010.fecyt.es/Resources/documentos/GSabstracts/ResearchDataManagementInTheLab.pdf>. [29.07.2010].
- Richter, Thomas u.a.: *LiLa: A European Project on Networked Experiments*. REV 2009. June 2009. Bridgeport, CT. http://www.lila-project.org/resources/files/REV_2009-en_final.pdf. [29.07.2010].
- Richter, Thomas u.a.: *Library of Labs (LiLa): Ein Europäisches Projekt zur Vernetzung von Experimenten*. In: Nicolas Apostolopoulos u.a. (Hrsg.): *E-Learning 2009. Lernen im digitalen Zeitalter*, Münster 2009, S. 268-278. <http://www.lila-project.org/resources/files/RichterBoehringerJeschke2009.pdf>. [29.07.2010].